

SPRINGER
REFERENCE

Robert Karlicek
Ching-Cherng Sun
Georges Zissis
Ruiqing Ma
Editors

Handbook of Advanced Lighting Technology

 Springer

Handbook of Advanced Lighting Technology

Robert Karlicek • Ching-Cherng Sun
Georges Zisis • Ruiqing Ma
Editors

Handbook of Advanced Lighting Technology

With 540 Figures and 75 Tables

 Springer

Editors

Robert Karlicek
Smart Lighting Engineering Center
Rensselaer Polytechnic Institute
Troy, New York, USA

Ching-Cherng Sun
Department of Optics and Photonics
National Central University
Jhongli, Taiwan

Georges Zissis
Toulouse University
Toulouse, France

Ruiqing Ma
OLED Lighting
Universal Display Corporation
Ewing, USA

ISBN 978-3-319-00175-3 ISBN 978-3-319-00176-0 (eBook)
ISBN 978-3-319-00177-7 (print and electronic bundle)
DOI 10.1007/978-3-319-00176-0

Library of Congress Control Number: 2016955455

© Springer International Publishing Switzerland 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

Introduction

This handbook illustrates the development of electrical lighting over more than one and a half centuries. The accomplishments of scientists, product- and process development engineers, application engineers, product and systems designers, architects, lighting designers, specifiers, entrepreneurs, business people, etc., have resulted in a remarkable continuous flow of breakthroughs that have shaped and reshaped the way in which mankind lights its world. It is amazing to see how for each lighting application area dedicated light sources, luminaires, and systems have been developed that are most suited and give the best performance for the users of that space.

The incandescent and the gas discharge based light technologies have dominated the twentieth century. Continuous developments from carbon to tungsten to halogen incandescent technology increased the energy efficiency and the lifetime of these light sources. In this book, you will read how materials are applied and pushed to the boundaries of their abilities in order to create the most effective and efficient light sources. The development of the technology of glass and quartz; electrical current feedthroughs through glass, vacuum, and clean gases; and the drawing of metal wires to ultra-thin coils are described by the world's most renowned scientists and industrial development engineers in the world of electrical lighting.

Also in the past century, discharge technologies combined with phosphors have added another order of magnitude to energy efficiency in combination with a significant increase in color quality, life time, and light output. These developments enabled new application areas like road lighting, industrial lighting, studio and theater, etc.

And the technology continued to develop. Ultra-high pressure mercury lamps enabled the commercialization of beamers and digital projectors. Gas discharge lamps for car headlights changed the face of cars. The last breakthrough development in conventional light sources was the metal halide lamp with ceramic technology. This innovative light source solved the color shifting phenomena of metal halide lamps in retail lighting. And with its high efficiency, long lifetime, and high lumens it became the lamp of choice for the demanding applications of shop lighting, hospitality, and other indoor usage.

It is fascinating to read how a relatively small group of scientists and product and process development engineers worldwide has been able to create such a wide range of lighting solutions for every lighting application.

Today, the importance and application of these twentieth century “analogue” lighting technologies are rapidly decreasing. During the last decade, solid state lighting has quickly started to replace conventional lighting technologies. Light emitting diodes have been on the market, as indicator lights, since the 1960s but have only just recently surpassed all mainstream conventional lighting technologies in performance. This book describes the various technology breakthroughs that were needed to create reliable and powerful efficient light sources that can be used in all application areas. Today, the LED lighting technology has turned into a game changer and is beating the conventional technologies in all aspects that are relevant for users: in basic lighting performance, such as efficiency, color quality, and lifetime; in versatility of use and possibilities to embed the light in other structures than luminaires; and last but not least in price. It is therefore widely anticipated that in the not too distant future, all of our electric lighting will be LED based.

We are witnessing a transition from the conventional “analogue” lighting technologies to “digital” lighting. LEDs are easy to control, dim and switch, and easy to combine with sensors into intelligent lighting systems. This book also describes many of the technologies that are required for smart sensors, wired and wireless connectivity, and a whole new kind of electrical, mechanical, optical, and connectivity interfaces. Intelligent lighting systems will become the back bone for smart homes, smart buildings, and smart cities. In this way, lighting will become the heart of the “Internet of Things.” Many new applications will evolve from this, creating more efficient and especially more human centric lighting systems. But also in adjacent territories, like indoor positioning, the lighting system will start playing a crucial role.

Our knowledge of what light does to the human eye and brain has developed fast in the last decade as well. Light does much more than just create energy efficient light for vision. Light has the ability to improve cognitive performance, it can energize, increase alertness, or relaxation. It can improve mood, as well as stabilize the sleep-wake cycle of people. Intelligent lighting systems will enable Human Centric Lighting and enhance “Quality of Life,” well-being, and performance of humans, by combining the visual, biological, and emotional benefits of light.

This handbook covers conventional as well as the solid state lighting technologies and unveils the consequences and opportunities of the transition from analogue to digital in the field of systems, application, energy efficiency, human factors, measuring methods, etc. It will serve as an encyclopedia for scientists, engineers, application designers, and all those who are interested in lighting technology. This handbook, with contributions from world leading experts, is truly unique in the breadth and the depth in which it explores 140 years of development in lighting technologies.

Jan W. Denneman

President Global Lighting Association

President Lighting Europe

Vice-President Philips Lighting

Preface

The Handbook of Advanced Lighting Technology is a major reference work on the subject of light sources science and technology, with particular focus on solid-state light sources – LEDs and OLEDs – and the development of ‘smart’ or ‘intelligent’ lighting systems. This last implies the integration of advanced light sources, sensors, and adaptive control architectures to provide tailored illumination which is ‘fit to purpose’ and ‘best service’ to the end-user.

The concept of smart lighting goes hand-in-hand with the development of solid-state light sources, which offer levels of control not previously available with conventional lighting systems. This has impact not only at the scale of the individual user, but also at an environmental and wider economic level.

These advances have enabled and motivated significant research activity on the human factors of lighting, particularly related to the impact of lighting on healthcare and education, and the Handbook provides detailed reviews of work in these areas. The potential applications for smart lighting span the entire spectrum of technology, from domestic and commercial lighting, to breakthroughs in biotechnology, transportation, and light-based wireless communication. Whilst most current research globally is in the field of solid-state lighting, there is renewed interest in the development of conventional and non-conventional light sources for specific applications.

This Handbook comprehensively reviews the basic physical principles and device technologies behind all light source types and includes discussion of the state-of-the-art. The book essentially breaks down into five major parts: Part 1: The physics, materials, and device technology of established, conventional, and emerging light sources, Part 2: The science and technology of solid-state (LED and OLED) light sources, Part 3: Driving, sensing and control, and the integration of these different technologies under the concept of smart lighting, Part 4: Human factors and applications, Part 5: Environmental and economic factors and implications.

Contents

Volume 1

Part I Introduction	1
History of Light Sources	3
John F. Waymouth	
History of Solid-State Light Sources	41
Oleg Shchekin and M. George Craford	
Part II Light-Emitting Diodes	71
LED Materials: Epitaxy and Quantum Well Structures	73
Zhen-Yu Li, Hao-Chung Kuo, Chen-Yu Shieh, Ching-Hsueh Chiu, Po-Min Tu, and Wu-Yih Uen	
LED Materials: GaN on Si	123
Armin Dadgar and Alois Krost	
Thin-GaN LED Materials	149
Ray-Hua Horng	
Phosphors for White LEDs	181
Chun Che Lin, Wei-Ting Chen, and Ru Shi Liu	
Component-Level Reliability: Physical Models and Testing Regulations	223
Cher Ming Tan	
Thermal Management: Component to Systems Level	239
Te-Yuan Chung	
Optical Design: Chip and Packaging	269
Ching-Cherng Sun	

Part III OLEDS/PLEDS	291
White OLED Materials	293
Yonghua Chen and Dongge Ma	
White OLED Devices	321
Dongge Ma	
OLED Optics	363
Wooram Youn, Sai-Wing Tsang, and Franky So	
White OLED Lighting Panel Manufacturing Process	385
Jeffrey P. Spindler, John W. Hamer, and Marina E. Kondakova	
OLED Manufacturing Equipment and Methods	417
Jeffrey P. Spindler, John W. Hamer, and Marina E. Kondakova	
Part IV Intelligent Lighting System Integration	443
Dimming	445
Joseph Denicholas	
Conventional IR and Ultrasonic Sensor Systems	465
J. P. Steiner	
Ambient and Spectral Light Sensors	515
Sajol Ghoshal	
Adaptive Distributed Sensing and Control Methods	535
Zhenhua Huang, Fangxu Dong, and Arthur C. Sanderson	
Lighting Control Protocols and Standards	559
Maulin Patel and Satyen Mukherjee	
Adaptive Control Technology for Lighting Systems	583
Francis Rubinstein	
Ambient Light Sensor Integration	607
Frangiskos V. Topalis and Lambros T. Doulos	
Optical Wireless Applications	635
Z. Zhou and M. Kavehrad	
Indoor Localization and Applications	665
Shinichiro Haruyama	
Integration of RF and VLC Systems	683
Michael B. Rahaim and Thomas D. C. Little	

Volume 2

Part V Applications 701

Agricultural and Horticultural Lighting 703
 Paulo Pinho and Liisa Halonen

Museum and Exhibition Lighting 721
 Jean-Jacques Ezrati

Landscape Lighting 737
 Janet Lennox Moyer

Part VI Human Factors and Performance 755

Human Vision and Perception 757
 Mahalakshmi Ramamurthy and Vasudevan Lakshminarayanan

History of Color Metrics 785
 Wendy Davis

**Color Rendering Metrics: Status, Methods, and Future
 Development 799**
 A. Žukauskas and Michael S. Shur

**Photoreception for Human Circadian and Neurobehavioral
 Regulation 829**
 George C. Brainard and John P. Hanifin

Lighting and the Elderly 847
 Eunice Noell-Waggoner

Photobiological Safety 865
 Christophe Martinsons

Educational Lighting and Learning Performance 897
 Thorbjörn Laike

Ethnic and Social Aspects of Lighting 907
 Shin Ukegawa

Part VII Energy Efficiency 919

**Energy Consumption and Environmental and Economic Impact of
 Lighting: The Current Situation 921**
 Georges Zissis

Life Cycle Assessment of Lighting Technologies 935
 Leena Tähkämö and Heather Dillon

Impact of Lighting on Flora and Fauna	957
Sibylle Schroer and Franz Hölker	
Light Pollution Reduction	991
Sibylle Schroer and Franz Hölker	
Part VIII Conventional Light Sources	1011
Incandescent Lamps	1013
Maxime F. Gendre	
Low-Pressure Gas Discharge Lamps	1065
Graeme Lister and Yang Liu	
Mercury-Vapor Lamps	1079
Heinz Schöpp and Steffen Franke	
High-Pressure Sodium-Vapor Lamps	1097
Heinz Schöpp and Steffen Franke	
High-Pressure Xenon Lamps	1105
Heinz Schöpp and Steffen Franke	
Metal-Halide Lamps	1111
Steffen Franke and Heinz Schöpp	
Ceramic Metal Halide Lamps	1125
Stuart A. Mucklejohn	
Electrodeless Lamps and UV Sources	1141
Graeme Lister	
Index	1173

About the Editors



Robert F. Karlicek, Jr. is a professor of Electrical, Computer and Systems Engineering at Rensselaer Polytechnic Institute where he also directs the Center for Lighting Enabled Systems & Applications (LESA) engineering research center. Prior to joining Rensselaer, he spent more than 30 years in industrial opto-electronics research and development, with positions at AT&T Bell Labs, EMCORE, GE, Gore Photonics and Microsemi. He has authored or co-authored over 45 journal papers and holds 40 U.S. patents in the field of opto-electronics device design, packaging and applications. He is a member of the IEEE, the Optical Society of America (OSA), the American Chemical Society (ACS), Radtech and IMAPS.



Ching-Cherng Sun received his B.S. in electrophysics from National Chiao Tung University in 1988 and his Ph.D. in Optical Sciences, from National Central University (NCU), in January 1993. In 1996, he joined the faculty of the Department of Optics and Photonics at NCU, and has been a Chair Professor since 2013. He founded the Institute of Lighting and Display in NCU, which is the first lighting-related institute in Taiwan. He is currently the Director of Optical Sciences Centre at NCU. Professor Sun has published more than 140 journal papers, and holds more than 60 patents. He is a Fellow of the International Society of Optical Engineering (SPIE) and Optical Society of America (OSA). His research interests include LED optics, optical modeling for phosphor, lighting design, volume holography, holographic data storage, optical information processing, optical system and optical engineering.



Georges Zissis is a Professor of Electrical Engineering at Toulouse 3 University – Paul Sabatier (France). He is President of the 63rd section of the National Council of Universities (Power Electronics, Electronics, Photonics and Systems) and Vice President IEEE Industrial Application Society. He graduated from Physics department of University of Crete in general physics in 1986. He received his M.Sc. and Ph.D.

degrees in Plasma Science in 1987 and 1990 from Toulouse 3 University (France). His primary area of work is in the field of Light Sources Science and Technology. Georges Zissis is Director of the “Light & Matter” research group at LAPLACE that employs 20 full researchers. In December 2006 he won the 1st Award of the International Electrotechnical Committee (IEC) Centenary Challenge for his work on normalization for urban lighting systems (jointly with IEEE, IET and the Observer). In 2009 he won the Energy Globe Award for France and he was awarded the Fresnel Medal by the French Illuminating Engineering Society. In 2011 he was named Professor Honoris Causa at Saint Petersburg State University (Russian Federation). He has authored or co-authored over 120 journal papers and to date has supervised 27 Ph.D. students.



Dr. Ruiqing (Ray) Ma is Director of Flexible OLED Lighting R&D at Universal Display Corporation located in New Jersey, USA. Prior to joining UDC in 2007, he was a Program Manager with Honeywell International, a Senior Research Scientist with Corning Incorporated, and a Student Intern at IBM T.J. Watson Research Center. Dr. Ma received his Ph.D. degree in Chemical Physics from the Liquid Crystal Institute at Kent State University in 2000. He has over 70 publications including 2 book chapters, and over 90 issued and pending US patents in the fields of flat panel displays and solid state lighting. He has given many invited talks, tutorials, and workshop presentations at various conferences organized by SID, OSA, MRS, SPIE, and SVC. He serves as an Associate Editor for *Journal of SID*, a Guest Editor for *Information Display*, and a Program Committee Member for SID.

Contributors

George C. Brainard Department of Neurology, Thomas Jefferson University, Philadelphia, PA, USA

Wei-Ting Chen Department of Chemistry, National Taiwan University, Taipei, Taiwan

Yonghua Chen Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, China

Ching-Hsueh Chiu Department of Photonics and Institute of Electro-Optical Engineering, National Chiao Tung University, Hsinchu, Taiwan

Te-Yuan Chung Department of Optics and Photonics, National Central University, Jhongli City, Taiwan

M. George Craford Lumileds, San Jose, CA, USA

Armin Dadgar Fakultät fuer Naturwissenschaften, Abteilung Halbleiterepitaxie, Otto-von-Guericke Universität Magdeburg, Magdeburg, Germany

Wendy Davis University of Sydney, Sydney, Australia

Joseph Denicholas Texas Instruments, Dallas, TX, USA

Heather Dillon Donald P. Shiley School of Engineering, University of Portland, Portland, OR, USA

Fangxu Dong NSF Engineering Research Center for Smart Lighting, Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA

Lambros T. Doulos Laboratory of Lighting, National Technical University of Athens, Athens, Greece

Jean-Jacques Ezrati Centre de restauration et de recherche des musées de France - C2RME, Paris, France

Steffen Franke INP Greifswald, Greifswald, Germany

Maxime F. Gendre Helmond, The Netherlands

Sajol Ghoshal ams AG, Unterpremstaetten, Austria

Liisa Halonen Department of Electrical Engineering and Automation, Lighting Unit, School of Electrical Engineering, Aalto University, Espoo, Finland

John W. Hamer OLEDWorks LLC, Rochester, NY, USA

John P. Hanifin Department of Neurology, Thomas Jefferson University, Philadelphia, PA, USA

Shinichiro Haruyama Graduate School of System Design and Management, Keio University, Kohoku-ku, Yokohama, Japan

Franz Hölker Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany

Ray-Hua Horng National Chung Hsing University, Taichung, Taiwan

Zhenhua Huang NSF Engineering Research Center for Smart Lighting, Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA

Mohsen Kavehrad Department of Electrical Engineering, The Center for Information and Communications Technology Research (CICTR), Pennsylvania State University, University Park, PA, USA

Marina E. Kondakova OLEDWorks LLC, Rochester, NY, USA

Alois Krost Otto-von-Guericke Universität Magdeburg, Magdeburg, Germany

Hao-Chung Kuo Department of Photonics and Institute of Electro-Optical Engineering, National Chiao Tung University, Hsinchu, Taiwan

Thorbjörn Laike Faculty of Engineering, Department of Architecture and Built Environment, Lund University, Lund, Sweden

Vasudevan Lakshminarayanan School of Optometry and Vision Science and Departments of Physics and Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada

Department of Physics and Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA

Zhen-Yu Li Department of Photonics and Institute of Electro-Optical Engineering, National Chiao Tung University, Hsinchu, Taiwan

Chun Che Lin Department of Chemistry, National Taiwan University, Taipei, Taiwan

Institute of Organic and Polymeric Materials, National Taipei University of Technology, Taipei, Taiwan

Graeme Lister Graeme Lister Consulting LLC, Wenham, MA, USA

Thomas D. C. Little Boston University, Boston, MA, USA

Ru Shi Liu Department of Chemistry, National Taiwan University, Taipei, Taiwan
Department of Mechanical Engineering and Graduate Institute of Manufacturing Technology, National Taipei University of Technology, Taipei, Taiwan

Yang Liu Fudan University, Shanghai, China

Dongge Ma Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, China

Christophe Martinsons Lighting and Electromagnetism Division, Centre Scientifique et Technique du Bâtiment, St Martin d'Hères, France

Janet Lennox Moyer Janet Lennox Moyer Design, Troy, NY, USA

Stuart A. Mucklejohn Ceravision Limited, Sherbourne Drive, Tilbrook, UK

Satyen Mukherjee Intelligent Enterprises, Current, Powered by GE, San Ramon, CA, USA

Eunice Noell-Waggoner Center of Design for an Aging Society, Portland, OR, USA

Maulin Patel Intelligent Enterprises, Current, Powered by GE, San Ramon, CA, USA

Paulo Pinho Department of Electrical Engineering and Automation, Lighting Unit, School of Electrical Engineering, Aalto University, Espoo, Finland

Michael B. Rahaim Boston University, Boston, MA, USA

Mahalakshmi Ramamurthy Department of Psychology, Developmental and Brain Sciences, University of Massachusetts, Boston, MA, USA

Francis Rubinstein Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Arthur C. Sanderson NSF Engineering Research Center for Smart Lighting, Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA

Heinz Schöpp INP Greifswald, Greifswald, Germany

Sibylle Schroer Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany

Oleg Shchekin Lumileds, San Jose, CA, USA

Chen-Yu Shieh Department of Optics and Photonics, National Central University, Jhongli City, Taiwan

Michael S. Shur Department of Electrical, Computer, and System Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA

Franky So Department of Material Science and Engineering, University of Florida, Gainesville, FL, USA

Jeffrey P. Spindler OLEDWorks LLC, Rochester, NY, USA

J. P. Steiner Lutron Electronics Inc., Coopersburg, PA, USA

Ching-Cherng Sun Department of Optics and Photonics/Institute of Lighting and Display Sciences, National Central University, Jhongli, Taiwan

Leena Tähkämö Lighting Unit, School of Electrical Engineering, Aalto University, Espoo, Finland

Cher Ming Tan Department of Electronic Engineering, Chang Gung University, Taoyuan, Taiwan

Frangiskos V. Topalis Laboratory of Lighting, National Technical University of Athens, Athens, Greece

Sai-Wing Tsang Department of Physics and Materials Science, City University of Hong Kong, Kowloon, Hong Kong

Po-Min Tu Department of Photonics and Institute of Electro-Optical Engineering, National Chiao Tung University, Hsinchu, Taiwan

Wu-Yih Uen Department of Electronic Engineering, College of Electrical Engineering and Computer Science, Chung Yuan Christian University, Chung-Li, Taiwan

Shin Ukegawa Lighting Business Group, Panasonic Corporation, Kadoma, Osaka, Japan

John F. Waymouth GTE Lighting Products, Marblehead, MA, USA

Wooram Youn Department of Material Science and Engineering, University of Florida, Gainesville, FL, USA

Z. Zhou Department of Electrical Engineering, The Center for Information and Communications Technology Research (CICTR), Pennsylvania State University, University Park, PA, USA

Georges Zissis LAPLACE, UMR 5213 CNRS, INPT, UPS, Université de Toulouse, Toulouse, France

A. Žukauskas Institute of Applied Research, Vilnius University, Vilnius, Lithuania

Part I

Introduction

History of Light Sources

John F. Waymouth

Contents

Introduction	5
The Role of Science in Light-Source Development and Optimization	5
The Importance of Materials in Light-Source Development and Optimization	12
The Contribution of Automated Light-Source-Manufacturing Machinery	18
The Symbiosis of Light Sources and the Electric Power System	23
The Electric Lamp Market Through the Years	31
References	39

Abstract

There are many histories of discovery, invention, and development in the electric lamp industry. One of the best I have found, from the earliest days to 1947, is a book by Arthur Bright, “The Electric Lamp Industry” (Bright 1949). Several more recent ones are listed in the references (Zissis and Kitsinellis 2009; Gendre 2003). A Google™ search under the heading “History of Light Sources” turns up a number of websites, of which a sampling is listed (<http://www.mts.net/~william5/history/hol.htm>; <http://www.invsee.asu.edu/Modules/lightbulb/meathist.htm>; <http://www.en.wikipedia.org/wiki/Light>; <http://www.edisontechcenter.org/incandescent.html>; http://www.wired.com/gadgets/miscellaneous/multimedia/2008/11/gallery_lights; <http://www.inventors.about.com/library/inventors/blight2.htm>; <http://www.ies.org/lighting/history>; <http://www.nelt.co.jp/english/products/useful/01.html>).

John F. Waymouth is retired.

J.F. Waymouth (✉)

GTE Lighting Products, Marblehead, MA, USA

e-mail: jfwaymouth@waymouth.org

Glossary and Abbreviations

Glossary

Exhaust tube	A small diameter tube fused into the stem, through which air is exhausted from the interior of the lamp and any gas filling is introduced
Flare	A short piece of tubing of which the bottom end is flared out in a skirt that is fused to the envelope the lamp in the sealing operation (usually fabricated from leaded glass)
Hard glass	A glass having a softening temperature greater than 700 °C and a service temperature ca 200 °C or less
Leaded glass	A “soft” glass containing lead oxide (commonly referred to as “lead glass” in the industry; this term is used to avoid confusion with the lead-in wire)
Lead-in wire	The connection from the circuit into the interior of the lamp (commonly referred to as the “lead wire” in the industry; this term is used to avoid confusion with the element Pb)
Mount	The completed assembly incorporating the stem, filament, and filament supports, ready to be sealed to the envelope of the lamp
Press	The part of the flare that is fused around and pressed on to the lead-in wires to make a hermetic seal
Soft glass	A glass having a softening temperature ca 700 °C and a normal service temperature ca 100 °C or less
Stem	A glass assembly comprising a flare, lead-in wires, and exhaust tube fused together in the press

Abbreviations

AEG	Allgemeine Elektrische Gesellschaft (Germany)
AEI	Associated Electrical Industries (England)
ANSI	American National Standards Institute (USA)
CFL	Compact fluorescent lamp
EMI	Electric and Musical Industries (England)
GE	General Electric (Co) (USA)
GEC	General Electric Company (England)
HPS	High pressure sodium (lamp)
LED	Light-emitting diode
LPS	Low pressure sodium (lamp)
LTE	Local thermodynamic equilibrium
MH	Metal halide (lamp)
NEMA	National Electric Manufacturers Association (USA)
OEM	Original equipment manufacturer
PCA	Polycrystalline alumina, branded “Lucalox™” by GE
RGB	Red-green-blue
TH	Tungsten halogen (lamp)

Introduction

The world does not need another compendium of “who discovered/invented/developed which lamp, where, and when?” which would necessarily be just a copying from earlier sources. Therefore, I propose to discuss that history from several much broader perspectives:

Section “[The Role of Science in Light-Source Development and Optimization](#)”

Section “[The Importance of Materials in Light-Source Development and Optimization](#)”

Section “[The Contribution of Automated Light-Source-Manufacturing Machinery](#)”

Section “[The Symbiosis of Light-Sources and the Electric Power System](#)”

Section “[The Electric Lamp Market Through the Years](#)”

Because space is limited, this discussion will focus primarily on the history involving the major light sources in use today: incandescent, fluorescent, HPS, mercury, MH, TH lamps, and LEDs. Readers interested in other sources are referred to the bibliography, particularly Bright’s book. In addition, there is nothing in this summary about the history in Russia and China, because they made little or no contribution to what is discussed in sections “[The Role of Science in Light-Source Development and Optimization](#),” “[The Importance of Materials in Light-Source Development and Optimization](#),” “[The Contribution of Automated Light-Source-Manufacturing Machinery](#),” “[The Symbiosis of Light Sources and the Electric Power System](#),” and “I” know nothing about their internal markets.

Since I had the good fortune to participate in some of (and observe more of) that history in the second half of the twentieth century, I trust it will not be taken amiss if I include a few personal reminiscences.

A word about references: the references listed herein do not necessarily indicate priority. They are frequently simply those which best support the point I wish to make.

The Role of Science in Light-Source Development and Optimization

For the most part, the role of science in discovery, invention, and development of light sources has been indirect, background, or post-development explanation and optimization. For example, it has been known since the human race emerged from the Stone Age into the Bronze Age that heated objects emit light; blacksmiths have since that time used for temperature control the fact that that light shifted from dim and red at moderate temperatures toward bright and white at high temperatures. It

was known since the work of Ohm that conducting wires got hot from the passage of current, to a degree that increased with their resistivity. Thus, early experimenters seeking to produce light by incandescence from electrically heated metals would have sought out conductors of high melting temperatures and high resistivities. Platinum and graphite were the obvious choices at the time: platinum because it could be used in air and graphite because it had a higher resistivity and high melting temperature.

By far the most important contribution of science to the electric lamp industry was the discovery of the principle of electromagnetic induction by Michael Faraday in 1831, which was the foundation of the development of the “dynamo,” the DC electric generator, through the inventions of Hippolyte Pixii (1832), Antonio Pacinotti (1860), Werner Siemens (1867), Charles Wheatstone (1867), Zenobe Gramme (1871), and Charles Brush (1876) (<http://www.en.wikipedia.org/wiki/Dynamo>). Absent the dynamo, the only source of electric power was the chemical battery, and electric lighting would have been a mere laboratory curiosity.

The numerous individuals active in inventing the incandescent light bulb were for the most part either gifted tinkerers or engineers, aware of the technical knowledge of the time, but far from scientists. They did experiments in their homes, in schools, and, in Edison’s case, in a working industrial laboratory: they repeated experiments to correct flaws and problems in previous experiments, over and over again. It is, after all, the light-source industry that has given the world of technology the term of “Edisonian Research.”

Despite the relative absence of scientific research in the initial development of incandescent lamps, it contributed significantly to optimization and refinement. Research in Germany and Austria led to means of production of refractory metals which eventually resulted in a significant optimization of the commercially successful carbon-filament incandescent lamp by the replacement of graphite with tungsten (Bright 1949, pp. 183ff).

Further optimization of the incandescent lamp resulted from the scientific work of Irving Langmuir at GE Research Laboratories in the convective heat loss from heated bodies in a gaseous atmosphere and the concomitant discovery of the “Langmuir sheath” (1912). This resulted in the development of the gas-filled incandescent lamp employing a coiled filament (to shorten the length of filament losing heat by convection). The reduction in evaporation rate contributed by the gas filling permitted higher-temperature operation of the filament for the same life and a corresponding significant improvement in luminous efficacy.

The development of the fluorescent lamp traces back to the work of Geissler in 1856 and Faraday, Crookes, and others who discovered that AC current passed through low-pressure gases would generate light of a color specific to the gas being used (Bright 1949, pp. 218ff). The Moore tube followed, employing nitrogen for golden-yellow light or carbon dioxide for white light, but significantly more efficient than the carbon-filament incandescent lamps of the time. Cooper-Hewitt used

mercury at a pressure of some torr to emit a ghastly blue-green light. Again, all of these were predominantly experimentation by technologists rather than the outcomes of scientific research.

When the development of techniques for liquefaction of air and separation into components made neon available, neon filling was aggressively pursued by Georges Claude in Europe, who invented a hollow cold cathode that had extremely long life and made a commercial reality of the use of “neon lamps” for advertising, first in Europe and then in the USA. Mercury vapor and internal phosphor coating (first innovated by Jacques Risler in France) for a greater variety of colors then set the stage for the invention at GE in the USA of the hot-cathode fluorescent lamp (Bright 1949, Chap. XIV), announced in 1938.

GE engineers had explored experimentally ranges of values of parameters such as tube diameter, fill gas, fill gas pressure, mercury pressure, and discharge current to optimize the performance of the lamps. However, there was little or no guidance from science for this optimization or, for that matter, the reason for the astonishing efficiency of generation of 254-nm mercury resonance radiation by the rare-gas-mercury discharge (60+ %). That did not become available until the pioneering work of Carl Kenty, Mary Easley, and Bentley Barnes at GE (1950–1951). Using experimental probe measurements of electron temperature in the plasma by Easley, Kenty calculated excitation and quenching rates of the mercury energy levels to show the importance of the 3P_0 and 3P_2 metastable states of mercury; his calculations showed that essentially all of the excitation into these states was transferred to the 3P_1 radiating state, effectively tripling its excitation cross section (Kenty 1950). Experiments and calculations by Kenty, Easley, and Barnes determined the elastic collision energy loss by electrons and showed the importance of the Ramsauer minimum in the elastic collision cross section of argon: it coincides with the maximum in the electron energy distribution, thus minimizing the elastic collision energy loss by the electron gas (Kenty et al. 1951). These two factors contribute notably to the efficiency of resonance radiation by the discharge.

Kenty’s work was extended in 1956 (18 years after the commercial announcement) by Waymouth and Bitter, who developed a complete ab initio model of the positive column of the rare-gas-mercury discharge (employing two adjustable constants), with which a specification of the tube diameter and length, rare gas and its pressure, mercury vapor pressure, and discharge current produced results within 10 % of experiment for electron temperature, electron density, UV output, power consumption, and maintaining electric field (Waymouth and Bitter 1956). Whereas the W-B model treated radially averaged quantities, a later model by Cayless calculated the full radial dependence of all quantities directly (Cayless 1960, 1963).

The most important aspect of these model calculations was that they permitted insight into the factors controlling efficiency. Waymouth used the W-B model to demonstrate that the sublinear dependence of UV output on power input of the

standard fluorescent lamp (and consequent decrease in efficiency with increasing power) resulted from approach of the mercury 3P manifold to thermodynamic equilibrium with the electron gas. Therefore, by increasing the electron temperature, lamps could be made with 2.5 times the output while still maintaining 80+ % of the efficiency. This was accomplished by reducing fill pressure and substituting neon and/or helium for argon, to increase the ambipolar diffusion loss rate and the required ionization rate. This insight resulted in the development of a line of high-powered fluorescent lamps that found wide application in high-bay industrial and commercial applications until they were supplanted by metal-halide lamps (Waymouth et al. 1957; Gungle et al. 1967). The same principle is still used in compact fluorescent lamps (CFL's) which are necessarily operated at high power density.

The development of phosphors for fluorescent lamps has been almost entirely empirical, guided in general by scientific knowledge, but mainly the result of patient experiments, testing activators in a variety of host crystals, and testing host crystals with a variety of activators.

An even greater passage of time (50 years!) was required before the #218 formula and process of Aladar Pacz at GE (1917) for producing non-sag tungsten wire was explained by the investigations of Ronald Koo at Westinghouse Research Laboratories (1967). See section "[The Importance of Materials in Light-Source Development and Optimization](#)" for details.

The high-pressure mercury vapor lamp was also developed empirically in the 1930s, initially at Philips in Holland, Osram in Germany, and GEC in England, first as a one-atmosphere-pressure discharge in an aluminosilicate hard-glass tube, with a luminous efficacy of ca 40 lum/W. The applied research of Willem Elenbaas at Philips through the latter half of the 1930s and the 1940s and summarized in his 1951 book (Elenbaas 1951) provided a complete understanding of this commercially important light source and demonstrated that efficiency increases with increasing power per unit length. Practically, this was realized by replacing the hard-glass burner by a shorter and smaller diameter quartz tube, resulting in an efficacy of 50 lum/W.

In the course of his applied research, Elenbaas developed the Elenbaas-Heller equation for calculating the radial temperature distribution of a high-pressure plasma in local thermodynamic equilibrium (LTE), essentially a continuity equation for the transport of heat from discharge to the wall. The divergence of the radial heat flux is equal to the net local production of heat, the difference between local electrical power input and local radiation loss:

$$-\text{Div}\{\text{grad}(\kappa T)\} = P(\text{heat}) = \sigma(T)E^2 - \epsilon(T)$$

in which κ is heat conductivity, T is local temperature, $\sigma(T)E^2$ is electrical power input per unit volume, and $\epsilon(T)$ is net radiation per unit volume (emission less absorption). This equation became very important in later years in the analysis of the

more complex plasmas of high-pressure-sodium (HPS) lamps and metal-halide (MH) lamps.

Another instance of science contributing many years after the fact to the improvement of a light source comes from the chemical photographic flashbulb, originally developed in the 1930s with aluminum combustible in oxygen by Philips in Holland. Licensed to Wabash in the USA, it was the foundation for a significant business for Sylvania, which acquired all the assets of Wabash in 1946. In the 1960s at the Bayside, New York, laboratories of Sylvania (by then a subsidiary of GTE), unpublished thermochemical calculations by Bernard Kopelman showed that the adiabatic reaction temperature of zirconium and oxygen was several hundred degrees higher than that of aluminum and oxygen. A following development program at Sylvania confirmed the superiority of Zr combustible, resulting in flashbulbs having equal output to the then-common Press-25 standard, but having less than one-quarter the volume. The new bulbs were small enough that they could be packaged in disposable magazines, of which the battery-ignited “FlashcubeTM” was the first (Fink and Shaffer 1971) of a series, followed by the “MagicubeTM,” using percussively ignited bulbs (Brooks and Kopelman 1970), and the “Flip-FlashTM,” employing piezoelectric ignition, a GE development (Blount 1976; Weber 1976).

Together with inexpensive “point-and-shoot” cameras equipped with sockets for the magazines, these developments revolutionized popular photography and permitted taking of excellent pictures by complete photographic novices. The flashbulb business was very profitable for a number of years until the chemical flashbulbs were replaced by even smaller xenon-flash lamps integral to the camera.

The original low-pressure-sodium lamps, introduced to the market by Philips in 1932, employed a discharge tube made of a special sodium-resistant glass, a neon fill at some tens of torr pressure, and a charge of sodium metal and operated at more than 100 W. Since a cold-spot temperature ca 220 °C was required to vaporize the sodium, the burner was enclosed in an unsilvered Dewar flask. The relatively high power density and fill pressure of neon was deliberately employed to *increase* the elastic collision loss by electrons in order to get the bulb hot. As a result, the luminous efficacy was only about 50 lm/W in spite of the high specific efficacy of the sodium yellow emission.

Post-World-War-II research at Philips led to greatly improved thermal insulation: the burner was enclosed in a vacuum outer jacket, which was coated with an indium tin oxide IR-reflecting film. As a result, the burner could be completely redesigned with lower fill pressures and operated at a much lower power density and still reach the 220 °C cold-spot temperature. These changes yielded much greater efficiency of generation of sodium resonance radiation and lamps delivering 200 lm/W (Elenbaas et al. 1969).

The very large resonance broadening of the sodium resonance lines at high pressure was not anticipated by the early investigators of discharges in sodium vapor, but was capitalized on when a suitable discharge tube material (translucent polycrystalline alumina, “PCA”) became available as a result of studies at GE

Research Laboratories (see Part The Importance of Materials in Light-Source Development and Optimization for details). The commercial high-pressure-sodium (HPS) lamp introduced by GE in 1964 was optimized empirically as to tube dimensions, sodium pressure, discharge current, and fill gas and pressure; the final design also incorporated mercury vapor at ca one atmosphere as a buffer gas to increase arc resistance and reduce heat conduction loss to the wall (Schmidt 1966). Despite considerable subsequent research into its plasma properties and energy balance, summarized by de Groot and van Vliet (1986), its design has changed little since.

It is somewhat difficult to track the influence of science on the development of tungsten-halogen (TH) lamps. There is an extensive summary of the scientific literature on the thermochemistry and vapor pressures of metal halides included in reports prepared as part of the Manhattan project and published as a book by L. L. Quill in 1950 (Quill 1950). In addition, van Arkel and deBoer had used an iodine cycle to purify metals in the 1920s; in a heated bulb containing a reservoir of impure metal, iodine transported metal as an iodide to decompose on a central hot filament (van Arkel and de Boer 1925). In addition, in the early days of carbon-filament lamps, the so-called “Novak” lamp employed an atmosphere of bromine to keep the bulb wall free of deposits of evaporated carbon. Necessarily, there is no mention of any of these sources in the earliest patent on TH lamps (Fridrich and Wylie 1959), nor in the first published article (Zubler and Mosby 1959), so that it is hard for an outsider to tell if any of these sources were used to guide the development.

Metal-halide lamps were invented almost simultaneously at Osram in Germany by Kuhl and Krense (1964) and by Reiling (1966) at GE in the USA. Kuhl’s application date was Aug 12, 1960, whereas Reiling’s was Jan 23, 1961. There was prior art in the area of additions of halogens and halides to discharge lamps that both patents had to get around, but it was hardly scientific. Steinmetz (1911) patented the addition of halides to mercury-pool-cathode discharges to add additional spectral lines in 1911, Neunhoeffler and (Schulz 1954) patented the addition of a halogen to high-pressure-xenon arc lamps to keep the bulb clean, and Beese and Henry (1956) patented the use of metal halides in a non-mercury-containing vapor arc to generate ultraviolet light for signaling in 1956. Since the US patent laws at the time regarded a development predictable by scientific knowledge to be obvious to one of ordinary skill in the art and therefore not patentable, it was fortunate that there was none.

Reiling (<http://www.Americanhistory.si.lighting> (click on: “Invention Factory twentieth century lighting,” then on “Laboratory: caution, inventors at work” then on “Metal halide scientific training”)) is quoted as saying that he was concerned that sodium iodide would be unacceptably corrosive to quartz until he used scientific knowledge to make thermochemical calculations convincing him that reaction would be miniscule. However, Kuhl’s patent makes a point to say that it was a surprise that sodium halides do not corrode quartz. Moreover, science may have been a bit misleading here. Although the chemical equilibrium at the quartz-salt interface may only result in a few ppm, of Na^+ in the quartz, it doesn’t stay there. Waymouth (1971) showed experimentally that negative charging of the outer surface of the quartz arc tube by photoelectrons emitted from the arc tube mounting frame caused the positive sodium ions to migrate from the inner surface to the outer

surface, become neutralized by capturing photoelectrons, and evaporate away as neutral atoms. The depletion of Na^+ ion concentration at the inner surface upset the thermochemical equilibrium and required further reaction of sodium iodide. By this process, essentially all the elemental sodium in the iodide dose would be lost in a few thousand hours, with disastrous consequences.

In subsequent applied research and development of the enormous variety of possible halide additives, two sources of scientific information were extremely valuable (to me at least). One was the aforementioned book by Quill (1950) and the so-called JANAF tables of thermochemical data (Stull et al. 1971), and the other was the NBS tables of spectral line wavelengths and intensities by Meggers et al. (1961). The former provided grist for modeling calculations in determining radial temperature profiles and species concentrations, while the latter was useful in selecting elements to try. Even so, it could be misleading as well. Frederic Koury at Sylvania initially tested scandium in the belief that it would be an excellent UV emitter based on line intensities in the Meggers tables. However, most of the UV emission comes from Sc^+ , while Koury tested it in combination with sodium iodide; the sodium reduced the arc temperature to the point that the Sc^+ concentration was low, and the emission came primarily from neutral scandium, which is a lovely multiline spectrum of white light (Koury and Waymouth 1968). The sodium-scandium combination proved to be the preferred combination in the US market, because of its $\sim 4,000$ K color temperature, matching cool white fluorescent, whereas rare-earth blends of $\sim 6,000$ K color temperature were preferred in Europe.

The widespread application of metal-halide technology generated a great deal of research (too voluminous to enumerate here) into thermodynamics and vaporization data of halide compounds, in particular into vaporization of complexes of mixed salts. There was a great deal of applied physics research in the 1970s, 1980s, and 1990s into the behavior of metal-halide discharges: highly sophisticated spectroscopic diagnostics and complex modeling calculations marrying determinations of species concentrations, including diffusional and convective transport, with Elenbaas-Heller temperature profile calculations incorporating detailed accounting of radiation emission and absorption processes. While these have provided improved understanding of the multiplicity of processes active in such discharges, I am not aware that any of them resulted in lamps of improved performance in the marketplace. Those improvements mostly resulted from empirical investigations of additives, improved processing and materials, geometric tweaks to arc tube design, as well as the substitution of PCA for quartz as the arc tube material (see section “[The Importance of Materials in Light-Source Development and Optimization](#)”).

Much of this work has been presented at conferences and published in the scientific and technical literature. Prior to World War II, the lamp industry presented papers at conferences of, and published papers in the journals of, the several Illuminating Engineering Societies around the world, mostly using these channels to advertise new or improved products while revealing as little as possible of the technical details. Postwar, especially in the discharge lamp area, presentations at scientific conferences such as the Gaseous Electronics Conference in the USA and the periodic International Conferences on Ionization Phenomena in Gases provided

avenues to exchange information with other technologists in the industry. Publications in journals such as the *Journal of Applied Physics* in the USA and the *Journal of Physics D: Applied Physics* in Europe served to advertise scientific and technological expertise and were not discouraged by the lamp companies. Other important sources are the in-house technical journals, such as the “Philips Technical Review,” the “Philips Research Reports,” the “Technische Wissenschaftliche Abhandlung der Osram Gesellschaft,” the “General Electric Review,” and the “GEC Journal.” Articles in these journals may be especially useful, since they are not so tightly constricted by page budgets as those in the general literature and hence able to discuss the material in more detail. Since 1977, the International Symposium on the Science and Technology of Light Sources has provided a venue where the science of all light sources may be discussed by a truly international gathering of scientists, technologists, and yes, inventors. The first meeting at Loughborough University in England, organized by Prof. John Raffle and senior technologists at Thorn Lighting, has led to a continuing series meeting at intervals or 2–3 years, in 2012 for the thirteenth time at Rensselaer Polytechnic Institute in Troy, New York, USA.

In sharp contrast to all of the above, the development of light-emitting diodes (LEDs) has rested entirely on the foundation of 50 years of semiconductor research and the technologies developed through that research. The development has proceeded almost entirely outside the traditional light-source companies, by people trained in semiconductor physics, chemistry, and technology. Without that background of science, LEDs would not have been possible. In recent years, the lamp companies have awakened to the threat to their established businesses and have taken a more active role in development and incorporating LEDs into marketable products.

The Importance of Materials in Light-Source Development and Optimization

The performance of every light source is limited by the capabilities of at least one of the materials of which it is composed. Accordingly, the electric lamp industry has been involved in the search for improved materials since its birth; it has been practicing “materials science” long before that has been recognized as a specialty.

Although carbon in the form of graphite was selected as an emitter by nearly all the early inventors of incandescent lamps, there was a wide variety of precursor materials used to obtain the graphite filaments; Edison carbonized “Bristol Board” paper strips and Swan carbonized cotton thread and later, in 1883, carbonized extruded nitrocellulose. Although Edison’s original commercial lamps were derived from carbonized paper, he very shortly switched to bamboo fibers, which were used commercially as the filament precursor for a dozen years, because he could make more efficient longer-lived lamps with it.

As a personal aside, some years ago on a visit to Japan, I was taken by my host, Makoto Toho, to a small park outside of Kyoto, a grove of bamboo. In the center of the park, reached by a short path, was a pedestal atop which was a bust of Thomas

Edison. Toho-san told me that this commemorated the fact that this was the grove from which Edison got his original bamboo (Hachiman bamboo) (<http://www.nelt.co.jp/english/products/useful/01.html>).

Incandescent bulbs in the early days were hand-blown from a leaded soft glass; this glass had a very wide working temperature range which made it ideal for manual glass-working activity, since it allowed the glass-blower a long time to complete the work. This was not an ideal glass for automatic glass-blowing processes because it takes too long for the glass to set before it can be discharged from the mold, making a high-speed machine uncomfortably large. After World War I, GE and Corning jointly developed an alternate soft glass from silica sand, soda ash, and limestone, “soda-lime glass” or just “lime glass,” with a narrower working temperature range much more suitable for automatic bulb-blowing machinery (see section “[The Contribution of Automated Light-Source-Manufacturing Machinery](#)”). This glass is still in use today for incandescent and fluorescent lamps.

The most important material improvement in incandescent lamp technology was the replacement of carbon by tungsten as a filament material. This was merely the last refractory-metal substitute in a chain including osmium, tantalum, molybdenum, and their alloys. The early refractory-metal filaments were made by extruding metal powders in a binder into filaments, burning out the binder, and sintering the metal to a solid. Although they were extremely fragile, commercial lamps using them were manufactured in Europe by a number of companies (the “Osram” firm’s name is a composite of “Osmium” and “Wolfram,” the German word for tungsten). Sintered-tungsten-filament lamps were introduced in the USA by GE in 1907 under patent rights purchased from the German Welsbach company (Bright 1949, p. 190).

They were rather quickly superseded by filaments made from ductile tungsten drawn into fine wires. William D. Coolidge at GE Research Laboratories developed a process for preparing tungsten wires by hot-swaging a sintered-tungsten ingot to produce a fibrous grain structure oriented along the axis of the billet. The resulting rod could then be drawn in a narrow range of temperature (temperature decreased as drawing proceeded) through diamond dies to diameters less than a few micrometers. The drawing process further enhanced the fibrous grain structure to make the wire as flexible as a rope but which becomes rigid upon recrystallization by heating to service temperature. Lamps using this wire went on the market in 1911, and a patent covering this high-level-blacksmith process was issued in 1913 (Coolidge 1913). The primary advantage of tungsten over carbon is its lower vapor pressure, which permits the filament to be operated at a higher temperature for equal life, yielding a higher efficacy.

An equally important materials development (although from a cost rather than performance perspective) was one of the earliest composite materials, the so-called “dumet” wire for glass-to-metal sealing, introduced into lamps in 1913 by GE in the USA and used around the world to this day for sealing to soft glass. Developed by Colin G. Fink of the GE Research Laboratory, this consisted of a core of a nickel-iron alloy having a thermal expansion coefficient less than that of glass, surrounded by a copper cladding having an expansion coefficient greater than that of the glass, in such proportions that the composite has the same expansion as the glass (Coolidge

1913). Prior to this development, the only satisfactory lead-in wire material was platinum, which was so expensive that burned-out bulbs were recycled to salvage the precious platinum wires.

The introduction of the gas-filled incandescent lamp as a result of the work of Langmuir required that the tungsten filament be coiled to shorten the length of filament to reduce the loss of heat by convection. The ductile-tungsten filament produced by the Coolidge process had a distressing tendency to “sag” in operation, causing turns to short out, develop hot spots, and fail. This was due to grain growth at high temperatures with grain boundaries extending perpendicular to the axis of the wire. Slip could occur along these boundaries, “offsetting” the crystals causing the wire to elongate and sag. This problem was solved through the work of Aladar Pacz, at the GE Lamp Division in Cleveland, who tested many additives to prevent sag, in a classic “Edisonian Research” procedure.

On the 218th test, incorporating a dopant including potassium and silica, he achieved success (Bright 1949, pp. 207, 325). Upon heat treatment at the recrystallization temperature, the crystals that grew were elongated along the axis of the wire and had irregular side walls which “interlocked” with adjacent crystals. Thus, there were no perpendicular grain boundaries, and the locking of the crystals together prevented sag. There are several apocryphal tales about this development. In one, Pacz had written the formula on the cuff of his shirt sleeve at the time of the experiment; the shirt had been sent to the laundry by the time he received the results of the wire test and had to be hastily retrieved before the formula was washed out. Another tale has the formula scribbled on the inside of a matchbook cover that had to be retrieved from the trash heap. In any case, the number “218” was used by GE as the commercial designation of its “non-sag” tungsten wire. Although Pacz received a US patent for this work (Pacz 1922), the patent was later invalidated in the courts, which claimed that the invention was anticipated in the Coolidge patent (1913). In the specification of the Coolidge patent, the inventor described as an essential part of his process the heating of tungsten oxide in “Battersea crucibles” for the express purpose of contaminating it with substances evolved from the crucible. If the degree of contamination was less than 0.8 %, a repeat firing in a new crucible was required. Since the Battersea crucibles were glazed with a potassium silicate glaze, it was held that Coolidge had anticipated the invention of Pacz.

There was no fundamental understanding of how the Pacz process worked until the work of Ronald Koo at Westinghouse Research Laboratories in 1967 (Koo 1967) (ca 50 years later). By transmission electron microscopy of thin sections of the recrystallized tungsten, Koo was able to demonstrate the presence of “strings” of nano-voids (ever after known as “Koo Voids”) or “bubbles” containing traces of potassium in the metal. The inclusion of potassium silicate in the doped tungsten oxide resulted in small “globules” of elemental potassium in the tungsten ingot after sintering in hydrogen. Potassium is not soluble in tungsten and does not form alloys with it. In subsequent swaging and drawing steps, the spherical globules are elongated into “needles” extending along the axis of the rod or wire. When the tungsten is annealed, these needles break up into strings of potassium drops. Further drawing and annealing operations elongate the drops into fibers and the fibers into nano-

drops, continuing in repeated cycles of drawing and annealing; the resulting strings of potassium bubbles in the finished wire prevent grains from growing in the transverse direction when the wire is recrystallized and forces the crystals to grow along the axis to form the necessary interlocking structure. The voids themselves also serve as sinks for dislocations in the tungsten crystals themselves, increasing the integrity of the crystal structure.

As noted already, the discovery and optimization of luminescent materials (“phosphors”) for fluorescent lamps is primarily an empirical process, testing various compounds (“host crystals”) and various luminescent centers (“activators”). In such materials the only limitation on the ingredients is that the host crystal must be transparent to the 254-nm ultraviolet radiation of the mercury-rare-gas discharge, whereas the activator must absorb it. There are literally millions of possible combinations. One of the most important phosphor developments was the calcium halophosphate phosphor activated with antimony and manganese, discovered by McKeag and Ranby at GEC in England in 1942 (McKeag and Ranby 1949). It literally saved the fluorescent lamp industry.

Prior to this development, the phosphors employed were calcium tungstate (blue) and zinc beryllium silicate (yellow) blends to make white of various color temperatures. At the time of the commercial introduction of the fluorescent lamp, the toxic properties of beryllium compounds were not recognized. Hundreds of workers preparing phosphors and manufacturing lamps were afflicted with pulmonary beryllicosis (similar to silicosis but much more severe) and many died as a result. In addition, disposal of lamps at the end of life created a breakage hazard, releasing the phosphor to the environment. Had not the halophosphate become available, it is extremely likely that production of fluorescent lamps would have had to shut down. The halophosphate compounds are no more toxic than the enamel on your teeth, to which they are chemically similar. By varying the ratio of antimony (which fluoresces blue) to manganese (which fluoresces yellow), the same range of “white” color temperatures as before could be achieved. Moreover, the quantum efficiency of halophosphate phosphors was greater than that of the prior phosphors, and they had better maintenance of light output. Within a short time, all manufacturers changed to the new phosphor, and the episode remains mostly forgotten except by the relatives of the victims.

Another major phosphor development was the so-called “three-band” phosphor system at Philips by Vrenken and co-workers in the mid-1970s (Vrenken 1978). These were aluminate compounds activated with rare-earth ions: Eu^{2+} for blue, Tb^{3+} for green, and Eu^{3+} for red. Characteristically these luminescent centers radiate only in narrow bands located at or near the wavelengths for maximum stimulus of the RGB cones of the human visual system. In addition, the quantum efficiency of all of these phosphors was greater than 95 % and they responded well to the 185-nm resonance radiation of mercury. In consequence, lamps employing them had much higher efficacy and much improved color quality than the then-standard halophosphate lamps. Efficacy of the standard 40-W lamp could be increased from 80 to 100 lum/W. These phosphors were also much more durable, having excellent lumen maintenance even at high wall loadings. A variety of other host-crystal:Eu,Tb

combinations have since been developed, mostly in Japan, and this family of phosphors has largely replaced the former halophosphate types despite their higher cost. CFLs in particular would not be practical without the rare-earth phosphors because of their high wall loadings.

Another seminal materials development was that of translucent polycrystalline alumina (PCA) by Robert Coble at the GE Research Laboratories (Coble 1962). He discovered that by adding a small quantity of a sintering aid such as MgO to alumina, it could be sintered to full density without excessive grain growth. Essentially all pores could be eliminated, greatly increasing the transparency of the alumina, although it remained translucent because of scattering from boundaries between randomly oriented birefringent crystal faces. The small crystallites gave the material excellent strength and good thermal shock resistance, and its nonreactivity to high-temperature molten and gaseous sodium made it a perfect arc tube material for high-pressure-sodium (HPS) lamps. GE trademarked its family of PCA products as "LucaloxTM" and designated its versions of HPS lamps with the same trademark. All HPS lamps today use this material. It is also used in many MH lamps because it can operate at higher temperatures than quartz and is not permeable to sodium.

Practical commercial implementation of metal-halide (MH) lamps required a definite improvement in the halide salt materials. Hydrogen is a detrimental impurity in many types of lamps, but in MH lamps it is fatal. It reacts with halide salts to produce hydrogen halides, which remain in the vapor state down to extremely low temperatures. Accordingly, at ignition a hydrogen-contaminated arc tube has a gas filling containing electronegative hydrogen halide vapors which capture free electrons from the Townsend avalanche, increasing the required ignition voltage to a value greater than the open-circuit voltage of the ballast, and the lamp does not ignite. In the early days of MH lamp development, the best salts available were just "chemically pure"; all are strongly hygroscopic and readily absorb water from the environment. Moreover, this water cannot be removed by simple vacuum baking. When the salt is heated, the water reacts with it to form the hydroxide plus hydrogen halide. Although the hydrogen halide could be pumped off, the vapor pressure of the hydroxide (also containing hydrogen!) is comparable to that of the halide salt and it cannot be separated by distillation.

A great deal of effort was put into developing processing techniques to try to remove water and hydroxide from the salt additives; in a word, they were cumbersome, suitable only for laboratory implementation and not amenable to automated manufacturing.

One of my greatest contributions to MH lamp technology was the discovery of Scott Anderson, founder and principal investigator of the Anderson Physics Laboratory of Champaign-Urbana, IL. Anderson had developed a process for dehydrating sodium iodide (for scintillation-counter crystals) which reversed the above reaction by passing a carrier gas plus HI through molten NaI. The HI reacted with the hydroxide, reforming NaI and water vapor, and the carrier gas flushed the water vapor out. In this way, all traces of water and NaOH could be removed;

by subsequent handling of the material only in a dry box, further contamination could be avoided.

Anderson advertised his product in *Physics Today*, where I saw it, got some, and tried it with great success. Subsequently, on consulting contracts with Sylvania (which literally kept him in business; he had no other customers at the time), Anderson developed an additional process of draining the dehydrated molten salt through a funnel vibrated at a specific frequency by a loudspeaker. This broke up the flowing stream into droplets whose size decreased with increasing frequency. The end product was a hydroxyl- and water-free salt in the form of a pourable powder (Anderson 1972), which permitted processing of MH lamps on a standard mercury-vapor-lamp exhaust machine, just by the addition of a simple dry-nitrogen-purged salt-powder dispenser. The resulting much-reduced cost of manufacture was an important factor in Sylvania's commercial success in the MH market.

Today, APL Engineered Materials, the successor to Anderson Physics Laboratory, sells its dehydrated salts to MH lamp manufacturers around the world, as well as amalgam pellets for HPS and CFL lamps, and many other specialty chemicals used in lamp-making and other industries.

In an LED, electrons and holes are injected into a semiconductor material in which they recombine to produce light. The color of the light depends on the energy width of the forbidden band in the recombination layer. Materials must be selected so that the width is greater than the photon energy. The bandgap can be adjusted by varying the ratios of elements of which the crystals are composed. Although internal quantum efficiencies may be greater than 50 %, the escape of light is controlled by total internal reflection dependent on the index of refraction. The nearer the photon energy to the bandgap, the higher the index of refraction and the greater the total internal reflection. Because the only satisfactory materials to date are quite expensive, the necessary layers are grown epitaxially on crystalline substrates. So long as the lattice spacing of the substrate and epitaxial layers are the same or nearly so, there are few dislocations in the lattice of the epitaxial layer; this is extremely important for efficient lamps because dislocations provide sites for non-radiative recombination of the electrons and holes. There is a suitable substrate for InGaN, a satisfactory material for blue emission, and for AlInGaP, a satisfactory material for red emission. There is neither an efficient green-emitting material nor a suitable substrate. Therefore green LED's achieve only about one-third the internal quantum efficiency of red or blue LED's. So-called RGB LEDs have spectra similar to those of three-band fluorescent lamps and could have high luminous efficacy and excellent color quality; the poor efficiency (the "green gap") of current green LEDs prevents achieving this result. Therefore, the only efficient white LEDs at present use a blue diode and a yellow phosphor and have a color quality similar to that of halophosphate-phosphor fluorescent lamps (Protzman and Houser 2006). One may be quite sure that these materials problems are being aggressively investigated in the laboratories of LED producers and may well have been overcome by the time this article is published.

The Contribution of Automated Light-Source-Manufacturing Machinery

Thomas Edison employed glass-blowers to produce the bulbs and stems required for the carbon-filament lamps the Edison companies were selling. There would not have been enough glass-blowers in the world to produce the bulbs for the 531,000,000 incandescent lamps sold in the USA alone in 1926 (Bright 1949, Table XIV, p. 247). It is plain that automated machinery was required for lamp-making, not only for reasons of cost, but simply to produce the volumes the market required.

Except for glass-blowing, much of the earliest machinery was developed in-house by GE, simply because the machine requirements were unfamiliar to anybody else, and GE's profits (see section "[The Electric Lamp Market Through the Years](#)") could support a captive machine-development group. 1907 saw the development of an automatic machine for welding lead wires, 1914 automatic stem and mount machines and the first rotary-indexing sealing machine. (An indexing machine is a rotating assembly with work-holders around its periphery that advances the work in steps through a series of stations followed by a pause during which operations at each station are carried out on the work, followed in turn by another advance to the next station.) The first automatic rotary-indexing exhaust machine was developed in 1915 (Bright 1949, Appendix I, p. 498).

The heart of such a machine is the "plate valve," a pair of contacting thick steel plates, one affixed to the chassis, the other rotating with the armature. The rotating plate has a set of ports, one for each work-holding head and connected to it by steel, glass, or rubber tubing (the "sweep"). The stationary plate has a set of ports ducted to the vacuum system, the flush gas system, or the fill gas system, respectively. As the machine indexes, the upper ports are brought into register successively with the stationary port appropriate to the process step being carried out at that station. The very thin space between plates is lubricated and vacuum-sealed by a film of low vapor-pressure vacuum oil feeding from a central reservoir and seeping to the outside edge. The "health" of the machine is determined by the flatness and register of the plates and is monitored by the consumption of vacuum oil. Any oil which escapes into a sweep is prevented from reaching the corresponding head by an oil trap.

Prior to 1921 the individual steps in fabrication and processing were carried out in individual departments: stem-making, mount assembly, sealing, exhausting, basing, and packing. This meant moving work in process in buggies from one department to another, often on different floors of the factory, a highly inefficient work flow. Under the leadership of W. R. Burroughs at the GE plant in Harrison NJ, the departments were broken up, and the machinery assembled in "groups." A stem machine, mount machine, sealing machine, exhaust machine, and basing machine were grouped together, with work going in stages from one to another. This required that many of the machines had to be redesigned to operate in synchronism. The function of the operators was to load and unload the machines and transfer the work to the next machine. Little or no skill was required, and the physical demands were small, so

that most of the operators could be women. With this reorganization, a group could produce 300–400 lamps per hour, a vast improvement in productivity (Bright 1949, pp. 349–350). European manufacturers, particularly Philips, developed equally capable machinery as work-flow systems.

Machinery for the manufacture of glass was developed primarily by glass manufacturers, who used essentially similar machinery for bottle-blowing (at the time a much larger business); an early such machine, a rotary “turret” machine, was developed in 1910–1911 at the Westlake Co, a subsidiary of Libbey Glass Co (Kadow 1925). In its ultimate embodiment in 1926, this was a 48-head machine that could produce 5,000 bulbs an hour with one operator and mechanic for each two machines. Empire Machine Company, a subsidiary of Corning, built similar machines.

The 800-lb gorilla in this playground was the Corning “Ribbon Machine” (<http://www.files.asme.org/communities/history/landmarks/5520.pdf>). Starting with a concept of William Woods in 1921, in collaboration with David E. Gray, a prototype was completed in 1925. A stream of glass exiting from the hearth of a melting tank was flattened into a ribbon of glass which flowed out onto a belt of linked individual plates carried on a looped chain driven by sprockets. Each of the plates had a hole in it through which molten glass sagged down into clamshell molds carried on a second chain-and-sprocket system nested within the first, while a third chain-and-sprocket assembly above the first carried hollow plungers that contacted the top of the ribbon, and through which blowing air was fed into the sagging glass “gather.” The molds were internally coated with burnt cork, which was moistened during the return path to the front of the machine. As the clamshell molds closed around the sagging glass, they were rotated, and the heat of the glass vaporized the moisture into steam. The blowing air forced the glass to assume the shape of the molds, but the layer of steam prevented it from actually contacting the mold. Thus, the blown ware acquired the shape of the rotating mold, but exhibited no mold marks.

After progressing a suitable distance for the glass to “set,” the clamshell molds were opened and retracted, and their chain returned them to the front of the machine. The bulb chain continued on for a few additional feet before the bulbs were cracked off at the plates and transferred into an annealing oven. The importance of quick-setting lime glass (section “[The Importance of Materials in Light-Source Development and Optimization](#)”) to the success of this system is apparent. This machine could produce as many as 2,000 bulbs *per minute*. Naturally, it displaced all other types of machines used for bulb-blowing in volume, although turret machines still continue to be used for low-production-volume specialty products. It was declared in 1983 to be an “International Historical Mechanical Engineering Landmark” by the American Society of Mechanical Engineers (ASME). At that time, it was reported that 15 ribbon machines were supplying all the bulbs used in the world for incandescent lamps.

A 1950 movie showing this machine in operation can be viewed at [youtube.com/watch?v=MD1BGUrk9M](https://www.youtube.com/watch?v=MD1BGUrk9M).

It can also be accessed by a Google™ search under “Corning Ribbon Machine” selecting “Movie-Glass Bulbs Ltd.-The Ribbon Machine -1950-YouTube.”

Two other glass-forming machines that would prove very important to the fluorescent lamp industry were the Danner (1916) and Vello (1935) tube-drawing machines. The Danner machine, invented by Edward Danner of Libbey Glass Co., employed a hollow conical rotating mandrel angled at about 45° downward onto which glass flowed in a stream from the forehearth of the glass tank. The rotation and the gravity flow down the mandrel to the narrow end formed the flowing glass into a tube which was prevented from collapsing by air pressure injected through the hollow center of the mandrel. Gravity then directed the hollow glass tube onto a horizontal track 120 m long, at the end of which a “tractor” pulled the solidified glass tube. Production speeds of a variety of tube diameters could be as high as 400 m per minute. The Vello machine, developed in Holland and later licensed to Corning by Philips, dispensed with the mandrel and flowed glass vertically downward through an annular orifice into a tube (again prevented from collapsing by internal air pressure) which was bent at right angles to deposit on a horizontal track essentially similar to that of the Danner draw. A central feature of the Vello draw is that the center core of the annular aperture is eccentrically mounted, displaced by an amount sufficient to compensate for the fact that the glass at the top of the tube is stretched less in going around the bend to the track than the glass at the bottom and would otherwise be thicker. Since the Vello machine is less complex and produces tubing somewhat faster than the Danner, it has supplanted the Danner in many glass plants. <http://www.eurotherm.com/industrial/glass/tube-glass> is a website showing diagrams of these machines.

As part of its licensing agreements, GE sold lamp-making machinery equal to its own to its licensees, with an option to buy it back if the licensee violated any terms of the license. Small non-licensees had to buy less-efficient slower machinery from independent suppliers such as Hoffman and Eisler. When Sylvania (a “class B” licensee; see section “[The Electric Lamp Market Through the Years](#)”) jumped into the fluorescent lamp market in 1939, it knew that its machinery would be repossessed and therefore established its own Equipment Development Group and later a tungsten wire-drawing facility at its phosphor plant at Towanda, PA. Fortunately, the intervention of World War II delayed the inevitable legal challenges by GE, which gave the Sylvania group time to develop machinery so that its production was able to continue without a hitch.

A major accomplishment of this group was the development in 1945 of a unique “automount” machine for producing stems and mounts at high speed on the same machine (Gardner 1953). Prior to this time, stems were produced on one machine, sent through an annealing oven to relieve the residual stresses caused by unequal cooling rates of the different parts of the press, and then passed on to the mount machine. Sylvania’s machine incorporated a special process step in which the hot glass of the press was deliberately cooled by air jets. The surface of the glass was therefore “set” while the inside was still cooling; this resulted in the surface stress being entirely compression, in which glass is very strong. In consequence, no annealing was required to prevent stem cracks, and the entire stem/mount operation could be performed on one machine. A second feature was that this was a sprocket-and-chain-driven indexing machine. Because of the low mass of the chain, indexing

could be carried out at much higher speeds than on a rotating machine. Prior to this time, chain systems had been deemed impractical for indexing machines because slack and wear in the chain would prevent the accurate location of the work holder required for precise lead-bending and filament mounting. The Sylvania machine took advantage of the fact that precision location was required at only a limited number of stations; some of these could be located at the sprocket, which would accurately locate the work holder. Others were equipped with movable jaws on the chassis which could grasp the work holder at the end of the index and fix it in position for the operation at that station. Unlike any previous machine, this one could produce complete stems and mounts, for either incandescent or fluorescent lamps, at index speeds of 3,600 per hour and above.

Rotary-indexing machines require a large mass of metal to be accelerated rapidly and then stopped equally rapidly. This requires the application of considerable force for both operations. Force times distance equals work and energy, and work divided by time equals power. Thus, power consumption limits the maximum index rate to ca 1,200–1,500 indexes per hour, with consequent limits on production rates. This was solved by machine designers by developing “dual-index” machines: each index moved two lamps in process from one station to the next. By this technique, reasonably sized machines could produce 2,500 lamps per hour.

Rotary-indexing exhaust machines for fluorescent lamps were even more cumbersome and slower than those used for incandescent lamps. They had to be much larger, simply because the lamps were larger, and they had a much larger volume of lamp to evacuate and fill with the proper fill gas. Indexing speeds are typically in the range of 600–1,000/h. An industry-leading development at Sylvania was the so-called “horizontal exhaust machine” a continuous-motion machine operating at 3,600+ lamps per hour (Dodge and Kimball 1955).

Continuous-motion permitted much faster rotation speeds than any indexing system, but drastically reduced the time available to exhaust the air from a large-volume lamp and replace it with fill gas, from ca 3 min on a 36-head machine indexing at 600/h to less than 30 s for a machine operating at 3,600 lamps per hour. A drastic modification of lamp-processing technique was required. This was obtained by equipping the lamp with an exhaust tube at each end and flowing an inert gas starting at atmospheric pressure in one end while pumping at the other to push the air and contaminating gases out of the lamp. Because the pressures were high, the throughput in liter-atmospheres was high and a 2-l volume 8-ft long tube could have air and contaminating gases removed and replaced by the fill gas at 2–3 torr in the required time.

The high speed of the system made manual transfers from mount machines and sealing machine impractical, so this was designed as an integrated system from the beginning. Coated bulbs are fed into a horizontal continuous-motion baking oven to remove the coating binder and are immediately conveyed from there automatically to a horizontal sealer where they are united with stems, automatically fed from an automount at each end. From the sealer, they are fed automatically to the horizontal exhaust machine and from there to a basing and testing reel and thence to automatic packaging. Despite the accelerated exhaust processing, this system produces higher-

quality lamps than the former indexing machines, because the bulbs never cool below ca 250 °C after emerging from the baker before they are sealed off at the conclusion of exhaust. This prevents moisture or contaminants from the factory atmosphere from condensing on the phosphor as it does in the indexing-machine system. Indeed, the bulbs in that system have to be forcibly cooled to permit them to be manually loaded into a sealing machine and then to the exhaust machine.

Sylvania has used the horizontal system for all its high-volume production in the USA and internationally while retaining indexing-type machinery for lower-volume types. Because of the very large reduction in total direct cost (materials plus direct-labor plus direct-labor overhead) permitted by the high production speeds and the degree of automation, other manufacturers have developed similar machinery.

In the climax phase of incandescent lamp manufacture, the machine “groups” were also integrated with automatic transfer of work in process from one machine to the next, further minimizing direct-labor costs. Very complex transfer systems are required (Gardner 1955). A typical arrangement would have an automount machine and a “smoke-coating” machine (applying nanoparticle silica electrostatically to the inside of the bulb as a light-diffusing layer) both operating at 3,000/h automatically feeding a continuous-motion sealing machine operating at 2,750/h which automatically feeds a dual-index exhaust machine operating at 1,250 indexes per hour. This then feeds a basing machine followed by a testing machine and an automatic packaging machine. The upstream machines operate at slightly higher speeds in order to insure that every work-holding head would be filled on the slowest machine. This means that the transfer systems have to function as storage banks as well as conveyors. If a storage bank is full, heads are locked out on upstream machines until the overflow is rectified. Operators are needed only to keep the parts hoppers full at the input. Typically, the total direct cost of manufacture with such a system is less than 10 % of the retail price of the product. All the rest is overhead and distribution costs.

A major “revolution” in coiling machines for primary coiling of tungsten wire was developed at Philips, the so-called “pot-flyer” machine (Govaert 1970). A vitally important factor in the fabrication of tungsten coils is the maintenance of constant tension on the wire as it is coiled, because the wire is stretched slightly in the process. If the tension varies, the stretch varies and the wire diameter varies. Thin spots along the wire overheat in service and develop “hot spots” at which the filament ultimately fails. Earlier coiling machines used mechanical arrangements to pull wire from a spool and apply tension as it was coiled around a mandrel. In the pot-flyer, the spool itself is spun at high speed, throwing the wire off by centrifugal force. The wire contacts the surface of a surrounding cylinder (the “pot”) and is led to the point of coiling at which it is directed radially inward to the mandrel. In this machine the tension is controlled by the centrifugal force on the straight radial segment of wire between the mandrel and the pot plus the force of friction between the wire and pot; the tension can be adjusted by the rotation speed of the spool and by adding rotation to the pot, either counter-rotating and forward-rotating. Coiling speeds of this system are 2.5 times greater than any prior coiling machine, and tension is much more uniformly maintained.

After 1970, the patent literature on lamp-making machinery has been sparse; I think not because there has been less activity, but because lamp companies have tended toward protecting intellectual property in this area as trade secrets rather than by patents. A GoogleTM search under “lamp-making machinery” 1950–1960 turns up 52 US patents; 1960–1970, 8 patents; and 1970–1980, 9 patents. Machine and process patents are difficult to enforce unless the lamp construction itself provides evidence of infringement. Unless you know a patent is being infringed, you can’t get a court order to enter a competitor plant to prove it. Thus, all a patent accomplishes is to provide a competitor with detailed instructions on how to infringe it. My knowledge of lamp-making systems since 1970 has been acquired only under conditions where confidential-disclosure agreements would apply and therefore cannot be discussed. Suffice it to say that continuous development and refinement of manufacturing machinery make electric lamp manufacture of all types among the most efficient and high quality for any product family.

Light-emitting diodes can make use of similar equipment to that used in high-volume semiconductor processing: automated chambers for epitaxial deposition, plasma processing, dry-etching, etc., with fully automated transfers from station to station. In volume production, therefore, the major item of direct cost is likely to be the cost of the materials: gallium is expensive.

Despite the importance of lamp-making machinery to the production of quality products at minimal cost, both Sylvania and GE in the USA have abandoned the field. Sylvania discontinued its Equipment Development Group prior to its acquisition by Osram, and GE spun its group off as an independent company as well. In consequence, neither developed automated machinery for fabricating the spiral CFLs for residential replacement of incandescent lamps. These lamps are almost entirely made in China with low-wage labor and according to anecdotal reports vary widely in quality. I do not know the status of this activity elsewhere in the world.

The Symbiosis of Light Sources and the Electric Power System

As already noted, electric lamps without a cost-effective reliable system of delivery of electric power would be only a laboratory curiosity. Thomas Edison won the incandescent lamp battle not because his lamps were first, or even the best, but because he recognized this from the beginning. Series-connected arc lamps had been used for street lighting as early as 1858 in England and 1863 in France, each installation with its own generator. Edison recognized that for widespread use, especially residential, it would be necessary to be able to switch any individual lamp on or off without disturbing the operation of others; ergo, they had to operate in parallel. Moreover, it would be ridiculous to expect every home to have its own generating system. Electricity had to be generated in central power stations and delivered to the customer’s premises by a distribution system. He realized as the result of back-of-the-envelope calculations that this meant the lamps had to be high resistance, absorbing power at high voltage and low current (Bright 1949, pp. 30 & ff, 62 & ff). Other carbon lamps, consisting of a graphite rod as the radiator operating

at high current and low voltage, would require much too large diameter for the copper wire distribution mains.

Hence, his experiments focused from the beginning only on high-resistance filament designs, eventually settling on lamp designs to operate at 110 V. Since the only generators of the time were direct current, for which convenient transformation of voltage is not available, this meant that the generators had to provide 120 V, allowing 10 V for line losses; the central generating stations could then be several miles apart.

From the beginning he saw electric lighting as a system. Rather than trying to sell light bulbs through retail channels, he established Edison Electric Companies in various cities to generate power, distribute it, and provide the bulbs to the customers. This mode of distribution persisted for many years. As late as the 1950s, Chicago Edison was providing bulbs to its residential customers as part of the electric bill. In the early days of manufacture, carbon-filament lamps were not sufficiently uniform and had to be sorted into as many as six voltage ranges; necessarily, the generating and distribution system had to provide the proper voltage for the range bin of bulbs it was supplying (Bright 1949, p. 197 footnote). Even at the end of the carbon-filament era, when graphite was replaced by tungsten, there were still two voltage ranges.

In Europe, on the other hand, the system voltage was fixed at 220–240 V to further reduce transmission losses. However, incandescent lamps are less efficient at the higher voltage, because the thinner filament cannot be operated at as high a temperature if the same life is to be achieved.

Although Edison showed great insight in treating electric lighting as a system, he did not show similar perspicacity in the “War of the Currents.” The first efficient AC transformer system was invented in Hungary in 1884 by the “ZBD team” of Charles Zipernowsky, Otto Blathy, and Max Deri. This permitted generation at one voltage, long-distance distribution at a much higher voltage and lower current, and local distribution and application at a third. The Edison Electric Light Co. got exclusive rights to use Zipernowsky et al.’s patents in the USA and paid \$20,000 for an option to purchase any US patents which issued. Edison himself vehemently opposed the use of alternating current and the Edison company never exercised its option. The advantages of AC distribution in reducing transmission losses were widely recognized and capitalized on in Europe and quickly supplanted DC there. In the USA, the prime mover was Westinghouse. Using patents of William Stanley, Westinghouse installed the first AC lighting system in Great Barrington, MA, in 1886. Westinghouse also acquired the US rights to Nikola Tesla’s AC induction motor, which also increased the utility of AC current (Bright 1949, pp. 98–100).

Despite Edison’s opposition, by 1889 there were 150 alternating-current systems serving 300,000 lamps installed by Westinghouse. In Europe, at the 1891 International Electrotechnical Exposition, 225 kW of 3-phase AC power generated in Lauffen am Neckar was delivered 176 km to the Westbahnhof in Frankfurt. Edison emphasized that AC electric shock was much more hazardous than DC because it caused heart fibrillation where DC did not and lobbied to have the first electric chair, installed at Sing Sing prison in New York, powered by AC as a public relations play.

Nevertheless, by 1896 the battle was lost in the USA, when the Niagara Falls power station went on line with alternating current. Pockets of DC distribution persisted in some cities for many years later, a fact I became rudely aware of in 1949 when I blew out the fluorescent lamps in my desk lamp upon moving into an apartment near Symphony Hall in Boston. The choke ballast offered little impedance to the flow of DC current.

In the USA, 60 hz was settled on as a system frequency, with local distribution with 120/240 V three-wire systems. In Europe the choice was for 50 hz at 220–240 V; although there was no longer any distribution-loss reason to maintain the higher local voltage, the large installed base of high-voltage lamps and motors dictated the choice. Canada had a part of its distribution system at 25 hz, at which even incandescent amps flicker. In other parts of the world, the choice of frequencies seems to have depended on whether generating and distribution systems were bought from the USA or Europe, but 220–240 V is the most common local distribution voltage. As a personal aside, when I was in Japan in the US army of occupation in 1945–1946, one part of the country had bought from Siemens and was 50 hz, and another part had bought from Westinghouse and was 60 hz. The two grids could not interchange power. Further information about the war of currents may be found in a book by McNichol (2006).

From the turn of the nineteenth century until shortly before World War II, the relationship between electric lamps and power lines was simple and peaceful, with minor exceptions. Power companies generated power, incandescent lamps were connected to it, and most people were happy. Well, not everybody. The life of incandescent lamps varies inversely as the 13th power of the applied voltage. Voltage 10 % below rated increases incandescent lamp life by a factor 3.5 over rated; 10 % above, a corresponding factor 3.5 reduction in life. During the depression years (1930–1940), capital was in short supply, and utility companies had trouble expanding their capacity to meet what was still a growing demand. As a result, line voltage at the end of a long local loop could be 20 % below that at the transformer; utilities compensated by jacking up the transformer voltage by 10 %, so that everybody's voltage was within the specified plus-or-minus 10 %. Customers at the end of the line enjoyed phenomenally long lamp lives. Customers near the transformer got lousy life and switched bulb brands frequently, cursing as they did so, and legislative bodies mounted investigations into the “conspiracies” resulting in the short life of incandescent lamps.

The second deviation from this relatively benign symbiosis was the growing importance of discharge lamps: “neon” sign lamps, high-pressure mercury street-lighting lamps, and late in the 1930s, fluorescent lamps. These all share the characteristics that they will not ignite without assistance at the 120-V lines in the USA, and they have “negative resistance,” i.e., they cannot themselves regulate the current that flows through them. They require an auxiliary apparatus to provide, first, the voltage to ignite and, second, an impedance to limit the current, a device called a “ballast.” The ballast comprises a step-up transformer to boost the line voltage and an impedance to limit the lamp current once ignited. For about the first 50 years of discharge-lamp application, this was an electromagnetic “copper-and-iron” device,

using the fact that current was AC to employ inductive and capacitive reactance to provide current limitation.

In Europe, the higher line voltage eliminated the necessity of a step-up transformer for fluorescent lamps. A circuit switch called a “starter” connected the ballast impedance through the cathode filaments in series to the current return lead. After an interval during which the filaments were heated to emitting temperature, the switch opened. The interruption of current through the inductive impedance generated a voltage pulse which ionized the gas, and the open-circuit line voltage drove current through the series combination of inductor and hot-cathode lamp. This very simple circuit could operate the fluorescent lamp industry’s “bread-and-butter” type (“4-ft (1.22 m) 40-W”) and predominates to this day. Similar “choke”-type ballast (without starters) can ignite and operate many types of high-pressure mercury lamps and are widely used wherever 220–240 V line voltages are available. A difficulty with this circuit is that lamp and line current lag line voltage by ca 50°, and the power factor is only about 60 %. Utilities object strongly to such loads, and hence, power factor must be corrected by the addition of a parallel capacitor.

In the USA through the 1940s, the same “switch-start” technology was used in combination with a step-up transformer. The most popular circuit was a two-lamp “lead-lag” arrangement with one lamp operated from an inductive impedance, the other by a series combination of capacitor and inductor in such proportions that capacitive impedance was twice the inductive impedance. The out-of-phase components of current canceled, and power factor was greater than 90 %. The circuit had the added advantage that current zeroes in the two lamps were out of phase, which minimized the 120-hz modulation of light output and the consequent “stroboscopic effect.” For shorter lamps, such as the 2-ft 20-W, no step-up transformer is needed and the choke-starter ballast is common.

An important ballast development in the USA was the so-called “Rapid Start™” system, invented by Eugene Lemmers of GE (Lemmers 1956). In this system, two lamps are operated in series, with a small capacitor shunting one of the lamps. The open-circuit voltage required to ignite is only about 1.5 times that of a single lamp because the unshunted lamp has infinite impedance in comparison to the capacitor, so all the open-circuit voltage appears across it; once it ionizes, its impedance is much less than that of the capacitor and much of the open-circuit voltage appears across the shunted lamp, ionizing it. A second major feature of the system is that all four cathode filaments of the pair of lamps are connected to filament transformers energized by the primary winding so that filament heat and open-circuit voltage are applied simultaneously. The lamps initially ignite in a cold-cathode high-cathode-fall “glow” discharge which is converted to a low-cathode-fall hot-cathode “arc” when the cathodes heat up. The cathode filaments themselves are completely redesigned to reach the desired temperature at the low voltage (3.75 V) of the filament windings within less than 1 s, to minimize the duration of the glow discharge phase, which would otherwise cause severe sputtering damage. The filament transformers are powered continuously during operation of the lamps. Although this represents a continuous power consumption of about 1 W per filament, it is partially compensated for by a reduction in the cathode fall of the discharge,

reducing the overall “end loss” per lamp from ca 19 to ca 15 V or about 1.7 W. The lamps give low-level light from the beginning of their ignition in the glow phase, which increases rapidly to full light output when the cathodes warm up. There is no “dead time” or “blinking” as is common in switch-start circuits, hence the name “Rapid Start™.”

With properly designed cathodes and short glow discharge phase, the damage to cathodes during starting is substantially less than that of switch-start or instant-start ballasts; because the cathode fall is lower, sputtering damage during operation is minimized or even eliminated. Lamp lives on the rapid-start system were immediately increased from ca 12,000 to 16,000 h on a 3-h-of-operation-per-start cycle. In the later days of magnetic ballast usage, 3-h-cycle lamp lives increased to 24,000 h. The ballast secondary impedance is of the “leading” type with capacitive impedance ca twice the inductive impedance. This is reflected as a leading current in the primary winding, which is compensated for by reducing the primary impedance to increase the lagging magnetizing current. Therefore, the lamp and ballast system naturally has high power factor. The reduced primary turns means that fewer secondary turns are required for a given voltage step-up; moreover, the total volt-ampere to be handled by the magnetic circuit is only 75 % of that of a lead-lag ballast. On both counts, the whole ballast is smaller. As a result, the rapid-start system is no more and perhaps less costly than the lead-lag switch-start system it replaced while providing significant performance and life advantages. Essentially, this system drove all others from the market in the USA, except for choke-starter ballasts for short lamps and instant-start systems for 8-ft T8 lamps.

Despite its many advantages, it was never adopted in Europe, where it was felt that it was still more expensive than the choke-starter system in use there.

For many discharge-lamp ballasts, it is an advantage to have an open-circuit voltage with a high peak-to-rms ratio. Ignition depends primarily on peak voltage, but the lower the rms voltage, the smaller need be the current-limiting impedance. This is accomplished by including a gap in the magnetic core which is bridged by a small number of laminations (“splines”). At the phase angle of maximum voltage, the magnetizing current is nearly zero, so that the laminations in the gap have high permeability and the magnetic coupling to the secondary is large, generating a high secondary voltage. At other phase angles where the magnetizing current is different from zero, the splines saturate and the permeability of the gap is that of air; therefore, the magnetic coupling to the secondary is smaller and the secondary voltage is lower. The open-circuit voltage waveform is therefore sinusoidal over most of the half cycle surmounted by a peak approximately double that of the base sinusoid. Such auto-transformers are commonly used with an inductive-capacitive current-limiting impedance and are typically referred to as “lead-peak” ballasts.

This combination has another useful feature, namely, it can be designed in such a manner that lamp current is nearly independent of line voltage. Such “regulator” ballasts were widely used for high-pressure mercury lamps in the USA, and a similar ballast was used for metal-halide lamps. In addition to having higher starting-voltage requirements, metal-halide lamps may briefly exhibit during warm-up a high reignition voltage at the zero-current crossover. This was demonstrated by Franke

et al. at Sylvania to result from vaporization of HgI_2 along with mercury as the lamp warmed up (Franke et al. 1967). These workers also demonstrated that because of phase relationships in the ballast, the voltage available in the circuit to overcome the high reignition voltage requirement was, counterintuitively, the instantaneous voltage across the capacitor *minus* the instantaneous open-circuit voltage at the zero-current crossover. In an inductive ballast circuit, the phase angles are such that instantaneous open-circuit voltage provides the reignition voltage. Since current zero occurs near the maximum of open-circuit voltage, inductive ballasts have excellent reignition performance. The 220/240 choke ballast with electronic ignitor was widely used in Europe for metal-halide lamps for this reason. Modification was required in the design of “lead-peak” inductive-capacitive ballasts to achieve equal reignition capability, and many early installations had lamp-ballast incompatibility problems. It was an advantage that Sylvania at the time was in the ballast business, so that the problems could be solved cooperatively in-house, rather than in an arms-length, finger-pointing relationship with outside suppliers. The resulting modifications later formed the basis for ANSI and NEMA specifications.

Despite the reignition problems, the lead-peak circuit has remained the preferred one in the USA for MH lamps because it has a much lower cold-lamp current than the inductive ballast. The higher current during warm-up of the inductive ballast leads to higher electrode temperatures and higher tungsten evaporation and wall blackening and therefore poorer lumen maintenance. In addition, the inductive ballast permits unequal current flow on the two half cycles, a not uncommon phenomenon with aged lamps, of which the two cathodes are quite likely to have a different cathode fall. The DC component of current flow can increase the level of magnetic saturation of the core, especially in cheaper ballasts, exaggerating the current difference. This can also be a serious problem during ignition if the two cathodes do not convert from the “glow” phase to the “arc” phase simultaneously. Moreover, unequal lamp current in the two half cycles introduces a 50/60-hz flicker to the light output, which is annoyingly perceptible to many people.

High-pressure-sodium (HPS) lamps exhibited a steady increase in lamp operating voltage during life, because end-darkening increased the cold-spot temperature, and they were operated with excess Na-Hg amalgam. The ballast was required to tolerate the rise in lamp voltage without exceeding a specified maximum lamp power level. Only inductive ballasts had this characteristic, so the common circuit was a choke ballast and electronic ignitor in Europe and the same with a step-up transformer in the USA. A fuller discussion of magnetic ballasts and lamp-ballast interactions is given in Waymouth’s book (Waymouth 1971, Chap 12, pp. 307ff).

The latter half of the twentieth century saw semiconductor electronic devices initially make inroads into, and subsequently take over, much of the task of current control in electric lamps. The first such device was the so-called “phase-control” dimmer for incandescent lamps and tungsten-halogen (“TH”) lamps. This simply operates the lamp in series with a “TRIAC” AC electronic switch. This switch is normally off, but can be triggered into an “on” state by application of a voltage pulse to a trigger electrode. It remains on until the following current zero between half cycles at which it turns off. Dimming is accomplished by varying the phase of the

trigger pulse. Full current and light output are obtained by having a trigger pulse at the beginning of the half cycle; dimming is accomplished by delaying the phase further and further into the half cycle, reducing the rms lamp current. This is a very simple device, hardly more expensive than a standard off-on switch, and is widely used.

However, it has two serious faults as a circuit element. First of all, it has a low power factor in the dimmed mode, because the phase of the current waveform is delayed relative to that of the line-voltage waveform. Second, the line current has a high peak-to-rms ratio ("crest factor") because the peak current is whatever the lamp resistance would draw at the turn-on phase, while the rms current is reduced by the absence of current flow prior to turn-on. High-crest-factor line current has a high third-harmonic component. Power companies hate third-harmonic almost as much as low power factor; in a three-phase wye-connected circuit, third-harmonic currents add in the neutral line rather than cancel like the fundamental. This risks overloading the neutral line. Probably because lighting load is only one-quarter the total electric power load and dimmed-incandescent and dimmed-TH loads are only a fraction of that, I am not aware of any effort on the part of the utilities to ban the phase-control dimmer.

The energy crisis of the 1970s resulted in serious attention being given to solid-state electronic ballasts for discharge lamps as a means of improving efficiency. Not only are semiconductor power supplies more efficient than electromagnetic, efficiency of fluorescent lamps is higher for high-frequency currents. The essential components of an electronic ballast for use with standard AC power lines are a diode rectifier and/or voltage doubler, a filter to smooth the DC output of the rectifier, an oscillator to generate ca 25 khz AC, and an impedance to limit the lamp current. Auxiliary transformers generate filament heating voltages for starting and operating modes.

Two standard filter circuits are the choke + capacitor, or capacitor-only, possibly with a small series resistance to limit peak diode current. Since chokes are large and heavy, all practical ballasts use only a capacitor filter. There are numerous problems with this circuit. First, the rectifier diodes only conduct while instantaneous line voltage exceeds capacitor voltage, i.e., near line voltage peak. Hence, line current is highly peaked, with current flow only near line voltage maximum, with corresponding high third-harmonic component. Second, line power factor is low, because there is no line current for much of the cycle. Early electronic ballasts attempted to solve this problem by eliminating the filter capacitor and feeding the oscillator with unfiltered pulsating DC. However, the oscillator output voltage depends on the instantaneous value of the pulsating DC voltage and therefore does not become high enough to restrike the lamp or lamps until partway into the 60-hz half cycle, and the lamp goes out before the end of the 60-hz half cycle. Thus, lamp current only flows during a portion of each half cycle of pulsating DC. This results in the lamp current crest factor (peak-to-rms ratio) being significantly larger than the 1.7 specified by the lamp manufacturer, as great as 2.0 or more. The high crest factor leads to cathode fall greater than the sputtering threshold, and extremely short lamp life results. An early test installation of semiconductor ballasts sponsored by the US

Department of Energy employed such ballasts with double the lamp failure rate of the prior magnetic ballast system (Jewell et al. 1980).

Later versions of electronic ballasts necessarily incorporated capacitance filtering of the DC rectifier output to ripple ratios ca 10 % and were much more successful. Circuitry was incorporated in the ballast to ameliorate the power factor and third-harmonic problems, at considerable expense; nearly half the circuit elements in the ballast may be devoted to this requirement (Knoll 1978). Incorporating dimming into an electronic ballast adds only marginally to the cost; most electronic ballasts for fluorescent lamps today are dimming ballasts. The latest versions of electronic ballasts incorporate a slow ramp-up of oscillator output voltage together with a prompt application of filament heating voltage (Kenner and Moisan 1992). In consequence, the cathodes are already at thermionic-emitting temperature when the voltage becomes high enough to ionize the lamp, and there is no period of glow discharge. Fluorescent lamps operated on such ballasts experience almost no starting damage to the cathodes and achieve steady-burning life of 40,000 h and more on 3-h cycles.

In common with mercury lamps, metal-halide lamps experience acoustic resonance when operated at frequencies in the kilohertz range. The plasma temperature fluctuates with the varying high-frequency lamp current, leading to pressure fluctuations generating sound waves. When the sound wave frequency coincides with the resonance frequency of an acoustic mode of the arc tube, violent fluctuation of the arc column results. This may either cause the arc to extinguish when contact with the arc tube wall drastically cools the plasma or in extreme cases may cause the wall to shatter. This has been treated in two ways: first, to energize the arc at a few-hundred-hz square-wave current, with the switching interval short enough that there is no time for the plasma temperature to change (Keijser 2001), and, second, to frequency-modulate the high-frequency current, so that it does not dwell at a resonance frequency long enough for the amplitude of plasma oscillations to grow to destructive levels (Faehnrich Rasch 1988). The primary advantage of solid-state HID ballasts is lighter weight in comparison to electromagnetic ones; this can be extremely important for pole-mounted outdoor systems.

In the 1970s Fusion Systems Co. of Rockville, MD, USA, developed a family of high-pressure mercury and metal-halide lamps energized by microwave power generated by magnetrons (Spero et al. 1975). These lamps have been widely used to provide UV radiation for curing photopolymerizable inks in a variety of industrial applications. The central controlling factor is the skin depth of the plasma for the microwave radiation; despite the high electron density, the high electron collision frequency reduces the conductivity sufficiently to permit the microwave radiation to penetrate sufficiently to energize the high-pressure plasma. A review of this technology area has been given by Waymouth (1992). More recently, microwave-excited compact high-pressure metal-halide sources emitting primarily in the visible have been developed and test marketed for outdoor lighting (Gilliard 2011). Other recent discussions of this technology area may be found in the proceedings of the 13th International Symposium of Science and Technology of Light Sources (Posters CP021, CP922, CP023, CP024 2012).

Inductive coupling of electromagnetic energy at lower frequencies has also been used to energize low-pressure mercury argon discharges in a phosphor-coated envelope having the size and shape of an incandescent bulb. In common with the microwave-excited high-pressure lamps, these share the advantage of having no electrodes, thus eliminating a major source of lamp failure. They can be reasonably efficient, but the only commercial applications I am aware of are in hard-to-reach locations where extreme lamp life is desirable. The earliest reference to this technology was by Hollister, supported by the US Department of Energy (Hollister 1977). A review of the RF-excitation technology area has been given by Wharmby (1990).

Light-emitting diodes have from the beginning been powered by electronic circuitry that the LED community calls “drivers.” However, they are really nothing but electronic ballasts. LEDs are Schottky diodes, with forward current rapidly increasing with voltage above a threshold; a few tenths of a volt increase in voltage can double or triple the current flow. Moreover, among a lot of nominally identical diodes, threshold voltage of individual diodes may vary by a few tenths of a volt. Thus, like discharge lamps, they must be energized from a current-limiting source rather than a voltage source. A light source to be used for general lighting service must employ multiple diodes to provide the desired 1,000+ lumens. These would be commonly arranged in series, possibly with parallel strings of series-connected diodes. Each series string must be individually powered from a DC source through a current-limiting impedance (the “ballast”), which may be either a resistance or a chopper current regulator (Jacobs 2012). If the diodes in the string are randomly selected, the percentage variance of the total voltage of an N-fold string will be the percentage variance of the individual diode voltage divided by the square root of N. Therefore, for total string voltages of 60 V and higher, the resistor voltage drop needs to be only about 10 % of the total, and ballast efficiency can be high.

In common with an electronic ballast for discharge lamps, the LED driver must have a rectifier and capacitor-input filter (with the same power-factor and third-harmonic problems as the electronic ballast) preceded or followed by a voltage converter to provide the desired open-circuit voltage for powering the series-string-plus-resistor light engine. For chopper current regulators, feedback circuits may be provided to maintain constant light output or color and dimming capability.

The Electric Lamp Market Through the Years

Two words characterize the market for electric lamps prior to World War II: “monopoly” and “cartel.” It was a monopoly in the USA, owned by General Electric, and a cartel elsewhere. The key step in the USA was the formation of the General Electric Company in 1892 in a merger between the Edison General Electric Company and the Thomson-Houston Electric Company of Lynn, Massachusetts. The Edison General Electric Company was a consolidation of all the individual Edison companies and held the rights to the Edison patents, while the Thomson-Houston Electric Company had obtained the rights by acquisitions to most of the other

significant patents (Bright 1949 Fig 19, p. 85). The combined company then sued a number of other lamp manufacturers and effectively shut them down. This left Westinghouse, which had a number of potentially troublesome patents of its own, and some companies deemed “too small to bother with.” GE and Westinghouse cross-licensed each other, and the remaining small competitors were organized into a holding company, the National Electric Lamp Company, of which GE owned 75 % of the stock. As part of an antitrust consent decree in 1911, GE purchased the rest of it and folded it all into the GE lamp division, thereby obtaining an 80 % market share of the industry.

In Europe, there were a number of companies manufacturing incandescent lamps, no one of which had accumulated rights to a sufficient patent portfolio to accomplish what GE had. They instead formed a cartel, with all companies contributing to a patent pool and all acquiring licenses to the patents in the pool. The participants then divided up the continent of Europe, agreeing not to meddle in other participants’ “national markets.” Since those national markets included the colonies, this agreement effectively divided up the rest of the world.

Beginning in 1904, GE signed cross-licensing agreements with many foreign companies; the list, too numerous to elaborate here, essentially comprised all the important lamp producers of the world, including most of the members of the international cartel (Bright 1949, p. 308). These agreements included not only cross-licensing of patents and technical-information-exchange and mutual noncompete clauses but also investment and partial ownership. By 1940, GE owned 21.45 % of Osram, 11.85 % of Philips, 37 % of Compagnie des Lampes, 10.6 % of the Hungarian lamp company (which later became Tungsram), 10.86 % of AEI, and 28.14 % of Tokyo Shibaura (later Toshiba) (Bright 1949, footnote 11, p. 309). These ownership positions included seats on the boards of directors. Truly, the tentacles of this octopus extended around the globe. Many of these affiliations continued as technical-information-exchange and cross-licensing agreements well into the latter half of the twentieth century.

As a result of its affiliations with members of the cartel, GE gained early knowledge of the development of refractory-metal filaments, in particular tungsten. There were a number of different inventors and patent applications for such filaments in Europe and in the USA. In the USA, an interference was declared among a number of applications. GE bought the US rights to all of them for a grand total of \$760,000, in order to be sure to own whichever one came out on top. The eventual winner was the Just and Hanaman patent, which formed the basis of GE’s continuing monopoly position even after all the Edison patents had expired (Bright 1949, pp. 192–193, 255). To these of course, it added its own Coolidge patent on ductile tungsten, the Langmuir patent on gas-filled incandescent lamps, and the Pacz patent on non-sag tungsten wire (see section “[The Importance of Materials in Light-Source Development and Optimization](#)” for details).

Naturally, this monopoly attracted a suit by the US Department of Justice in 1924 which was finally settled by the US Supreme Court in 1926. The Court held that ownership of patents gave the holder the right to include restrictions of market share, price, or any other limitations in licensing agreements offered to competitors. This

decision formed the foundation for GE's establishment of two classes of licensee, A and B. Westinghouse was the only Class A licensee, with a maximum quota of domestic sales of 25 %, with cross-licensing of patents on both sides at a net royalty of 1 % to GE, essentially as a modification of the earlier cross-licensing agreement between them.

There were six class B licensees, of which Sylvania was the only one to become a significant factor in the post-World-War-II market. The name "Sylvania" is used herein to identify the lamp-making arm of a company that went through numerous changes of corporate identity: Hygrade Sylvania in the 1930s, Sylvania Electric Products during the 1940s and 1950s, GTE Sylvania after acquisition by General Telephone and Electronics, and Osram Sylvania Inc., after purchase from GTE by Osram. Its quota was 8 %, with a royalty rate of 3 %. Class B licensees received licenses under Westinghouse patents as well, as long as they remained class B licensees of GE, and were required to grant licenses to GE and Westinghouse under any patents they might hold. As noted earlier, class B licensees could purchase GE lamp-making machinery and lamp components from GE, with GE retaining an option to buyback if the terms of the license were violated. The licensees were not licensed for export nor for making their own bulbs or bases (Bright 1949, pp. 257–260).

The peace and quiet of GE's monopoly was rudely shattered by two events: the development of the fluorescent lamp by GE and World War II. Although as already noted GE had developed and announced with Westinghouse in April 1938 a commercial fluorescent lamp, with a "big-splash" demonstration at the New York World's Fair in 1939, it was rather diffident about marketing it. GE's utility customers (who bought electrical generating and distribution equipment) were outraged that their supplier was commercializing a light source that would produce the same amount of light with less electric power, thereby cutting their sales. Sylvania, however, had no such internal conflict. It had been aware of the work of Claude and others in Europe and had been exploring the technology area before the GE announcement. Moreover, it had a patent of its own, the Cox patent (1937), for a method of applying a coating of fluorescent phosphor to the inside of a tube, which it believed GE was infringing. Consequently, it reverse-engineered GE products and brought essentially identical lamps to market almost simultaneously.

World War II accelerated enormously the introduction of fluorescent lamps to the industrial-commercial market. Hundreds of new factories were being built in the USA to produce war materiel; all of them needed lighting, and electricity was in short supply. Consequently they were nearly all equipped with fluorescent lighting. Sylvania moved very aggressively to supply not only fluorescent lamps, but fixtures and ballasts; it could provide one source for the entire lamp-ballast-fixture package. In consequence, its share of the new US market quickly jumped to more than 20 %, far more than the 8 % it was limited to by its class B license for incandescent. In the last quarter of the twentieth century, it exited both the ballast and fixture businesses, officially on the grounds that the margins it could obtain were unsatisfactory.

Naturally GE sued and Sylvania countersued in May of 1940. GE alleged infringement of two patents, one by Albert Hull (1931) and one by Meyer et al. (1939). The Hull patent claimed a hot-cathode discharge device with a method

to hold the cathode fall below a critical value, and the Meyer patent claimed an elongated tube with at least one thermionic cathode and a luminescent material coated on the inside wall. The Meyer patent (based on an original German application) had been bought by GE when it was declared in interference with a GE application by Buttolph covering much the same technology. Sylvania alleged infringement by GE of the Cox patent as well as of patents by C. G. Smith and C. J. LeBel, to which it had bought the rights from Raytheon (Smith 1940; LeBel 1941). The US government entered the fray by filing a companion antitrust suit against GE and nine other defendants including Westinghouse and Philips in December 1942. All three trials were postponed at the request of the War Production Board until after the war in order not to distract the attention of company executives from the job of producing armaments for the war effort.

As a purely personal note, the only time I was ever prevented by the Sylvania legal department from presenting a paper or submitting a technical article on my work was in 1951, when a paper I proposed for the MIT Physical Electronics Conference would have reported that the peak cathode fall in a fluorescent lamp was indeed less than 16 V, contrary to the position Sylvania was arguing in court: that it did not infringe the Hull patent. The whole affair was finally settled with GE required to abandon its scheme of restrictive licenses, and all parties agreeing to license each other on reasonable terms. Sylvania, of course, paid royalties, but a number of years later finally achieved full technical equality in the 1970s with a royalty-free cross license with GE. I took some satisfaction that metal-halide-lamp patents of mine figured in this settlement.

The post-World-War-II marketplace could not have been more different. To begin with, all restrictions on international competition were removed as a result of the antitrust action. No longer were the cozy little “don’t mess around in my market, and I won’t mess around in yours” gentlemen’s agreements in effect. Sylvania jumped into the fray by acquiring a 49 % interest in Thorn Lighting of England with an option to buy the rest. It sold the interest back to Thorn in the 1970s but still retained technical-information-exchange and cross-licensing agreements until Thorn Lighting was bought by GE (together with Tungsram) in the 1990s. Thorn itself in the 1970s acquired the lighting business and facilities of AEI and later acquired a much bigger fish, the non-lighting company EMI.

Sylvania also established a European marketing division, “Sylvania International,” which originally sold products exported from the USA. In the 1960s, it established fluorescent manufacturing facilities with a horizontal exhaust machine in Erlangen, Germany, and a TH lamp manufacturing facility in Tienen, Belgium. In Japan, it established a joint venture with NEC with a fluorescent plant (again equipped with a horizontal exhaust machine) near Kyoto and acquired partial ownership of Kondo, a TH lamp manufacturer in northern Honshu. The horizontal exhaust machines made it a low-cost producer in both European and Asian markets, but it eventually sold its 49 % interest back to NEC in Japan. Needless to say, the other European producers were quite annoyed with this upstart underselling them in their own market.

In the early 1970s GE and Philips severed their technical-exchange agreements, and Philips and Sylvania signed one. At a somewhat later date, GE and AEG discontinued their agreements and as a result GE severed its contacts with Osram, which at that time was part of AEG. Osram in turn bought back the Osram brand name from GEC in England, which had expropriated it during World War I. There were discussions between Osram and Sylvania relative to a technical-information-exchange, but I am not aware of the signing of any formal agreement. With the unification of Germany, Osram reacquired the properties of the East German company Narva, which had been Osram's prior to World War II. Osram then purchased the Sylvania lighting business outright from GTE in the 1990s; this required spin-off of the Sylvania Lighting International group, which was purchased by "a consortium of individual investors" and continues as a major lighting manufacturer in Europe to this day.

Philips had moved the headquarters of the company to the USA during World War II to escape German control, creating the North American Philips Company in 1942; the headquarters was moved back to Eindhoven after the war, but North American Philips remained. This was primarily an electronics company until it bought the lighting division of Westinghouse in the 1980s. The Philips and Osram purchases leave GE as the only major lamp company in the USA that is still domestically owned. Its lamp division has been rumored to be on the market from time to time as well. *Sic transit gloria mundi!*

There were a number of lamp manufacturers in Japan during the latter half of the twentieth century: Matsushita, Mitsubishi, Toshiba, Hitachi, Japan Storage Battery, Kondo, Ushio, and Iwasaki, to name a few. Not all sold a full line of products, but there was considerable competition among them. Unlike the USA or Europe, in which the residential market was almost entirely incandescent, circular fluorescent in ceiling-mounted fixtures was popular in homes in Japan.

The post-World-War-II residential lighting markets in the USA and Europe remained almost entirely incandescent. Spiral and other CFLs were developed by a number of companies during the 1970s oil crisis, but the necessary ballasts were either too heavy or too costly to integrate with the lamp. (Philips did develop the magnetically ballasted SL lamp, but it found significant application only in commercial downlight installations.) Manufacturers of residential table-lamp and floor-lamp fixtures vehemently refused to equip their wares with ballasts: "The residential customers wouldn't buy them, and even if they did, they wouldn't pay any more for them, so we'd just be taking money out of our own pockets." In consequence, the 1970s oil crisis had little effect on the dominance of incandescent in the residential market.

In the 1990s electronic ballasts for CFLs became inexpensive and reliable enough to be integrally incorporated in lamps that could screw into incandescent sockets and fit inside the shades of portable lamps. However, these were retail priced at \$10 or more in the USA and did not sell into the residential market. Their long life and low power consumption did gain them entry into many hotel rooms. My local electric utility (a winter-night-peaking one) rented electronically ballasted CFLs to its

electricity customers for 20c a month, which I used in timer-controlled sockets (saving 70c worth of electricity per month from day one, without having to pay the up-front costs). The utility was buying increased peak capacity (from the demand reduction) at ca \$500/kw, much less than the \$1,000/kw it would have otherwise cost. I remain convinced that this, reverting to the *modus operandi* of the early electric lamp industry, would be a much more useful model for introducing LEDs to the residential market than outlawing incandescent lamps.

In the 2000s most manufacture of spiral CFLs was shifted to China, and a wide variety of integrally ballasted CFLs are now available in the USA at retail prices of \$2–\$3 and in Europe at equivalently low cost. Many of these carry the brands of the major manufacturers. The phosphors used are rare-earth types, and color and color-rendering are largely indistinguishable from incandescent. Market penetration has been much greater as a result. The US Congress has pointed to this availability to justify in part its banning the manufacture and sale of many wattages of incandescent lamps, seemingly oblivious to the fact that this has resulted in the closure of manufacturing plants and the loss of jobs during a period when unemployment is distressingly high. Other countries in Europe have enacted similar legislation.

There are certain unique characteristics of the current industrial-commercial electric lamp market that have had great influence on business strategy. To begin with, more than 80 % of lamp sales are for replacement purposes. Even though discharge lamps may have service lives of several years, the installations in which they are operated have service lives typically measured in quarter-centuries. Thus, over its life (determined as much by ballast life as any other factor) a typical industrial, commercial, or municipal-street-lighting installation will consume 8–10 bulbs in every fixture. Very few building owners will replace a lighting installation that is still working, except perhaps as a general renovation of the entire building. Owners of lighting installations are not like computer or cell-phone owners who will “upgrade” or even replace working systems as obsolete just to get the latest bells and whistles contrived by suppliers.

In consequence, the time scale for introduction of a new product (unless it will operate in an existing socket) can be decades. Even in the earliest days of the industry, it was more than 30 years before incandescent lighting really displaced gas lighting. More recently, in the USA, HPS lamps were introduced to the market in 1964 as replacements for high-pressure-mercury outdoor lighting. It was 1984 before unit sales of HPS lamps equaled those of mercury lamps. The same situation applied in Europe as well, even though in some Eastern Europe countries, replacement became a legislated requirement. An exception to this rule was the replacement of incandescent by TH lamps in new optical projector designs. Requiring only socket and optics redesign, this conversion doubled screen lumens, and customers did upgrade.

Numerous small companies financed by venture capital have over the years developed products which they believed were superior to anything currently available in the general lighting market. Their record of commercial success has been abysmal, because there is no market whose time constants are less appropriate to the needs of venture capital than the industrial-commercial lighting market. These forays seldom got beyond the test-installation stage before the venture capital ran out.

This long incubation time means that, unlike the semiconductor industry where second place is tantamount to failure, being a strong second has been a perfectly viable business strategy for introduction of new lamp products requiring new infrastructure. There was a saying in the USA (rooted partly in the psyche by national history) that “Pioneers get arrows; the second wave are the homesteaders, and they get the land.” Nevertheless, the introduction of totally new product families offers a golden opportunity to increase market share. Being a strong second in the fluorescent lamp industry gave Sylvania the opportunity to triple its market share. A similar opportunity arose in the introduction of metal-halide lamps, in which Sylvania’s strong second place allowed it to completely lap Westinghouse and significantly close the gap on GE. In Europe, Philips saw HPS lamps as a threat to its established low-pressure-sodium (LPS) business and aggressively pursued a leadership position in this market, while Osram focused on MH lamps.

A second unique characteristic of the industrial-commercial market is that the major lamp companies are not the prime movers of new installations. Fixture manufacturers, of which there are at least 50 in the USA (and if a visit to the Hannover Fair was any indication, in Europe as well), are the system integrators. They provide the housing, incorporating the ballast, socket, and the optics to control the light distribution, and most importantly the information about that distribution to the lighting designer (hired by the architect) needed to calculate the fixture layout to achieve a prescribed light level in a new installation at a building-code-required maximum energy consumption. Thus, the specifier buys from the fixture manufacturer, not from the lamp manufacturer. If the fixture is not shipped with a lamp in the socket, the lamps are drop-shipped to a construction site on order from the fixture manufacturer.

The lamps are sold to the fixture manufacturer at “original equipment manufacturer” (OEM) prices, little more than cost. It is important for a lamp manufacturer to have its product in the fixture, to take advantage of the fact that a building superintendent satisfied with the performance of the installation will most likely buy the same brand of lamps when replacements are needed. Therefore, lamp manufacturers will compete strongly for OEM business. Another party heavily involved in the replacement-lamp market is the “lighting maintenance contractor.” Because of the rather wide distribution of lamp lives (for 10,000 h rated life, first failures in an installation will likely occur before 5,000 h, and the last after 20,000 h), it is advantageous to replace all lamps (“group relamp”) before rated life, a service most economically provided for large installations by lighting maintenance contractors. Typically, these will clean the fixtures at yearly intervals (increasing foot-candle levels by 10–20 % depending on the dustiness of the facility) and replace all lamps at 2–4-year intervals, depending on rated lamp lives and operating cycle. As volume purchasers of lamps, these contractors will necessarily receive significant discounts.

It is a great temptation for lamp manufacturers to enter the fixture and maintenance contractor businesses to capture the value of these discounts in house. Those which have done so have generally found it to be a mistake. Both businesses are highly fragmented with no company holding more than a few percent share of the national markets, and the lamp manufacturers’ ventures have not typically

outperformed. Moreover, in fixtures especially, the large number of competitors selling nationally means that margins are small, much less than the lamp manufacturers are accustomed to enjoy. Finally, in establishing these businesses, lamp manufacturers are competing with their own customers, who generally have at least two other possible suppliers.

The low margins in the fixture business means that the companies in this field are eager for some way to distinguish themselves from competitors: design, performance, anything! At least a few fixture manufacturers will be avid promoters of a new light source, with which they can sell something their competitors don't have. Several of them were vigorous promoters of metal-halide lamps when they were first introduced and enthusiastically sold the new technology to lighting designers, architects, and specifiers. Their activity greatly accelerated the penetration of MH lamps into the market. The nearly 50 years that has elapsed since then without equally revolutionary new products from the lamp companies has left them most unhappy with the lack of new opportunity.

As a result, it is the fixture manufacturers who have been and will continue to be the drivers of introduction of LEDs to the general lighting market. The lamp companies have been most reluctant, citing all the disadvantages of LEDs in comparison to existing sources, but the advertisements in "Lighting Design and Application" (LD&A) and other trade magazines tell the story. Lamp companies are responding to the demands of their fixture customers for the new sources and are belatedly entering the market. However, they lack the automated facilities to manufacture the quantities of chips that will be required at the costs required and in most cases lack the expertise to construct such facilities in-house. They must, therefore, either acquire such capability as turnkey factories or outsource their production requirements to organizations that have the facilities, needing much lower investment (Guess how accountants would decide?). This is altogether a totally different commercial landscape from the one they have happily enjoyed for the last half-century. Why would fixture manufacturers buy from a lamp-company intermediary when they can buy direct from the original producer?

Should the LED revolution proceed to the projected conclusion, replacing incandescent, fluorescent, and HID lamps, there will be a further major change in the lighting market. The LED array required to reach a useful light level is an integral part of the fixture, which serves as the heat sink. So important is the heat sink that these arrays will probably not be designed like light bulbs for field replacement by untrained users. However, since the life of the chips is supposed to be ca 100,000 h, at 4,000 h per year of operation (when properly heat-sinked!), the array should last the 25-year life of the fixture and driver. There will be no replacement lamp sales, only replacement fixture sales. The role of the fixture manufacturer in such a market will be greatly amplified.

Lamp companies can still fight back. Inductively coupled fluorescent lamps also have 100,000-h lives and can be manufactured on presently idle incandescent lamp machinery with minor machine modifications, can be equipped with integral RF generators, and can fit in most incandescent sockets. Fluorescent lamps can be equipped with larger cathodes carrying more cathode coating and, when operated on mating electronic ballasts which delay application of igniting voltage until the

cathodes are hot, can also reach 100,000-h life. Thus, fluorescent lamps can reach with today's technology the life, efficacy, and color-rendering capabilities promised for LEDs at some future date, but at far lower cost. The major consequence is that exactly as with the LED revolution, there will be only one lamp sold over the expected life of the fixture and ballast, and the replacement lamp market will disappear. In either the LED scenario or long-lived-fluorescent scenario, the business model which has served the lamp companies so well over the last half-century will no longer be appropriate; if they wish to stay alive, they will have to enter the fixture business and drive the present participants out of business.

It will be interesting to see how this war plays out.

References

- Anderson S (1972) US Patent 3,676,534
Beese NC, Henry DE (1956) US Patent 2,765,416
Blount R (1976) US Patent 3,952,320
Bright AA (1949) The electric lamp industry: technological change and commercial development 1800 to 1947. Macmillan (Reprinted 1972 by Arno Press Inc.)
Brooks D, Kopelman B (1970) US Patent 3,517,182
Cayless M (1960) Resonance radiation in tubes of non-circular cross-section. *Br J Appl Phys* 11:492
Cayless M (1963) Theory of the positive column of Hg-Argon discharges. *Br J Appl Phys* 14:863
Coble RL (1962) US Patent 3,026,210
Coolidge WD (1913) US Patent 1,082,933
Cox JL (1937) US Patent 2,096,693
Danner E (1916–1917) US Patents 1,219,709 & 1,220, 201
de Groot JJ, van Vliet JAJM (1986) The high pressure sodium lamp. Kluwer Technical Books, Amsterdam
Dodge EH, Kimball LW (1955) US Patent 2,726,799
Elenbaas W (1951) The high pressure mercury vapour discharge. North-Holland, Amsterdam
Elenbaas W, van Boort HJJ, Spiessens R (1969) Improvements in low-pressure sodium lamps. *Illum Eng* 64:94
Faehrich H-J, Rasch E (1988) Electronic ballasts for M-H lamps. *JIES* 17:131
Fink W, Shaffer J (1971) US Patent 3,609,331
Franke A, Gungle WC et al (1967) *Illum Eng* LXII:204
Fridrich EG, Wylie E (1959) US Patent 2,883,571
Gardner R (1953) US Patent 2,960,285
Gardner R (1955) US Patent 2,960,285
Gendre MF (2003) Two centuries of electric light source innovation. <http://www.geocities.com/mfgendre>
Gilliard R (2011) Longitudinally-mounted light-emitting plasma in a dielectric resonator. *J Phys D Appl Phys* 44:224008
Govaert LE (1970) US Patent 3,493,017
Gungle WC, Waymouth JF, Homer HH (1967) Operating parameters of high-output fluorescent lamps. *Illum Eng* LII:262
Hollister DD (1977) US Patent 4,010,400
Hull AW (1931) US Patent 1,790,153
Jacobs R (2012) In: Light sources 2012. Devonshire R, Zissis G (eds) *Canopus Paper IL27* p 349
Jewell JE, Selkowitz S, Verderber R (1980) *LD&A* January:36
Kadow A (1925) US Patents 1,527,557, -558, -559 (Application 1910–1911)

- Keijser RAJ (2001) Electronic operation of HID lamps. In: Proceedings of the 9th International symposium on the science & technology of light sources. Cornell University Press, Ithaca, Paper 027:1, p 103
- Kenner JG, Moisan L (1992) US Patent 5,144,195
- Kenty C (1950) Production of 2537 Angstrom radiation and the role of metastable atoms in Ar-Hg discharges. *J Appl Phys* 21:1309
- Kenty C, Easley M, Barnes B (1951) Gas temperatures and elastic losses in Hg-Argon discharges. *J Appl Phys* 22:1006
- Knoll WL (1978) US Patent 4,109,307
- Koo RC (1967) Evidence for voids in annealed doped tungsten. *Trans Met Soc AIME* 239:1956
- Koury F, Waymouth JF (1968) US Patent 3,407,327
- Kuhl B, Krense H (1964) German Provisional Patent 1,184,008
- Langmuir I (1912) *Phys Rev* 34:40
- LeBel CJ (1941) US Reissue Patent 21,954
- Lemmers E (1956) US Patent 2,774,918
- McKeag AH, Ranby PW (1949) US Patent 2,388,733, (based on a British application dated 17 June 1942)
- McNichol T (2006) *AC/DC: the savage tale of the first standards war*. Wiley, New York
- Meggers WE, Corliss EH, Scribner BF (1961) Tables of spectral line intensities, part I: arranged by elements. NBS monograph 32, NIST, Rockville MD USA
- Meyer F, Spanner H, Germer E (1939) US Patent 2,162,732
- Neunhoeffler O, Schulz P (1950) US Patent 2,697,183
- Pacz A (1922) US Patent 1,410,499
- Posters CP021, CP922, CP023, CP024 (2012) In *Light sources 2012*. In: Devonshire R, Zissis G (eds) *Canopus*, pp 137–144
- Protzman JB, Houser KW (2006) *Leukos* 3:121, and references contained therein
- Quill LL (1950) The chemistry and metallurgy of miscellaneous materials. McGraw Hill, New York
- Reiling G (1966) US Patent 3,234,421
- Sanchez-Vello L (1935) US Patents 2,009,326 & -793
- Schmidt K (1966) US Patent 3,248,590 and (1968) US Patent 3,384,798
- Smith CG (1940) US Patent 2,201,817
- Spero DM, Eastlund BJ, Ury MG (1975) US Patent 3,872,349
- Steinmetz CP (1911) US Patent 1,006,021
- Stull DR et al (1971) JANAF thermochemical tables, 2nd ed NSRDS-NBS, vol 37 and supplements to June 1974
- van Arkel AG, de Boer J (1925) *Z Anorg Chem* 148:345
- Vrenken LE (1978) Fluorescent lamps with very high efficiency. *Light Res Tech* 10:181
- Waymouth JF (1971) Electric discharge lamps. MIT Press, Cambridge, MA, p 266 ff. and references contained therein
- Waymouth JF (1992) Applications of microwave discharges to high-power light sources. In: Ferreira CM, Moisan M (eds) *Microwave discharges, fundamentals and applications*. Plenum Press, New York and London, pp 427ff
- Waymouth JF, Bitter F (1956) Analysis of the plasma of fluorescent lamps. *J Appl Phys* 27:122
- Waymouth JF, Bitter F, Lowry EF (1957) Factors to be considered in the design of high-output fluorescent lamps. *Illum Eng LII*:257
- Weber K (1976) US Patent 3,937,946
- Wharmby DO (1993) Electrodeless lamps for lighting: a review. *IEE Proc A* 140:465
- Zissis G, Kitsinellis S (2009) State of art on science and technology of electric light sources: from the past to the future. *J Phys D Appl Phys* 42:173001
- Zubler EG, Mosby FA (1959) An incandescent lamp with virtually 100% Lumen maintenance. *Illum Eng* 54:734

History of Solid-State Light Sources

Oleg Shchekin and M. George Craford

Contents

Introduction	42
Early Pioneers	43
Evolution of Visible III-V LED Technology	46
GaN/AlInGaN Materials and Devices	49
Continual Improvement in LED Performance	52
Laser-Based Solid-State Sources	60
LED Packaging	60
Conclusion	65
References	66

Abstract

The second decade of the twenty-first century has become the period when light-emitting diodes (LEDs) are beginning to reach their promise of being the dominant technology for generating light. For those who have worked in this field, it is gratifying to see the enthusiasm of consumers for the technology and how the rate of adoption often outpaces expert projections. At the time of writing, leading manufacturers of lighting products achieve more than 40 % of their revenue by sales of products that use LEDs as the light source, and this portion is projected to be over 80 % by 2020. The rapid adoption of LEDs is largely due to the tremendous improvement in underlining technology which enabled cost reductions and increase in functionality of lighting systems. As novel as LEDs may

O. Shchekin (✉) • M.G. Craford
Lumileds, San Jose, CA, USA
e-mail: oleg.shchekin@lumileds.com

seem, the related discoveries and technology development have over a century-long history, with generations of researchers working on fundamental exploration without which the present day successes would not have been possible. At the same time, the present state of the art of LED technology and manufacturing lacks the homogeneity of the more established fields such as conventional light sources or silicon integrated circuits. This chapter will review the pioneering work, with more focus given to the developments of the most recent decades. An outlook to the directions that the field may take will be provided as well.

Introduction

The second decade of the twenty-first century has become the period when light-emitting diodes (LEDs) are beginning to reach their promise of being the dominant technology for generating light. For those who have worked in this field, it is gratifying to see the enthusiasm of consumers for the technology and how the rate of adoption often outpaces expert projections. At the time of writing, leading manufacturers of lighting products achieve more than 40 % of their revenue by sales of products that use LEDs as the light source (Philips Results Q3 2014), and this portion is projected to be over 80 % by 2020. The rapid adoption of LEDs is largely due to the tremendous improvement in underlining technology which enabled cost reductions and an increase in functionality of lighting systems. The technological advances have produced commercial components with luminous efficacy over 200 lm/W. Such high efficiency allows not only for energy savings but also for the reduction in system cost as fewer LEDs and less heat sinking are needed to produce a desired amount of light. The growing production volumes and the beginning of standardization of technology and manufacturing methods are reducing the cost of LED components, further lowering barriers for adoption of solid-state lighting. Besides the energy savings, the basic flexibility of LED technology invites new possibilities such as choice in light color and the natural integration with electronic components enabling control systems such as occupancy detection and communication via modulated light. The long life and high brightness offer design flexibility as well as added safety for automotive applications, while the microelectronic nature of LEDs is even utilized for data transfer and communication ([Philips creates shopping assistant with LEDs and smart phone](#); Elgala et al. 2011).

Such growing adoption and awareness of the LED technology by the society motivate reviews of the history of innovation in the field, which is the focus of this chapter. As novel as LEDs may seem, the related discoveries and technology development have over a century-long history, with generations of researchers, often working in challenging conditions on fundamental exploration without which the present-day successes would not have been possible. At the same time, the present state of the art of LED technology and manufacturing lacks the homogeneity of the more established fields such as conventional light sources or silicon-integrated

circuits. Therefore, we are writing this manuscript sometime half to three quarters of the way in LED evolution from a laboratory novelty to a common component, depending on application. This chapter will review the pioneering work in the field, where the reader will be referenced to some detailed accounts written previously (often by contemporaries of a period). More focus will be given to the developments of the most recent decades. An outlook to the directions that the field may take will be provided as well.

Early Pioneers

The first recorded observation of electroluminescence was made in 1907 by Henry Round. The early twentieth century was the time of rapid development in radio, and Round and others were experimenting with solid-state rectifying detectors (crystal detectors). He noticed that applying a voltage between two point contacts on the surface of carborundum (SiC) resulted in the crystal giving off a yellowish light. Round recorded his observations in a note to the editors of *Electrical World* (Round 1907), in which he inquired about the existence to references of similar observations and also theorized about the origins of the light. Round experimentally identified the dependence of the emission on the bias voltage and asserted that the emission is not connected with heating. Now we know that Round formed a Schottky contact and the light was generated by injecting carriers across the junction at bias.

The research on electroluminescence of SiC was independently restarted by Oleg Losev in 1922. A very informative account and analysis of Losev's work are given by Loebner (1976). Similar to Round, Losev observed luminescence from a point metal contact against the surface of SiC, but unlike Round, he continued the study of the phenomenon and published four patents and 16 papers related to the topic. Losev made a few very important observations. He showed that the wavelength of light emitted by the diodes was, effectively, the applied voltage minus the resistive voltage drop and postulated that the emission was an inverse of the photoelectric effect. He also understood that the mechanisms for light emission in forward and reverse bias were distinct and showed that the emission of light was coming from a thin layer beneath the surface of the crystal. He characterized the active part of the crystal to consist of the three distinct conductive layers, which, in retrospect, was an observation of a p-n junction, except he described all of the layers as n-type stopping short of discovering hole conduction in silicon carbide.

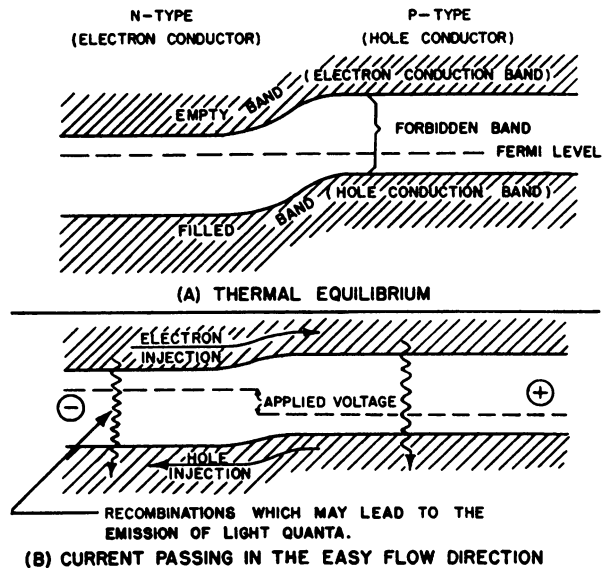
Besides the academic investigations, Losev envisioned practical uses for his discovery. He filed a number of patents outlining methods for the use and application for the light-emitting SiC crystal detectors. One particular patent for a light relay, filed in 1927, is truly visionary (Losev 1929). Here Losev teaches the use of the crystal detector as a high-speed modulated light source to record information by exposing a moving photographic plate to the emitted light. The detector itself is connected to a modulating circuit to which the signal is transmitted with or without a wire, and the fast modulating property of the light-emitting crystal detector is

leveraged for high-speed information transfer and recording. The invention is many decades ahead of the age of telecommunication and optical storage.

It is worth mentioning that Losev did not have a formal university education and was a 19-year-old technician in Nizhegorodskaya Radio Laboratory at the time of his discovery of luminescence in SiC. He continued to work as a technician during his carrier in various Soviet radio laboratories but was given a “Candidate of Science” degree (Soviet equivalent of a PhD) by the Ioffe Institute in 1938 without a proper dissertation. He passed away from hunger in 1942 during the WWII blockade of Leningrad at the age of 33. At the time, Losev was busy working on a three-contact semiconductor system to replace vacuum tubes (Zheludev 2007). He intended to publish a manuscript “on an important silicon device,” but due to the blockade, he could not send the document out to the evacuated office of the journal Soviet Physics. If the manuscript is ever found, we may see if Losev was on the path to inventing the transistor.

The modern explanation of the electroluminescence in SiC was formulated in 1951 by Kurt Lehovec, almost 20 years after Losev’s initial observations. It is important to note, though, that the understanding of light-emitting diodes would not be possible without the discovery and the description of transistor in 1947 by John Bardeen and Walter Brattain (Brattain et al. 1948). The transistor validated the existence and injection of holes and that the injection of current into a semiconductor creates non-equilibrium in electron hole populations, with carriers of each type being injected over the energy barrier (minority carrier injection). Lehovec et al. (1951) revisited Losev’s SiC luminescence work and postulated that the process of light emission consists of carrier injection of over a p-n barrier, with a direct reference to the transistor effect, while light emission results from the recombination of these carriers across the bandgap (see Fig. 1). A possibility of non-radiative recombination

Fig. 1 Reproduction of Fig. 8 from reference (Lehovec et al. 1951) illustrating the explanation for light emission in silicon carbide as a result of carrier injection over the p-n barrier followed by radiative or non-radiative recombination of the carriers across the bandgap (Copyright (1955) by The American Physical Society)



was also stated where the energy would be released as heat. This description of the mechanism of light emission in a diode has served as a guiding principle in the development of LEDs.

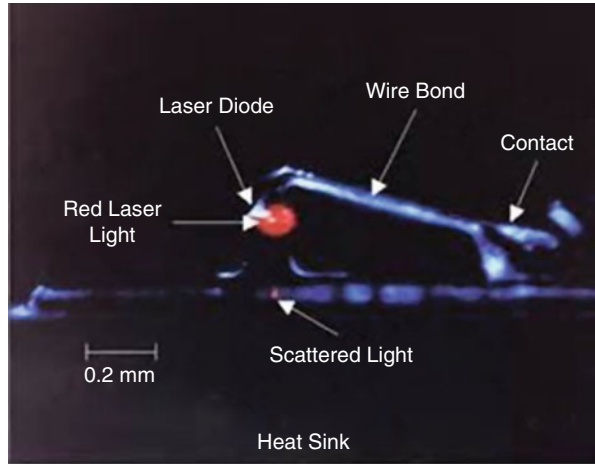
The decade of the 1950s was dominated by research on electroluminescent ZnS for possible use in flat television panels. Even though luminescent II-VI materials only found use in niche applications, the field did keep enough bright people employed for important progress to be made toward the present-day mainstream LED technology. A number of efforts particularly stand out. One is the work by Michael Schon in 1953, where Schon revisited electroluminescence of SiC and separated the efficiency of the process into injection efficiency of minority carriers and, separately, efficiency of recombination at and near the p-n junction. This distinction structured the way LED and laser designers go about optimizing the electrical efficiency of emitters. Also significant is the demonstration of electroluminescence in GaP by Wolff et al. in 1955. This marked the beginning of the work on III-IV semiconductor-based LEDs and motivated further investigations of GaP luminescence in a number of laboratories (Loebner 1976).

The 1960s is the decade of the first practical implementations of LEDs using materials and principles which are part of the basis of LED technology in use today. A number of demonstrations of GaAs-based infrared LEDs were made in 1962 (Biard and Pittman 1966; Hall et al. 1962; Pankove and Berkeyheiser 1962; Pankove and Massoulié 1962; Quist et al. 1962). Biard and Pittman, at Texas Instruments, filed a patent on August 8, 1962 in which they detailed the structure of a GaAs p-n junction LED (Biard and Pittman 1966). Biard and Pittman noticed infrared emission in 1961 when working on tunnel and varactor diodes based on GaAs. This led them to think about how to design the semiconductor and diode contact for efficient extraction of light, which was described in the 1962 patent. The same year Texas Instruments released the first IR LED product (TI-SNX-100) which found its first use in IBM punch card readers.

GaAs is a direct-bandgap semiconductor which, unlike earlier indirect-bandgap SiC and GaP LEDs, has potential to be a high quantum efficiency emitter. The missing link from the standpoint of lighting was the demonstration of a direct-bandgap visible emitter. The first direct-bandgap visible (red) LEDs were reported in 1962 by Holonyak and Bevacqua (1962) at General Electric using GaAsP, an alloy of GaAs and GaP. They also demonstrated a red current-injected laser at 77 K. GaAsP had been studied by Ehrenreich at General Electric (Ehrenreich 1960) who documented the bandgap behavior versus the As/P fraction. However, experts in the field had been doubtful that crystal quality suitable for electronic devices could be achieved. Holonyak and Bevacqua demonstrated that by using closed-tube vapor transport, they could achieve suitable crystals. Holonyak had the vision to recognize that this work could lead to a practical light source although “much more experimental work must be done” (Readers Digest 1963). All of the high-performance LEDs and injection lasers utilized today follow from Holonyak’s work and are based on semiconductor alloys (Fig. 2).

After this work was demonstrated, Holonyak was visited by R.A. Ruehrwein of Monsanto Company. Ruehrwein recognized that the closed-tube growth technique

Fig. 2 Image of the first red-emitting GaAsP alloy laser demonstrated by Prof. Nick Holonyak, Jr., at General Electric in 1962 (Image courtesy of Nick Holonyak)



used by Holonyak could be replaced with an open-tube chemical vapor deposition (CVD) process that would be much more scalable for manufacturing. In the following years, Monsanto implemented the large-scale epitaxial growth of GaAsP and in the 1970s became the world's largest supplier of compound semiconductor materials, particularly GaAsP for applications such as calculators and watch displays, prior to the development of LCDs.

Evolution of Visible III-V LED Technology

The indirect GaP and phosphorus rich GaAsP semiconductors were used also to generate light of other colors in addition to red. In 1965, Thomas et al. (1965) demonstrated that if GaP is doped with an optically active isoelectronic impurity, the spatial localization of carriers at the impurity spreads the wavefunction in momentum space, which increases the probability of an optical transition with one of the bands. Thomas and his team at ATT Labs used nitrogen-doped GaP to generate green emission and in 1967 Logan et al. demonstrated red-emitting GaP using Zn and O pair emission (Logan et al. 1967). In the early 1970s, the team at Monsanto used nitrogen doping in GaAsP to achieve orange and yellow emitters as well as $10\times$ brighter red than those based on direct GaAsP (Groves et al. 1971; Craford et al. 1972). GaAsP:N had higher performance than GaAsP primarily due to improved light extraction from the chip. The emission from the nitrogen sites was below the energy gap, and the indirect semiconductor had less absorption than direct GaAsP. The GaAsP:N was also grown on a “transparent” GaP substrate instead of GaAs. The CVD grown nitrogen-doped GaAsP:N and nitrogen- and ZnO-doped GaP, grown using LPE, were the main materials systems for visible LEDs through the early 1980s (Fig. 3).

In the early 1980s, AlGaAs heterojunctions became the highest performance red LEDs. AlGaAs had been studied using liquid-phase epitaxy (LPE) growth since the



Fig. 3 Examples of some of the first products using LEDs. The Hamilton Pulsar wristwatch was announced in 1970 and used GaAsP LEDs in its display. Hewlett-Packard and Texas instruments offered a series of programmable calculators GaAsP with LED displays. LED use was short lived for both of these applications as the LEDs consumed quite a bit of power and were not bright enough in direct sunlight. Liquid crystal displays eventually replaced LEDs in calculators and wristwatches

1960s (Alferov et al. 1969; Rupprecht et al. 1967). It was seen to have advantages over GaAsP by offering an opportunity to combine direct-bandgap heterostructures in a lattice-matched system. This combination was expected to result in devices with high internal efficiency due to the lattice-matched system and the confinement of carriers by heterostructures and high extraction efficiency since the confining layers would be of higher bandgap and would not absorb light from the active region. Unfortunately, the high-volume CVD could not be used because the aluminum attacked the hot quartz walls of the growth chamber. The main challenge with state-of-the-art LPE was creating high-quality crystals in high volume. The temperature-difference LPE technique developed by Nishizawa in Nishizawa et al. (1977) was used by Stanley Electric to introduce and manufacture high-performance red LEDs. The combination of excellent crystal quality, efficient direct-bandgap recombination, and good light extraction due to the heterostructure yielded devices with a 10× improvement in performance. Stanley Electric demonstrated luminous efficacies up to 10 lm/W, and other manufacturers soon followed.

The 10 lm/W barrier may not be impressive today, but it was an important milestone in the evolution of LEDs as this efficacy, along with the low voltage and size, made LEDs attractive for use in a variety of practical outdoor applications such as traffic lights, displays, and car taillights. However, AlGaAs materials, proved to

have severe reliability issues because the aluminum, at concentrations necessary for establishing a useful heterostructure, became readily oxidized, and LEDs failed.

The next improvement in visible LED technology came with the quaternary alloy AlInGaP. $(\text{Al}_x\text{Ga}_{1-x})_{0.5}\text{In}_{0.5}\text{P}$ heterostructures can be lattice matched to GaAs and made to emit from red to green. At Al compositions near 53 %, the semiconductor becomes indirect; therefore, the emission near green is inefficient. However, for longer wavelengths, the direct bandgap and confinement of the active region by the heterostructure promised high radiative recombination and photon extraction efficiencies. Additionally, the lower aluminum content was expected to alleviate the reliability issues associated with oxidation in AlGaAs LEDs.

In the mid-1980s, the AlInGaP alloy was being studied for use in semiconductor lasers (Kobayashi et al. 1985; Ikeda et al. 1986; Ohba et al. 1986; Itaya et al. 1990) but was not being pursued for LEDs. LPE and VPE were the dominant high-volume, low-cost crystal growth technologies in LED production with epi cost of around \$10 per square inch (Craford 2013). For thermodynamic reasons, LPE was found to be unsuitable for high-volume, high-yielding growth of AlInGaP, and as in the case of AlGaAs, CVD had an issue with aluminum attacking the quartz walls of the chamber. The technique which was being used to grow AlInGaP was metal-organic chemical vapor deposition, or MOCVD. MOCVD was pioneered by Manasevit at Rockwell (Manasevit 1968). Unlike in CVD, where the walls of the reactor are kept hot, MOCVD heats the growth substrate to facilitate cracking of metal-organic precursors and growth of epitaxial films. In 1977, R. D. Dupuis and P. D. Dapkus demonstrated room-temperature operation of AlGaAs lasers grown by MOCVD with the performance matching the best of those grown by LPE. The results were obtained by a greatly improved MOCVD process and showed that MOCVD could be used for the generation of high-performance semiconductor devices. In the mid-1980s, MOCVD was still a relatively expensive low-capacity technique thought to be useful for lasers but not commercial LEDs. Hewlett-Packard started a program developing AlInGaP LEDs believing that there were no fundamental barriers to developing MOCVD into a high-volume manufacturing technology. In 1990, the HP team announced AlInGaP LEDs operating at $10\times$ that of existing yellow LEDs and equivalent to that of best AlGaAs devices (Kuo et al. 1990). The first devices had the absorbing GaAs substrate but had a thick, transparent, and lattice mismatched GaP layer on top to give current spreading and improve light extraction.

The AlInGaP-based LEDs are the dominant technology for red/orange direct color emitters today. The epitaxial growth is based on AlInGaP lattice matched to GaAs, which creates challenges for maximizing emitter efficiency. The emission wavelength of the heterostructure is varied by adjusting the aluminum-gallium ratio and is limited to aluminum concentrations below 53 %, at which the bandgap becomes indirect. Additionally, the confinement of electrons worsens with higher aluminum content, resulting in strong temperature sensitivity of the internal efficiency. Therefore, for practical applications, the AlInGaP material system is restricted to wavelengths 580 nm and longer. In addition, in the early AlInGaP devices, the extraction of the photons was hindered by the absorbing substrate and the relatively high index of refraction of the AlInGaP semiconductor. A series of

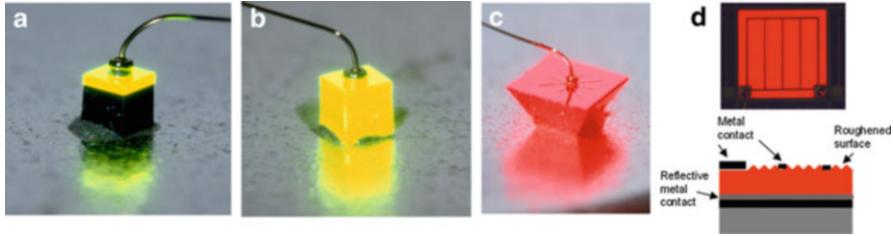


Fig. 4 Evolution of high extraction efficiency AlInGaP emitters: (a) thick GaP window layer with absorbing GaAs substrate; (b) diode with thick window layer and transparent substrate; (c) TIP chip architecture; (d) thin-film architecture

advances in the optical efficiency was made by teams from Hewlett-Packard. In the 1992 demonstration by Huang et al. (1992), the extraction efficiency of the LEDs was improved by $2\times$ by utilizing an even thicker GaP window, allowing for improved current spreading and extraction of photons before they are absorbed by GaAs substrate. In 1999, Kish et al. eliminated the absorbing substrate all together by etching it away and bonding AlInGaP to the GaP transparent substrate, raising the extraction efficiency to an estimated 20–24 % (Kish et al. 1994). In 1999, Krames et al. have further improved the extraction efficiency of the die, by introducing tapered inverted pyramid (TIP) chip architecture (Krames et al. 1999). The taper frustrated the TIR rays inside the LED and resulted in further 40 % improvement in efficiency with an estimated photon extraction efficiency of 60 %. The TIP chip was the first LED to achieve 100 lm/W.

The TIP chip and transparent substrate architectures have high extraction efficiencies and by design emit photons over a large area and into a wide solid angle. This makes these architectures not ideal for applications where secondary optics in a lighting system are used to collect the light from the LED and shape or focus the light with high flexibility. For such purposes, a two-dimensional surface emitter is much more optimal. The need to address such applications has resulted in the introduction of thin-film architectures by a number of manufacturers. Figure 4d schematically shows a thin-film architecture for a red AlInGaP die. Here an AlInGaP epi structure has a metal contact deposited and bonded to a carrier substrate. The carrier substrate is often GaAs, Ge, or Si. After bonding, the growth substrate is removed, and the exposed AlInGaP surface is roughened for enhanced photon extraction. The thin, high-index film creates challenges for extraction of photons, and significant effort has to be put into die and epi-layer design, but high extraction efficiencies have been demonstrated (Streubel et al. 2002; Broell et al. 2014).

GaN/AlInGaN Materials and Devices

In 1968, the Radio Corporation of America, the leading television manufacturer, launched an internal program to develop blue LEDs. At that time, the technology for red- and green-emitting LEDs had been demonstrated and was becoming available.

Developing a blue LED would allow for creation of a flat color television. Herbert Paul Maruska was a scientist at RCA who was asked by James Tietjen, the director of the company's central labs at the time, to look into gallium nitride as the basis material for blue LEDs. There are a number of reviews of the development by Maruska himself as well as by others (Schubert 2006; Maruska). By then, GaN had been prepared as powder, its crystal structure was documented, and a variety of dopants were experimented with. No one had tried the VPE growth of GaN, so Maruska, who had a reactor for growth of GaAsP, replaced the arsine bottle with ammonia. He chose sapphire as the substrate because it was available and would not react with ammonia. In 1968, after some failed attempts, Maruska succeeded in growing the first monocrystalline film of GaN by turning up the growth temperature to 850 C. The GaN films were strongly n-type without intentional doping, but the initial attempts for p-doping proved unsuccessful. Maruska and Tietjen published the work in 1969 (Maruska and Tietjen 1969). In 1971 Jacques Pankove and Ed Miller, Maruska's colleagues at the RCA labs, demonstrated a GaN-based LED using a metal-insulator-semiconductor (MIS) diode structure where the insulating layer was created by doping GaN with zinc following earlier work by Maruska. These devices emitted green light and were the first GaN-based LEDs (Pankove et al. 1971). At that time, Maruska was enrolled in a PhD program at Stanford, where he continued his work on the GaN-based LEDs, sponsored by RCA. There he experimented with magnesium doping of GaN, which he believed would be a better p-dopant to achieve blue emission. In July 1972 he successfully demonstrated the first blue/violet current-injected GaN-based LEDs emitting at 430 nm (Maruska et al. 1972). The efficiency of the emitters was very low since these LEDs were effectively metal-insulator-semiconductor diodes as the Mg-doped GaN did not show p-type conductivity. The Stanford-RCA team published a number of studies where they attributed the luminescence to impact ionization of the magnesium dopant by electrons injected by tunneling through the insulating regions (Maruska and Stevenson 1974; Pankove and Lampert 1974). Maruska returned to RCA the following year. Unfortunately, due to budget concerns at the time, the blue LED project was canceled. The LEDs created by the RCA effort were not yet ready for commercialization, and more discoveries were needed to make these LEDs practical, but the technology of GaN on sapphire and magnesium doping is the foundation of all of the GaN-based LEDs today (Fig. 5).

The work on GaN-based LEDs slowed considerably in the second half of the 1970s and the early 1980s. However, a few groups of researchers continued to focus on two main issues limiting the LED performance: crystal quality and p-doping. In 1983, Yoshida et al. reported on the effectiveness of using AlN buffer for the growth of GaN films on sapphire using reactive MBE (Yoshida et al. 1983). In 1986, Amano et al. (1986) demonstrated and quantified the improvement in the quality of MOVPE GaN films grown on sapphire by means of a thin AlN buffer layer. In the initial demonstration, followed later by a detailed study (Hiramatsu et al. 1991), the team showed that a thin AlN layer, deposited at relatively low temperatures, crystallizes, upon heating of the growth substrate, into columnar structures which serve as nucleation sites for the growth of GaN along the c-axis. As GaN is deposited, the

Fig. 5 The first blue LED based on Si- and Mg-doped GaN on sapphire substrate developed by Maruska (Photo: Herb Maruska)



multitude of growth islands coalesce, eventually forming a high-quality monocrystalline film. The quality of the crystal was reflected in the much narrower width of X-ray rocking curve peaks and lower background electron concentration. Later, high-quality growth of GaN was also achieved with low-temperature GaN buffer layers by Nakamura (1991) using MOCVD and by Moustakas using molecular beam epitaxy (MBE) (Lei et al. 1991).

The breakthrough on p-doping of GaN was announced in 1989 by Amano et al. who discovered that the conductivity of GaN doped with Zn or Mg dramatically increased after the material was irradiated with a beam of low-energy electrons. This demonstration was an important step toward GaN-based p-n junction LEDs which were also reported in the same paper. The mechanism responsible for the improved conductivity was explained by Nakamura et al. (1992) who later showed that a similar effect can be achieved by thermal annealing of Mg- or Zn-doped GaN. The low-energy electrons or thermal treatment break the Mg-H complexes which form during crystal growth in which the dopant is passivated by hydrogen. Annealing is a commonplace, large-volume manufacturing technique, and the demonstration by Nakamura was an important step toward industrialization of GaN p-n junction-based devices.

In 1993, Nakamura, who was employed as an engineer at Nichia Corporation, demonstrated, for the first time, high-quality InGaN heterostructures as well as GaN/InGaN double-heterostructure LEDs grown on sapphire with MOCVD (Nakamura et al. 1993). The blue-emitting LEDs had the external quantum efficiency of 0.22 %. A year later, Nakamura and his team followed up with an InGaN/AlGaN heterostructure LED, which now showed an EQE of 2.7 %. This demonstration of the broad range of nitride alloys allowed for subsequent engineering of GaN-based LEDs to much higher efficiencies. The use of MOCVD, sapphire substrates, and the thermal activation of magnesium enabled the leveraging of the existing manufacturing tools for the rapid development and proliferation of GaN-based blue, UV, and green LEDs. In 2014, Isamu Akasaki, Hiroshi Amano, and Shuji Nakamura were awarded Nobel Prize in Physics for their numerous key contributions to the development of high-efficiency blue LEDs.

The significance of the development of high-efficiency blue LEDs, and the motivation behind the recognition by the Nobel Prize committee, is due to their direct application to high-efficiency white-light sources. In 1970, Bell Labs was granted a US patent 3,691,482 (Pinnow and Gerard Van Uitert 1970), where the inventors described the use of YAG and various other down-converting materials to generate white as well as colored light. Generation of white light, in particular, was taught as a combination of light from a phosphor excited by blue laser and a portion of the laser light which was not converted. With the demonstration of efficient blue LEDs, the ideas for generation of white as well as different-colored LEDs using down conversion followed (Shimizu et al. 1999; Butterworth and Helbing 1998; Höhn et al. 2001; Baretz and Tischler 2003).

Continual Improvement in LED Performance

The introduction of white emitters based on blue LEDs started the proliferation of solid-state lighting into various applications which influenced LED designs and motivated improvements in efficiency. To better illustrate the development of LED sources through the 1990s, it is worth introducing the main elements comprising the luminous efficacy of LEDs. The LED luminous efficacy, η_L , can be defined as the product of the following efficiencies: internal quantum efficiency (IQE) of carrier recombination in the active region, extraction efficiency (EXE) of photons from the blue LED, electrical efficiency (ELE) of injecting carriers into the active region, and efficiency of converting (CE) blue photon into the desired LED spectrum (Krames et al. 2007):

$$\eta_L = \text{IQE} \times \text{EXE} \times \text{ELE} \times \text{CE}$$

The conversion efficiency (CE) can then be broken down further into “package efficiency” (PE), Stokes shift penalty (QD), luminous equivalent of emission spectrum (LE), and quantum efficiency (QE) of down conversion by phosphors. Figure 6 illustrates the main interactions between the building blocks of an LED (blue die, package, epi, and converter) and how these affect luminous efficacy.

The evolution of the GaN-based chip has been driven by applications and a general need for greater LED efficiency. Figure 7a shows the schematic of a blue LED of the type used in the original demonstration of the blue double-heterostructure LED by Nakamura (Nakamura et al. 1994). The main feature of this structure is the semitransparent Ni/Au contact which is used to facilitate uniform injection of current on the p-side. While helping to spread the current, the contact absorbs photons and limits the extraction efficiency of the structure. In the late 1990s and early 2000s, two directions in LED architectures emerged. One focused on low cost and low power per emitter for use in indicators and mobile displays. The other focused on high-power applications which valued the number of lumens per LED such as traffic and automotive signaling and, later, automotive forward lighting, illumination, projection, and camera flash. The chip in Fig. 7a was well suited for

LED building blocks	IQE	ELE	EXE	CE			
				PE	LE	QD	QE
epi (substrate, emission wavelength, heterostructure)							
die (type, p-, n- contacts, interconnects)							
package (type, encapsulants, interconnects)							
converter (phosphors, encapsulants)							

Fig. 6 LED building blocks and their impact on the efficacy. The *blue* cells indicate a component of luminous efficiency affected by a particular LED building block

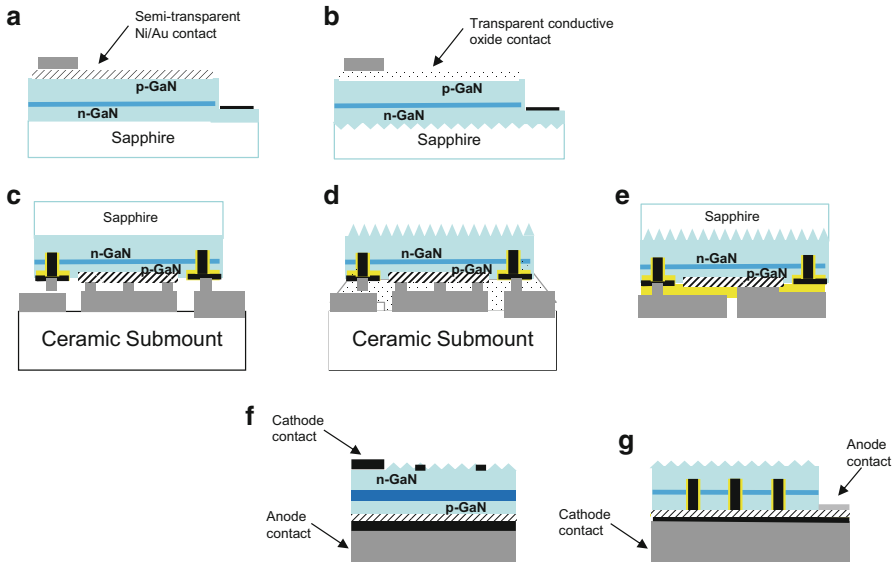


Fig. 7 Evolution of the GaN blue die and proliferation of die types: (a) conventional chip; (b) lateral die (*LD*); (c) flip chip; (d) thin-film flip chip (*TFFC*); (e) chip-scale package (*CSP*); (f) vertical thin film (*VTF*); (g) embedded-contact VTF (*EC-VTF*)

low-power applications. To increase the extraction efficiency of the die, certain changes were adopted and proliferated through the industry during the past decade. Figure 7b shows an example of a conventional architecture, which is similar to that in Fig. 7a, except that the semitransparent p-contact has been replaced with a transparent conductive oxide, such as indium tin oxide (ITO), and the epi is grown on patterned sapphire. Due to lateral direction of current spreading in this architecture, it is often called a lateral die (*LD*). Often, to reduce the parasitic interaction

between the die and a package, there is a reflector deposited on the sapphire surface. The patterned sapphire, combined with the lower loss from the transparent oxide, can yield architectures which achieve very high extraction efficiencies, upward of 88 % (Narukawa et al. 2010).

The LD in Fig. 7b came to dominate the low- and mid-power needs of the display industry. While low cost and straightforward, the die and the package, which will be discussed in detail later in the chapter, are limited in how much power density an LED can handle. Therefore, the high-power LEDs and the die followed different architecture directions.

The flip-chip architecture (FCLED), shown in Fig. 7c, allowed for excellent current spreading and heat extraction while enhancing extraction efficiency by utilizing a large area, highly reflective, silver-based p-contact to reflect photons out of the LED and also allow a large fraction of the photons to avoid any interaction with the metal contacts (Wierer et al. 2001). The extraction of heat was facilitated by means of metal interconnects between the die and a submount. The proximity of the p-contact to the quantum wells in FCLEDs was later used by Shen et al. to further enhance the extraction efficiency by tuning the dipole radiation pattern to maximize the energy coupled into the GaN/sapphire escape cone (Shen et al. 2003).

Roughening of the high-/low-index interfaces had been proposed and utilized for IR LEDs as the means to break the total internal reflection which was limiting the extraction efficiency of LEDs (Stern 1964; Joyce et al. 1974; Schnitzer et al. 1993). In 1998, Wong et al. demonstrated the separation of a sapphire substrate from a GaN film using a UV (excimer) laser pulse (Wong et al. 1998) and applied this technique to demonstrate thin-film LEDs (Wong et al. 1999). In 1996, Minsky et al. (1996) demonstrated that a GaN surface could be photo-electrochemically etched. Laser-assisted lift-off (LLO) and PEC etching of the consequently exposed n-GaN were then used to significantly enhance the extraction efficiency of GaN-based LEDs. In 2004, two teams, from UCSB and from OSRAM, demonstrated this approach in a so-called vertical thin-film (VTF) die architecture, shown in Fig. 7f. The VTF architecture was fabricated by bonding the metallized p-side of the GaN to a conductive submount while the substrate on which GaN has been grown is removed, followed by roughening of the exposed n-GaN and deposition of a patterned n-contact. At that time the teams reported an almost doubling of the extraction efficiency of un-encapsulated chips (Fujii et al. 2004; Haerle et al. 2004). Application of lift-off and PEC to a flip-chip LED enabled the realization of the thin-film-flip-chip architecture (TFFC) with an extraction efficiency over 80 % reported by a Philips Lumileds team (Shchekin et al. 2006; Krames et al. 2007). The TFFC parts had a higher extraction efficiency than VTF because the emitting surface was not occluded by the n-contact on top of the n-GaN as was the case with VTF. Also, because of the absence of wire bonds, TFFC chips could be combined into compact high-brightness arrays and were readily suited for use with high-performance ceramic phosphors. The need for higher extraction efficiency in the VTF architecture motivated introduction of what could be called an “embedded-contact VTF” (EC-VTF) shown in Fig. 7g. In this architecture, the n-contact features resemble those previously found in flip-chip and TFFC LEDs but still feature a wire bond

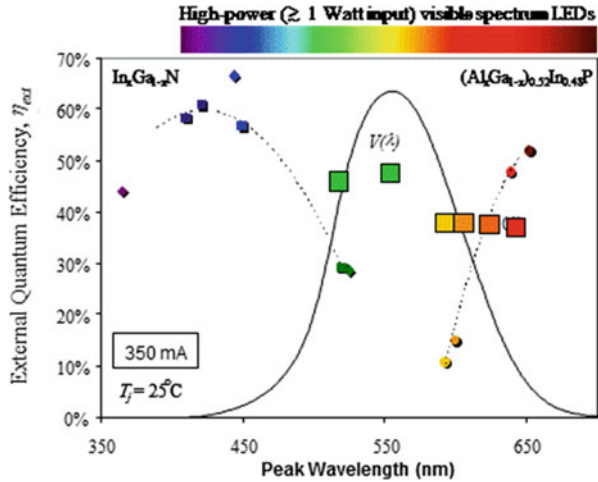
similar in appearance to VTF. It is important to note that in the industry, the thin-film architectures are also made in high volume with epi grown on substrates which can be chemically removed or thinned such as silicon, silicon carbide, and GaN. Due to the excellent heat dissipation ability and high extraction efficiency, flip-chip and the thin-film architectures (TFFC, VTF, and EC-VTF) have become the dominant die architectures for the high-power LEDs.

The drawback of the thin-film based emitters is that there is need for an additional substrate to provide mechanical support. As discussed, in case of the VTF or EC-VTF, the support comes from an intermediate substrate carrier and ultimately requires a separate substrate which provides electrical contacts via wire bonds. In the case of TFFC, there is a separate submount/substrate which provides electrical connections and also mechanical support. Figure 7e shows an example of a chip-scale package (CSP) approach, where a flip-chip architecture is fabricated from epi grown on patterned sapphire. The sapphire is left on and provides mechanical stability such that the die itself can have solder pads and be handled as a stand-alone package. Such a CSP approach combines the high-performance attributes of flip chip, robustness of patterned sapphire, and compactness without a need for an intermediate substrate.

In the late 1990s, it became apparent that for the InGaN/GaN LEDs grown on sapphire substrates, the internal quantum efficiency (IQE) of the radiative recombination of charge carriers peaked at fairly low current density and then monotonically decreased with current drive (Kim et al. 2001; Nakamura and Fosol 1998). The origins of this “droop” have been a rich topic of research and discussion. The phenomena has been attributed to spillover of carriers (Kim et al. 2007; Bochkareva et al. 2010; Özgür et al. 2010), recombination at dislocations (Monemar and Semelius 2007), and Auger recombination (Shen et al. 2007). Although the origin of the droop is still an open question, the Auger theory is finding growing support (Kioupakis et al. 2011). A number of directions are being pursued by researchers in the field to reduce droop, with the efforts focused on alternative growth substrates, such as nonpolar and semipolar GaN (Chakraborty et al. 2005; Zhao et al. 2011) and optimized active region design.

YAG:Ce has been one of the earliest and most enduring choices of phosphor materials for making white LEDs. It is efficient, stable, relatively low cost, and excitable in blue and has emission broad enough to produce light of reasonable quality. Introduced in products by Nichia Corporation around 1996, it is still found in LEDs for a broad range of applications. However, using YAG alone for making white LEDs allows only CCTs of ~ 4000 K and above with CRI ~ 70 . For warmer CCTs and higher CRI, approaches were investigated using combinations of at least two phosphors: green and red (Mueller and Mueller-Mach 2000). Some of the combinations considered early on included (Ca, Se)S:Eu for red and thiogallate for green. While these materials could produce spectra with excellent color rendering and high lumen equivalent, there was a need for robust red phosphors without some of the shortcomings of thiogallates and sulfides, such as poor performance at high drive and stability in humid environments. In 1995, Schlieper et al. (1995) reported the synthesis of nitride-silicate $\text{Sr}_2\text{Si}_5\text{N}_8$ and Ba_2SiN_8 , (258) europium-activated

Fig. 8 Illustration adapted from Ref. (Shchekin et al. 2010) showing how the yellow gap in the direct-emitter efficiency can be bridged by fully converted phosphor LEDs



nitride phosphors. In 2006, Uheda et al. (2006) reported nitride $CaAlSiN_3:Eu$ red phosphor for use in white LEDs. These nitride reds have proven to have higher quantum efficiency and greater stability than either sulfides or orthosilicate (Tasch et al. 2001) phosphors while offering a broad range of emission wavelengths accessible by tuning of the stoichiometry of the material. The nitride phosphors have also created opportunities for new applications and further improvements in lm/W.

In 2009, a team from Philips Lumileds demonstrated and productized a fully converted amber-color-emitting pcLED (Mueller-Mach et al. 2009). The device used a 258-nitrido-silicate phosphor sintered into a ceramic plate placed onto a TFFC blue pump LED. The resulting amber LED, using the higher efficiency blue pump, outperformed direct-emitting amber LEDs based on $AlInGaP$. Since the nitride can be prepared as a ceramic, it is possible to reduce the scattering of blue light from the phosphor, making this approach more efficient and resulting in color with greater saturation than fully converting blue using powder phosphors. Figure 8 shows how “the yellow gap” in the efficiency of direct-emitting color LEDs can be spanned by pcLEDs utilizing various nitrides (Bechtel et al. 2010).

Currently, the combinations of nitride reds and aluminum/gallium garnet yellow and green phosphors are the state of the art for efficient phosphor conversion in white LEDs. Quantum efficiency of conversion is over 90 % for 1W power LED emitters and higher for low-power emitters. The remaining quantum efficiency losses come from photon absorption in the package and are worsened by mechanisms impeding photon extraction out of the package, such as scattering and Fresnel reflections. Figure 9 shows the breakdown of efficiency components for a representative, as of this writing, warm white LED in comparison to practical limits. Here the practical limits for epi IQE and phosphor quantum efficiency can be debated, as the limiting mechanisms are not universally agreed on, while the quantum deficit and lumen equivalent are for LED spectra with CCT of 3000 K, CRI of 80 and red rendering index (R9) greater than zero.

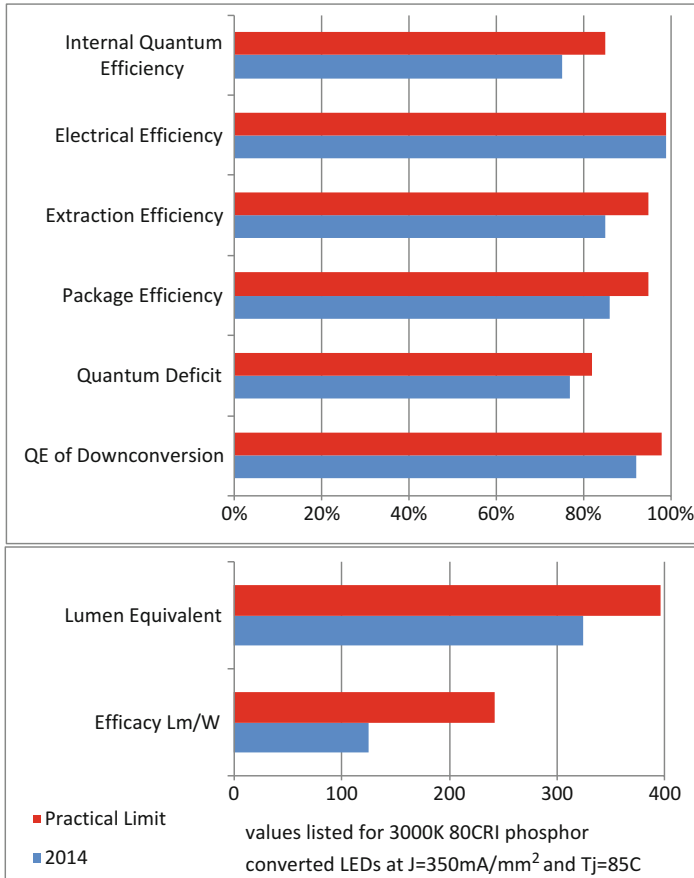


Fig. 9 Efficacy breakdown for a typical warm white phosphor-converted LED at $J = 350 \text{ mA/mm}^2$ and $T_j = 85 \text{ }^\circ\text{C}$

For phosphor-converted LEDs, the most significant areas for improvement in lm/w are in epi IQE and the shape (lumen equivalent) of the white emission spectrum. The limitations of the current red nitride phosphors lie in the width of the emission spectra. At 70–100 nm FWHM, these phosphors generate a considerable amount of light at wavelengths to which human eye has low sensitivity. Figure 10 shows the human eye photopic response curve against a comparison of the emission spectrum of a typical warm white LED with common nitride red phosphors and a hypothetical spectrum for the same CCT, but with the red phosphor with FWHM of 30 nm. The difference in lumen equivalent between the two spectra is 17 %.

The anticipated lumen gain depends on CCT and color rendering specification and is generally 10–20 %. The optimal narrow red phosphors, with high efficiency, stability, and emission wavelength for maximum lumen equivalent, do not exist yet, but there are some candidates emerging. Quantum dots, which, have been

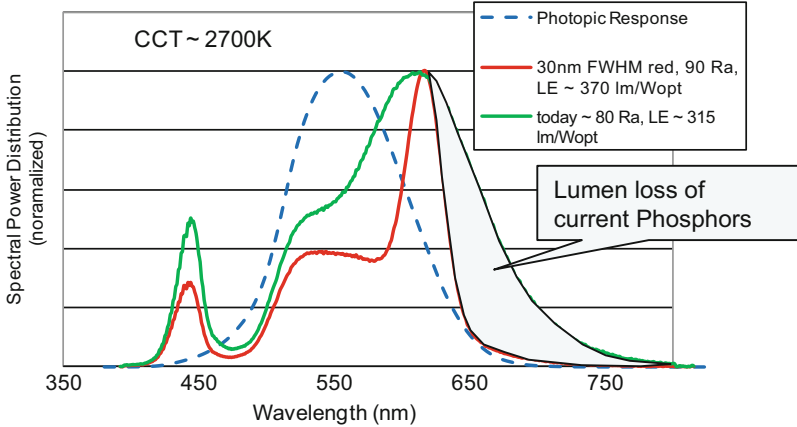


Fig. 10 An illustration of lumen benefit from reducing the width of the red phosphor spectrum

considered for use in LEDs for some time (Bawendi et al. 2002), have been introduced in display applications in configurations where the converting element is not in immediate contact with the blue die. Mn $4+$, activated fluorosilicates (Paulusz 1973; Setlur and Radkov 2010) have been proposed for use in lighting. Recently, a Eu $2+$ -activated nitride phosphor has been announced with FWHM of 50 nm (Pust et al. 2014). This material is particularly interesting as it has potential to be readily usable in all LED emitter architectures similar to the commonplace broad-band Eu $2+$ -activated nitride phosphors.

Other ways to generate white light with LEDs, besides phosphor conversion, is to mix direct red, blue, and green emitters or utilize so-called “hybrid” light engine, where a phosphor-converted LED with a color point slightly off-Planckian is combined with a direct-emitting AlInGaP-based red LED. These approaches potentially allow the reduction or elimination of energy losses associated with phosphor conversion (Fig. 11).

The RGB solution with direct color emitters is severely limited by the IQE of the AlInGaP material system in the green wavelengths. Additionally, for high color rendering, one would need to add a direct yellow emitter, for which there is no high efficiency option. Because of the limited efficiency in the yellow/green colors, the direct-emitter solutions are not used in the general lighting applications where high efficiency and color rendering are specified. However, the direct RGB emitters are widespread in architectural lighting applications, where the decorative properties of the three color LEDs offer significant reduction in power consumption over filtered light from conventional light sources. The hybrid approach, however, has been very successful in achieving high luminous efficacies, especially in applications which allow use of multiple LEDs and where there is need for significant red content and high color rendering. Recently, Philips reported a TLED prototype (LED replacement of fluorescent T lamp) with 200 lm/W at CCT 3000–4500 K and uncompromised color rendering CRI > 80 and R9 > 20 (Details of the 200lm/W

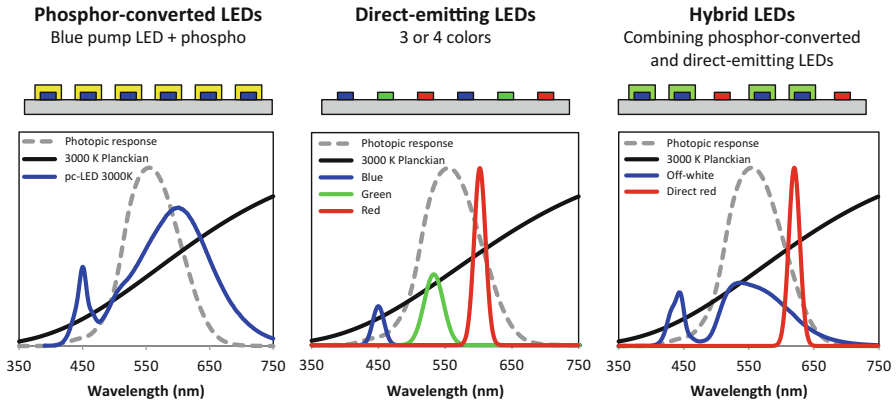


Fig. 11 Types of SSL sources for maximum luminous efficacy

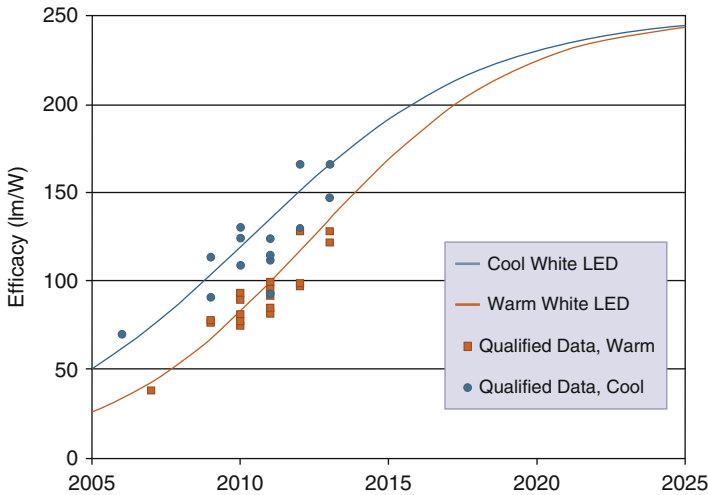


Fig. 12 US DOE white-light LED package efficacy projections for commercial product (US Department of Energy 2014)

TLED lighting technology breakthrough unraveled). The Hybrid approach has also been used in many commercial products, for example, in the DOE L-Prize winning 60 W-equivalent LED bulb from Philips (Rice 2011).

Figure 12 shows the lm/W efficacy projections published in the 2014 edition of the US Department of Energy Solid-State Lighting Research and Development Multi-year Program Plan (MYPP).

According to the MYPP, the maximum efficiency for pcLED and mixed/hybrid sources are expected to be similar at ~250 lm/W with the assumption of 25 °C and 35A/cm². Even though the hybrid light engines allow reduced phosphor conversion

losses, the IQE of the AlInGaP material system is limited, especially as operating temperatures increase. A similar projection but at 85 °C junction temperature and warm white taken at 3000 K CRI 90, R9 50 puts both pLEDs and hybrids at a practical maximum of 225 lm/W (Soer et al. 2014). The hybrid sources are expected to reach near-maximum efficacies sooner than pLEDs as the direct-emitting red LED sources are readily available, while the suitable narrow red phosphors for pLEDs require basic materials development and discovery.

Ultimately, with 100 % internal, electrical, and package efficiency, the luminous efficacy would be limited by the spectrum and the desired quality of color rendering. Per estimate by Phillips et al. (2007), one can place the ultimate limit for lm/W of a white LED at ~400 lm/W. To move past the 200–250 lm/W limit of present-day technologies, it is necessary to exclude phosphors and emit a white spectra directly, which would necessitate substantial, long-term investment into the existing compound semiconductor systems or introduction of new ones. Examples of technologies which may be capable of bringing greater efficacies are nitride-based micro- and nano-wires (Wang et al. 2014) or new materials systems based on quantum dots (Mashford et al. 2013) or perovskite crystals (Tan et al. 2014).

Laser-Based Solid-State Sources

Even though solid-state lighting is conventionally associated with LEDs, lasers have always been considered as possible options for sources of white light through mixing of directly emitted color light or utilizing blue light to pump phosphors. As has been mentioned (Pinnow and Gerard Van Uitert 1970), some of the initial ideas for phosphor conversion detailed the use of blue lasers and YAG phosphors. Lasers offer possibility of relatively small emitting surface and high efficiency at high-input power density, making them attractive for use in high-brightness applications. At the same time, the high price and low peak efficiency have kept semiconductor lasers out of use in general lighting. The examples of successful application of lasers include overhead and cinema projectors (Beck) and high-end car headlights, as demonstrated by BMW (Hanafi and Erdl 2015). In the case of overhead projectors and car headlights, the underlying technology is similar to LED-based lighting where a blue source, in this case a blue diode laser, is used to excite remote phosphor to produce blue and green/yellow light. For the cinema projection, the technology is more exotic, where red and blue light is delivered by red and blue laser diodes, while green colors are obtained by frequency doubling infrared diode light (Fig. 13).

LED Packaging

LED packaging has evolved driven by target applications, reliability requirements, and cost. Through the 1970s and the 1980s, LEDs were used primarily as indicators. The typical package of that time is the, very familiar, 5 mm lamp which schematic is

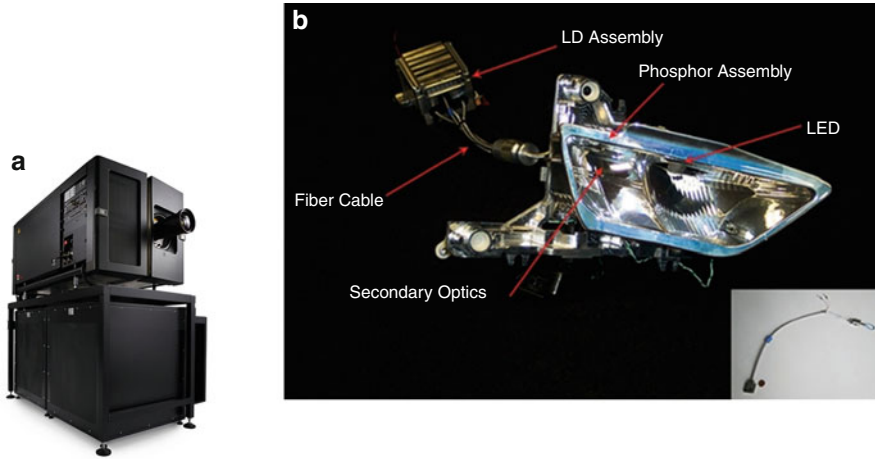


Fig. 13 Examples of products using semiconductor lasers: (a) laser-based cinema projector; (b) laser-based headlight (illustration: Compound Semiconductor) (Credit: Barco)

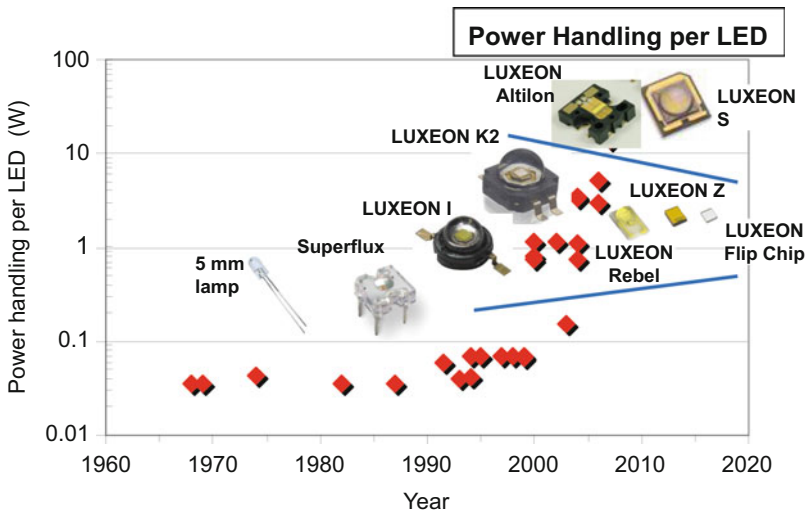


Fig. 14 High-power LED package evolution with examples of emitters introduced by HP/Philips Lumileds

shown in Fig. 14a. Such lamps were designed for 0.1 W input of electrical power. A small die was housed and connected to two leads held together by the hard epoxy lens. The lens provides an optical function, comes in a variety of shapes, and could include color filtering. As LED efficiency improved, higher-power LED packages were designed for emerging applications. An example of this is the 0.2 W SuperFlux

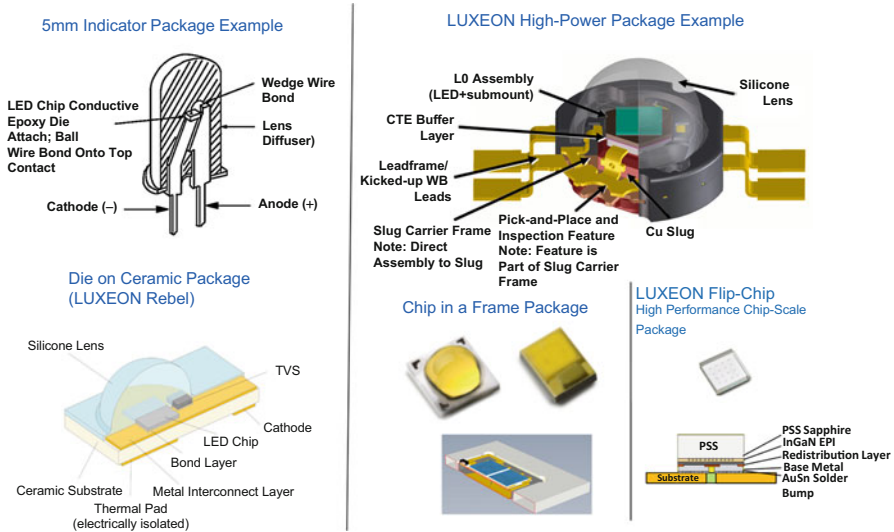


Fig. 15 The trend in high-power LED packages is toward smaller size and greater flexibility

LED introduced by Hewlett-Packard for the automotive signaling market. In 1998, Hewlett-Packard introduced a 0.5 W red power emitter designed for traffic lights. In 2001, Lumileds (formerly Hewlett-Packard) introduced the LUXEON I white high-power LED which used a 1 mm² flip-chip die with the package designed to dissipate the heat from the die through a dedicated heat sink. The design allowed up to 3 W electrical power input and junction temperatures up to 120 °C. To increase the power handling further, the LUXEON K2 package was released where a number of high-performance technologies were implemented, allowing the junction temperature to be raised up to 150 °C for white and 185 °C for direct emitters and drive current up to 1.5 Amps. In 2007, Philips Lumileds introduced the LUXEON Rebel package which started the LED industry switch to die-on-ceramic (DoC) technology. DoC allowed significant simplification and cost reduction over previous high-power packages while offering all of the benefits in a more compact form factor.

The technology in LUXEON K2 and DoC is leveraged for high-power array emitters used for illumination and automotive forward lighting. The improvements in efficiency, packaging materials, and the pressure to reduce component cost led to further miniaturization of the high-power LEDs. The use of aluminum nitride allowed reduction in the attach footprint as well as a choice of whether L1 optical elements were used. This flexible approach is illustrated in Fig. 15 where the chip in a frame concept offers the option to have a flat package instead of a dome for a lower cost emitter with smaller source size (LUXEON ZES). The chip-scale package (CSP) is the emerging concept in high-power LED architecture and is illustrated in Fig. 7e. Here the solder pads are on the die itself, and the optical elements can be

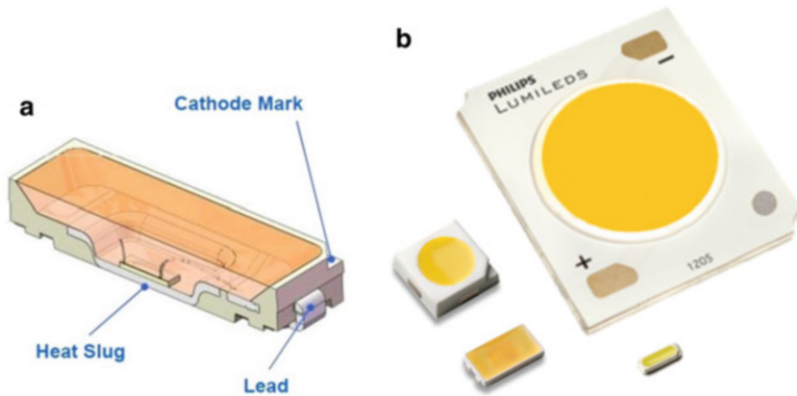


Fig. 16 (a) Schematic of a typical mid-power LED, (b) illustration of the scalability of the mid-power architecture from 10 lm to 5000 lm emitters

shaped into the die or molded around it. The heat management of CSP devices is achieved partly by the high efficiency of the emitter and by the board to which the CSP is attached. A CSP can be used by itself or as an emitter in a separate package with the latter adding thermal or optical functionality.

Another type of LED architecture, often referred to as low or mid power, is illustrated in Fig. 16. At the heart of the LED is, usually, a blue lateral die as shown in Fig. 7b, which is attached to an overmolded lead-frame package. Phosphor mix or silicone is then dispensed into the cavity (“cup”) formed in the package. This architecture originates from display applications, where similar packages were first used for backlighting low-power mobile displays. As their efficiency improved, the LEDs could be driven harder, allowing for use in large screen displays and ultimately in lighting applications. The primary attraction of this architecture is its low cost, since the die and the package leverage the high-volume manufacturing used for display applications. Also, such emitters are quite efficient. The efficiency of the blue die has been mentioned earlier in the chapter. The high efficiency of the package comes from a relatively large volume of phosphor, reflective package materials and lower current density at which the die is driven compared to a typical high-power emitter. The architecture is highly scalable as multiple dies can be placed in packages of various sizes, as is shown in Fig. 15b.

The disadvantage of the mid-power architecture is that the power density cannot be as high as that for specifically designed high-power emitters. While high lumen output has been demonstrated for the large CoB arrays, the surface brightness is lower than that of high-power emitters.

Figure 17 maps the increasing diversity of the LED packages across source luminance and lumen output. This is a useful view to help link the solid-state source to an application and also see where a need can be addressed with more than

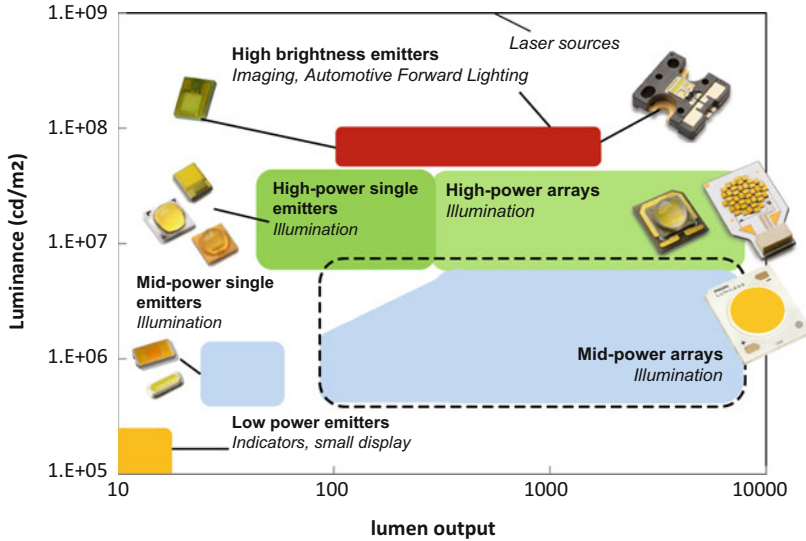


Fig. 17 Lumen output versus luminance map of the various LED sources

one type of emitter. For high luminance applications, for example, automotive forward lighting, mobile camera flash, directional outdoor illumination, spot lights, and down lights, which demand $>10^7$ cd/m², high-power LEDs dominate. Here the emitting surface (source) has to be small and bright to allow for narrow and/or well-defined beams. This usually requires power densities at which the reliability of mid-power architectures is compromised.

For applications such as nondirectional lamps, the mid-power and high-power emitters and arrays strongly overlap in the region of 5×10^5 to $\sim 5 \times 10^6$ cd/mm² (the dashed area in Fig. 16). Here cost and intended application determine the selection. Indeed, in applications where the luminance is not critical, high-power, mid-power, and, in some cases, low-power LEDs compete. In this case, the key metric becomes the light output at a given cost, expressed as lm/\$, and now performance (lm/W and W/mm²) and cost (\$/mm²) are inherently linked. As an example, during the last few years, the mid-power LEDs have almost entirely displaced high-power LEDs in A19 lamps, and low-power LEDs are displacing mid power in tube lights. The application landscape for the various LED architectures will continue to be dynamic as the efficacies increase and costs decrease. The low cost of the low- and mid-power emitters based on lateral die will continue to be attractive. As the luminous efficacy of the LEDs increase, these architectures will start addressing applications where the high brightness requirements of the source could previously be fulfilled only by high-power emitters. It is likely that the high-end camera flash, projection, and automotive forward lighting will continue to be

addressed by high-power emitters, as in those applications, there is value to going $10\times$ current levels of brightness that even today's high-power emitters cannot achieve yet.

Conclusion

This chapter reviewed the pioneering work which formed the technology foundation for solid-state light sources. The variety of technology building blocks, LED architectures, and lasers and the growing multitude of packages to address lighting applications were discussed. LED-based solid-state lighting is rapidly displacing the traditional sources, with efficiency levels $10\text{--}20\times$ that of incandescents. At the same time, the SSL technology has not reached the level of standardization and technology homogeneity as seen in other mature fields. There are five mainstream architectures for blue LED die and four crystal substrates in use for epitaxial growth. The contribution and respective limits of each LED building block to luminous efficacy can be clearly quantified. Therefore, the progress toward the practical limits of efficacy and the contribution of technology to emitter cost will lead to standardization of building blocks and chip design. Similarly, for LED packages and LED arrays, there are applications where there is clear preference for high-power or mid- to low-power emitter architectures. However, there are numerous applications which can be addressed by a variety of emitter types. Here also, the value of higher luminous efficacy and the emitter cost per lumen output will sort out the winning designs as the emitter technology progresses. One trend will dominate for high-volume applications, that is, the cost will remain at the forefront of requirements once the minimum system performance has been attained for any given application. Indeed, performance is traded off against cost by either reducing the number of LEDs or by cost reduction in other components (e.g., heat sinks) to result in a lower system-level cost. This then bounds the value of increased performance. The practical limit in lm/W for the phosphor-converted LEDs as well as hybrid arrays is estimated at $200\text{--}250$ lm/W. There are laboratory examples of retrofit lamps with efficacy of 200 lm/W and commercial offerings of LEDs with high color rendering near 140 lm/W. It is clear what is needed to reach the practical limits in luminous efficacy. To go beyond the practical limits, significant investments in new technologies are required. The motivation for such investments will be a topic of debate in the future. For some applications, such as automotive forward lighting, projection and camera flash, source brightness, and, therefore, maximum power density has definite value. For such applications, there is clear motivation to maximize luminous efficacy at high drive. This is where the role of laser-based sources will be significant. For general lighting, LED-based sources are expected to provide optimal lm/\$ and lm/W for most applications. It is still to be determined

how much motivation there will be to pursue new technologies to push beyond the efficacies that will be achieved as a result of optimizing the existing LED technologies over the next decade.

References

- Alferov ZI, Andreev VM, Korol'kov VI, Portnoi EL, Tret'yakov DN (1969) Injection properties of n-AlxGa(1-x)As-p-GaAs heterojunctions. *Sov Phys Semicond* 2:843
- Amano H, Sawaki N, Asaka I, Toyoda Y (1986) Metal organic growth of a high quality GaN film using an AlN buffer layer. *Appl Phys Lett* 48:353–355
- Baretz B, Tischler MA (2003) Solid state white light emitter and display using same. US Patent 6,600,175, 29 Jul 2003
- Bawendi MG, Heine J, Jensen K, Miller J, Moon R (2002) Quantum Dot white and colored light emitting diodes. US Patent 6,501,091, 31 Dec 2002
- Bechtel H, Schmidt PJ, Tücks A, Heidemann M, Chamberlin D, Müller-Mach R, Müller GO, Shchekin O (2010) Fully phosphor-converted LEDs with Lumiramic phosphor technology. In: Tenth international conference on solid state lighting, 77840W
- Beck B. Lasers: coming to a theater near you. *IEEE Spectrum*, 22 Feb 2014. [Online]. Available at <http://spectrum.ieee.org/consumer-electronics/audiovideo/lasers-coming-to-a-theater-near-you>
- Biard JR, Pittman GE (1966) Semiconductor radiant diode. US Patent 3,293,513, 20 Dec 1966
- Bochkareva NI, Voronenkov VV, Gorbunov RI, Zubrilov AS, Lelikov YS, Latyshev FE, Rebane YT, Tsyuk AI, Shreter YG (2010) Mechanism of the GaN LED efficiency falloff with increasing current. *Semiconductors* 44(6):794
- Brattain W, Bardeen J, Shockley WH (1948) *Phys Rev* 74:230–232
- Broell M, Sundgren P, Rudolph A, Schmid W, Vogl A, Behringer M (2014) New developments on high efficiency infrared and InGaAlP light emitting diodes at OSRAM opto-semiconductors light-emitting diodes: materials, devices, and applications for solid state lighting XVIII. In: Streubel KP, Heonsu Jeon, Li-Wei Tu, Strassburg M (eds) *Proceedings of SPIE*, vol 9003, 90030L
- Butterworth M, Helbing R (1998) Fluorescent dye added to epoxy of light emitting diode lens. US Patent 5,847,507, 8 Dec 1998
- Chakraborty A, Haskell BA, Keller S, Speck JS, DenBaars SP, Nakamura S, Mishra UK (2005) Demonstration of nonpolar m-plane InGaN/GaN light-emitting diodes on free-standing m-plane GaN substrates. *Jpn J Appl Phys* 44:L173–L175
- Craford MG (2013) From Holonyak to today. *Proc IEEE* 101(10):2170–2175
- Craford MG, Shaw RW, Herzog AH, Groves WO (1972) Radiative recombination mechanism in GaAsP diodes with and without nitrogen doping. *J Appl Phys* 43:4075
- Details of the 200lm/W TLED lighting technology breakthrough unraveled, 11 Apr 2013. [Online]. Available at <http://www.newscenter.philips.com/main/standard/news/articles/20130411-details-of-the-200lm-w-tled-lighting-technology-breakthrough-unraveled.wpd#.VJtLsF4AB4>
- Ehrenreich H (1960) Band Structure and Electron Transport of GaAs. *Phys Rev* 120:1951
- Elgala H, Mesleher R, Haas H (2011) Indoor optical wireless communication: potential and state-of-the-art. *Commun Mag* 49(9):56–62
- Fujii T, Gao Y, Sharma R, Hu EL, DenBaars SP, Nakamura S (2004) Increase in the extraction efficiency of GaN-based light-emitting diodes via surface roughening. *Appl Phys Lett* 84(6):855
- Groves WO, Herzog AH, Craford MG (1971) The effect of nitrogen doping on GaAsP electroluminescent diodes. *Appl Phys Lett* 19:184
- Haerle V, Hahn B, Kaiser S, Weimar A, Bader S, Eberhard F, Plössl A, Eisert D (2004) High brightness LEDs for general lighting applications using the new ThinGaN™-technology. *Phys Status Solidi A* 201(12):2736

- Hall RN, Fenner GE, Kingsley JD, Soltys TJ, Carlson RO (1962) Coherent light emission from GaAs junctions. *Phys Rev Lett* 9:366
- Hanafi A, Erdl H (2015) Lasers light the road ahead. In: Compound semiconductor, 14 July 2015. Available at <http://www.compoundsemiconductor.net/article/97529-lasers-light-the-road-ahead.html>
- Hiramatsu K, Itoh S, Amano H, Akasaki I, Kuwano N, Shiraishi T, Oki K (1991) Growth mechanism of GaN grown on sapphire with AlN buffer layer by MOVPE. *J Cryst Growth* 115:628–633
- Höhn K, Debray A, Schlotter P, Schmidt R, Schneider J (2001) Wavelength-converting casting composition and light-emitting semiconductor component. US Patent 6,245,259, 12 Jun 2001
- Holonyak N Jr, Bevacqua SF (1962) Coherent (visible) light emission from Ga(As_{1-x}P_x) junctions. *Appl Phys Lett* 1(82)
- Huang KH, Yu JG, Kuo CP, Fletcher RM, Osentowski TD, Stinson LJ, Craford MG, Liao A (1992) Twofold efficiency improvement in high performance AlGaInP light-emitting diodes in the 555–620 nm spectral region using a thick GaP window layer. *Appl Phys Lett* 61(9):1045–1047
- Ikeda M, Nakano K, Mori Y, Kaneko K, Watanabe N (1986) MOCVD growth of AlGaInP at atmospheric pressure using triethylmetals and phosphine. *J Cryst Growth* 77:380
- Itaya K, Ishikawa M, Uematsu Y (1990) 636 nm room temperature cw operation by heterobarrier blocking structure InGaAlP laser diodes. *Electron Lett* 26:839
- Joyce WB, Bachrach RZ, Dixon RW, Sealer DA (1974) Geometrical properties of random particles and the extraction of photons from electroluminescent diodes. *J Appl Phys* 45:2229
- Kim AY, Goetz W, Steigerwald DAWJJ, Gardner NF, Sun J, Stockman SA, Martin PS, Krames MR, Kern RS, Steranka FM (2001) Performance of High-Power AllInGaN Light Emitting Diodes. *Phys Status Solidi A* 15:188
- Kim MH, Schubert MF, Dai Q, Kim JK, Schubert EF, Piprek J, Park Y (2007) Origin of efficiency droop in GaN-based light-emitting diodes. *Appl Phys Lett* 91(18):183507
- Kioupakis E, Rinke P, Delaney KT, Van De Walle CG (2011) Indirect Auger recombination as a cause of efficiency droop in nitride light-emitting diodes. *Appl Phys Lett* 98(16):161107
- Kish FA, Steranka FM, DeFevre DC, Vanderwater DA, Park KG, Kuo CP, Osentowski TD, Peanasky MJ, Yu JG, Fletcher RM, Steigerwald DA, Craford MG, Robbins VM (1994) Very high-efficiency semiconductor wafer-bonded transparent-substrate (Al_xGa_{1-x})_{0.5}In_{0.5}P/GaP light-emitting diodes. *Appl Phys Lett* 64(21):2839–2841
- Kobayashi K, Kawata S, Gomyo A, Hino I, Suzuki T (1985) Room-temperature cw operation of AllnGaP double-heterostructure lasers. *Electron Lett* 239:931
- Krames MR, Ochiai-Holcomb M, Hoerfler GE, Carter-Coman C, Chen EI, Tan IH, Grillot P, Gardner NF, Chui HC, Huang JW, Stockman SA, Kish FA, Craford MG, Tan TS, Kocot CP, Hueschen M, Posselt J, Loh B, Sasser G, Collins D (1999) High-power truncated-inverted-pyramid Al_xGa_(1-x)_{0.5}In_{0.5}P/GaP light-emitting diodes exhibiting >50% external quantum efficiency. *Appl Phys Lett* 75(16):2365
- Krames MK, Shchekin OB, Mueller-Mach R, Mueller GO, Zhou L, Harbers G, Craford MG (2007) Status and future of high-power light-emitting diodes for solid-state lighting. *IEEE J Disp Tech* 3:160–175
- Kuo CP, Fletcher RM, Osentowski TD, Lardizabal MC, Craford MG, Robbins VM (1990) High performance AlGaInP visible light emitting diodes. *Appl Phys Lett* 57:2937–2939
- Lehovec K, Accardo CA, Jamgochian E (1951) Injected light emission of silicon carbide crystals. *Phys Rev* 83:603–608
- Lei T, Fanciulli M, Molnar RJ, Moustakas TD, Graham RJ, Scanlon J (1991) Epitaxial growth of zinc blende and wurtzitic gallium nitride thin films on (001) silicon. *Appl Phys Lett* 59:944
- Loebner EE (1976) Subhistories of the light emitting diode. *IEEE Trans Electron Dev* 23(7):675
- Logan RA, White HG, Trumbore FA (1967) P-n junctions in GaP with external electroluminescence efficiencies ~ 2% at 25 °C. *Appl Phys Lett* 10:206
- Losev OV (1929) Soviet patent 12191

- Manasevit HM (1968) Single-crystal gallium arsenide on insulating substrates. *Appl Phys Lett* 12:156
- Maruska H. A brief history of GaN blue light emitting diodes. LIGHTTimes Online – LED Industry News. [Online]. Available at http://www.sslighting.net/news/features/maruska_blue_led_history.pdf
- Maruska HP, Stevenson DA (1974) Mechanism of light production in metal-insulator-semiconductor diodes; GaN:Mg violet light-emitting diodes. *Solid State Electron* 17:1171
- Maruska HP, Tietjen JJ (1969) The preparation and properties of vapor-deposited single-crystal-line gan. *Appl Phys Lett* 15:327
- Maruska HP, Rhines WC, Stevenson DA (1972) Preparation of Mg-doped GaN diodes exhibiting violet electroluminescence. *Mater Res Bull* 7:777
- Mashford BS, Stevenson M, Popovic Z, Hamilton C, Zhou Z, Breen C, Steckel J, Bulovic V, Bawendi M, Coe-Sullivan S, Kazlas PT (2013) High-efficiency quantum-dot light-emitting devices with enhanced charge injection. *Nat Photonics* 7:407–412
- Minsky MS, White M, Hu EL (1996) Room-temperature photoenhanced wet etching of GaN. *Appl Phys Lett* 68(11):1531
- Monemar B, Sernelius BE (2007) Defect related issues in the “current roll-off” in InGaN based light emitting diodes. *Appl Phys Lett* 91:181103
- Mueller GO, Mueller-Mach R (2000) White light emitting diodes for illumination *Proc SPIE* 3938 (30):30–41
- Mueller-Mach R, Mueller GO, Krames MR, Shchekin OB, Schmidt PJ, Bechtel H, Chen C-H, Steigelmann O (2009) All-nitride monochromatic amber-emitting phosphor-converted light-emitting diodes. *Phys Status Solidi (RRL)* 3(7–8):215–217
- Nakamura S (1991) GaN growth using GaN buffer layer. *Jpn J Appl Phys* 30:L1705
- Nakamura S, Fosol G (1998) *The blue laser diodes*. Springer, Berlin
- Nakamura S, Iwasa N, Senoh M, Mukai T (1992) Hole compensation mechanism of P-type GaN films. *Jpn J Appl Phys* 31:1258
- Nakamura S, Senoh M, Mukai T (1993) P-GaN/N-InGaN/N-GaN double-heterostructure blue-light-emitting diodes. *Jpn J Appl Phys* 32:L8
- Nakamura S, Mukai T, Senoh M (1994) Candela class high brightness InGaN/AlGaIn double heterostructure blue light emitting diodes. *Appl Phys Lett* 64:1687
- Narukawa Y, Ichikawa M, Sanga D, Sano M, Mukai T (2010) White light emitting diodes with super-high luminous efficacy. *J Phys D Appl Phys* 43. Art. ID 354002
- Nishizawa J, Suto K, Teshima T (1977) Minority-carrier lifetime measurements of efficient GaAlAs p-n heterojunctions. *J Appl Phys* 48:3484
- Ohba Y, Ishikawa M, Sugawara H, Yamamoto T, Nakanishi T (1986) Growth of high-quality InGaAlP epilayers by MOCVD using methyl metalorganics and their application to visible semiconductor lasers. *J Cryst Growth* 77:374
- Özgür Ü, Liu H, Li X, Ni X, Morkoç H (2010) GaN-based light-emitting diodes: efficiency at high injection levels. *Proc IEEE* 98:1180
- Pankove JI, Berkeyheiser JE (1962) A light source modulated at microwave frequencies. *Proc IRE* 50:1976
- Pankove JI, Lampert MA (1974) Model for electroluminescence in GaN. *Phys Rev Lett* 33:361
- Pankove JI, Massoulié MJ (1962) Injection luminescence from gallium arsenide. *Bull Am Phys Soc* (7) 88
- Pankove JI, Miller EA, Richman D, Ber JE (1971) Electroluminescence in GaN. *J Lumin* 4:63
- Paulusz AG (1973) Efficient Mn(IV) emission in fluorine coordination. *J Electrochem Soc* 120 (7):942–947
- Philips creates shopping assistant with LEDs and smart phone. *IEEE spectrum*, 18 Feb 2014. [Online]. Available at <http://spectrum.ieee.org/tech-talk/computing/networks/philips-creates-store-shopping-assistant-with-leds-and-smart-phone>
- Philips Quarterly report Q3 2014

- Phillips JM, Coltrin ME, Crawford MH, Fischer AJ, Krames MR, Mueller-Mach R, Mueller GO, Ohno Y, Rohwer L, Simmons JA, Tsao JY (2007) Research challenges to ultra-efficient inorganic solid-state lighting. *Laser Photonics Rev* 1(4):307
- Pinnow DA, Gerard Van Uiter LD (1970) Display system. US Patent 3,691,482 A, 19 Jan 1970
- Pust P, Weiler V, Hecht C, Tücks A, Wochnik AS, Henß A-K, Wiechert D, Scheu C, Schmidt PJ, Schnick W (2014) Narrow-band red-emitting Sr[LiAl₃N₄]:Eu²⁺ as a next-generation LED-phosphor material. *Nat Mater* 13:891–896
- Quist TM, Rediker RH, Keyes RJ, Krag WE, Lax B, McWhorter AL, Zeigler HJ (1962) Semiconductor maser of GaAs. *Appl Phys Lett* 1(4):91–92
- Manchester H (1963) Light of hope-or terror. *Readers Digest*, (Feb, 1963) 97
- Rice A (2011) Bulb in, bulb out. *The New York Times*, 3 June 2011. [Online]. Available at http://www.nytimes.com/2011/06/05/magazine/bulb-in-bulb-out.html?pagewanted=all&_r=0
- Round HJ (1907) A note on carborundum. *Elec World* 49:308
- Rupprecht H, Woodall JM, Pettit GD (1967) Efficient visible electroluminescence at 300°K from Ga_{1-x}Al_xAs p-n junctions grown by liquid-phase epitaxy. *Appl Phys Lett* 11:81–83
- Schlieper T, Milius W, Schnick W (1995) High temperature syntheses and crystal structures of Sr₂Si₅N₈ and Ba₂Si₅N₈. *Z Anorg Allg Chem* 621:1380
- Schnitzer I, Yablonovitch E, Caneau C, Gmitter TJ, Schere A (1993) 30% external quantum efficiency from surface textured, thin-film light-emitting diodes. *Appl Phys Lett* 63(16):2174
- Schubert FE (2006) *Light-emitting diodes*, 2nd edn. Cambridge University Press, Cambridge
- Setlur AA, Radkov EV (2010) Energy-efficient, high-color-rendering LED lamps using oxyfluoride and fluoride phosphors. *Chem Mater* 22(13):4076–4082
- Shchekin OB, Epler JE, Trottier TA, Margalith T, Steigerwald DA, Holcomb MO, Martin PS, Krames MR (2006) High performance thin-film flip-chip InGa_N–Ga_N light-emitting diodes. *Appl Phys Lett* 89:071109
- Shchekin O, Mueller G, Mueller-Mach R, Chamberlin D, Bechtel H, Schmidt P, Steigelmann O (2010) Phosphor materials as key enabling ingredients in LED performance. In: *Phosphor Global Summit*, San Diego
- Shen Y-C, Wierer JJ, Krames MR, Ludowise MJ, Misra MS, Ahmed F, Kim AY, Mueller GO, Bhat JC, Stockman SA, Martin PS (2003) Optical cavity effects in InGa_NO_{Ga}N quantum-well-heterostructure flip-chip light-emitting diodes. *Appl Phys Lett* 82(14):2221
- Shen Y-C, Mueller GO, Watanabe S, Gardner NF, Munkhom A, Krames MR (2007) Auger recombination in InGa_N measured by photoluminescence. *Appl Phys Lett* 91:141101
- Shimizu Y, Sakano K, Noguchi Y, Moriguchi T (1999) Light emitting device having a nitride compound semiconductor and a phosphor containing a garnet fluorescent material. US Patent 5,998,925, 7 Dec 1999
- Soer W, Vampola K, Shchekin O (2014) The road to 250 lm/W. In: *2014 Solid-State Lighting R&D Workshop*, Tampa
- Stern F (1964) Transmission of isotropic radiation across an interface between two dielectrics. *Appl Opt* 3:111–114
- Streubel K, Linder N, Wirth R, Jaeger A (2002) High brightness AlGaInP light-emitting diodes. *IEEE J Sel Top Quantum Electron* 8:321–332
- Tan Z-K, Moghaddam RS, Lai ML, Docampo P, Higler R, Deschler F, Price M, Sadhanala A, Pazos LM, Credgington D, Hanusch F, Bein T, Snaith HJ, Friend RJ (2014) Bright light-emitting diodes based on organometal halide perovskite. *Nat Nanotechnol* 9:687–692
- Tasch S, Pachler P, Roth G, Tews W, Kempfert W, Starick D (2001) Light source comprising a light-emitting element. US Patent US 6,809,347 B2, 19 Nov 2001
- Thomas DG, Hopfield JJ, Frosch CJ (1965) Isoelectronic traps due to nitrogen in gallium phosphide. *Phys Rev Lett* 15:857
- U.S. Department of Energy (2014) Solid-state lighting research and development multi-year program plan

- Uheda K, Hirosaki N, Yamamoto H (2006) Host lattice materials in the system $\text{Ca}_3\text{N}_2\text{-AlN-Si}_3\text{N}_4$ for white light emitting diode. *Phys Status Solidi A* 203(11):2712
- Wang R, Nguyen HPT, Connie AT, Lee J, Shih I, Mi Z (2014) Color-tunable, phosphor-free InGaN nanowire light-emitting diode arrays monolithically integrated on silicon. *Opt Express* 22(S7): A1768–A1775
- Wierer JJ, Steigerwald DA, Krames MR, O’Shea JJ, Ludowise MJ, Christenson G, Shen Y-C, Lowery C, Martin PS, Subramanya S, Goetz W, Gardner NF, Kern RS, Stockman SA (2001) High-power AlGaInN flip-chip light-emitting diodes. *Appl Phys Lett* 78(22):3379
- Wolff GA, Herbert RA, Broeder JD (1955) Electroluminescence of GaP. *Phys Rev* 100:753–754
- Wong WS, Sands T, Cheung NW (1998) Damage-free separation of GaN thin films from sapphire substrates. *Appl Phys Lett* 72(5):599
- Wong WS, Sands T, Cheung NW, Kneissl M, Bour DP, Mei P, Romano LT, Johnson NM (1999) Fabrication of thin-film InGaN light-emitting diode membranes by laser lift-off. *Appl Phys Lett* 75(10):1360
- Yoshida S, Misawa S, Gonda S (1983) Improvements on the electrical and luminescent properties of reactive molecular beam epitaxially grown GaN films by using AlN-coated sapphire substrates. *Appl Phys Lett* 42:427–429
- Zhao Y, Tanaka S, Pan CC, Fujito K, Feezel D, Speck JS, DenBaars SP, Nakamura S (2011) High-power blue-violet semipolar (20-2-1) InGaN/GaN light-emitting diodes with low efficiency droop at $200/\text{Acm}^2$. *Appl Phys Express* 4:082104
- Zheludev (2007) The life and times of the LED — a 100-year history. *Nat Photonics* (1): 189–192

Part II

Light-Emitting Diodes

LED Materials: Epitaxy and Quantum Well Structures

Zhen-Yu Li, Hao-Chung Kuo, Chen-Yu Shieh, Ching-Hsueh Chiu, Po-Min Tu, and Wu-Yih Uen

Contents

Introduction	74
Growth and Characterization Methods of Gallium Nitride (GaN)-Related Materials	75
Improvement of Emission Efficiency and Droop by MQWs and EBL Design	76
Wider InGaN Quantum Well	76
Graded-Thickness Multiple Quantum Wells (GQWs)	81
Graded-Composition Multiple Quantum Barriers (GQB)	86
Graded-Composition Electron-Blocking Layer (GEBL)	91
Improvements of the GaN-Based LEDs' Main Material Qualities	96
Freestanding GaN Substrate (FS-GaN)	96
Low-Cost and High-Efficiency GaN-Based LEDs on Large-Area Si Substrate	105
Reference:	116

Abstract

This chapter describes how, in order to achieve low droop and high-efficiency light-emitting diodes (LEDs), we investigated the following multiple quantum wells (MQWs) and electron-blocking layer (EBL) design to enhance our LED

Z.-Y. Li (✉) • H.-C. Kuo • C.-H. Chiu • Po-Min Tu
Department of Photonics and Institute of Electro-Optical Engineering, National Chiao Tung University, Hsinchu, Taiwan
e-mail: chenyu.li@msa.hinet.net; hckuo@faculty.nctu.edu.tw; chinghsuehchiu@gmail.com; bomin.tu@gmail.com

C.-Y. Shieh
Department of Optics and Photonics, National Central University, Jhongli City, Taiwan
e-mail: jokohnson@gmail.com

W.-Y. Uen
Department of Electronic Engineering, College of Electrical Engineering and Computer Science, Chung Yuan Christian University, Chung-Li, Taiwan
e-mail: uenwuyih@ms37.hinet.net

devices: graded-thickness multiple quantum wells (GQWs), graded-composition multiple quantum barriers (GQBs), selectively graded-composition multiple quantum barriers (SGQBs), and graded-composition electron-blocking layer (GEBL). Besides, the crystal quality of the epitaxial layer was enhanced by introducing freestanding GaN substrate for the epitaxial growth of III-nitride epilayer. On the other hand, in recent years, the epitaxial growth of GaN-based materials on Si substrate has a great potential for applications in low-cost and high-efficiency LEDs. Hence, the properties of GaN-based LEDs on Si will also be described in this chapter.

Introduction

It is well known that the blue/green light-emitting diodes (LEDs) and laser diodes (LDs) have been successfully fabricated on sapphire by low-pressure metal-organic chemical vapor deposition (LP-MOCVD) using InGaN/GaN multiple quantum well (MQW) structures and have now become commercialized. However, the InGaN/GaN MQWs still contain a high density of threading dislocations (around 10^8 – 10^{10} cm⁻²) due to the large lattice mismatch and the difference in thermal expansion coefficients of III–V films and sapphire substrates (Cao et al. 2004a). Therefore, so far, many studies have attempted to grow InGaN/GaN MQWs with ultra-flat interfaces and low threading dislocation density (TDD) therein on a sapphire substrate (Son et al. 2006). Simultaneously, numerous investigations demonstrated that the device properties of InGaN/GaN MQW LEDs and/or LDs are affected by the TDD, such as the efficiency and lifetime of InGaN/GaN MQW LEDs and LDs decrease with an increase of TDD in InGaN/GaN MQW. In other words, improving the crystalline quality of InGaN/GaN MQW epilayer structure is required to dramatically enhance the performance of GaN-based LED devices. Additionally, as the efficiency of LEDs increasing, the upcoming challenge is the efficiency “droop” for high-power applications (Kim et al. 2007a). It means that the efficiency reduces rapidly when LED is operating under high carrier density. The major cause of efficiency droop is still a huge controversy. Various possible mechanisms of droop including carrier overflow (Vampola et al. 2009), nonuniform distribution of holes (Ding et al. 2009; Wang et al. 2010), Auger scattering (David and Grundmann 2010), and carrier delocalization (Monemar and Sernelius 2007) have been proposed. In recent years, great efforts have been made to reduce the efficiency droop. Most of them are focus on minimizing the carrier overflow by reducing or eliminating the polarization field in the active region, such as using polarization-matched multiple quantum wells (MQWs) (Schubert et al. 2008; Kuo et al. 2009), staggered InGaN quantum wells (Arif et al. 2007), and nonpolar or semipolar GaN substrate (Ling et al. 2010). But for improving hole distribution, only several approaches, such as p-type MQWs (Xie et al. 2008) or coupled quantum wells (Ni et al. 2008), are explored. However, in the p-type MQWs, the Mg dopant is very likely to diffuse into wells, while in the coupled quantum wells, electrons are tending to overflow by using thin barriers. These will result in reduction of radiative efficiency. In order to

improve the radiative efficiency, some approaches will be proposed in our study, including MQWs and EBL design, and will be studied and described in detail in this chapter. On the other hand, although the emission efficiency of GaN-based LEDs can be improved continuously, the cost of GaN-based LED devices had become the critical factor in recent years that determines their use in the general field. Many studies of the growth of InGaN-based epilayers on a variety of substrates, such as sapphire (Nakamura et al. 1994; Nakamura 1998; Jeong et al. 2010; Chiu et al. 2008a), SiC (Wetzel et al. 1994; Zehnder et al. 2001; Akasaka et al. 2004), freestanding GaN (FS-GaN) (Fang et al. 2012; Cao et al. 2004b; Chao et al. 2011), and Si (Dadgar et al. 2000, 2003a; Butter et al. 1979; Zhang et al. 2000; Krost and Dadgar 2002; Feng et al. 2002; Gong et al. 2003); Lu et al. 2003; Uen et al. 2005), have been published. Typically, InGaN-based LEDs are grown on sapphire using hetero-epitaxial techniques, such as metal-organic chemical vapor deposition (MOCVD) (Goldenberg et al. 1993; Liu et al. 2002; Davis et al. 2007). However, its low thermal conductivity and insulating properties make sapphire an imperfect substrate for InGaN-based epilayers. A high price and some mechanical defects reduce the acceptability of SiC as a substrate in the LED market. Recently, many works have used FS-GaN as a substrate for the epitaxy of InGaN-based LEDs. Although InGaN-based LEDs on FS-GaN substrate have a high emission efficiency and low droop, FS-GaN is too expensive, preventing end users from considering the purchase of LED devices that are grown on FS-GaN substrate. Among the aforementioned substrate materials, silicon (Si) is regarded as a relatively promising substrate for use in InGaN-based epitaxy because it has two advantages – a low manufacturing cost and the ability to form large size substrate (up to 8–12 in in diameter). The evolution of Si in the semiconductor industry is long: almost defect-free Si substrates can be fabricated, and these Si have many advantages over III–Vs, such as high heat dissipation, strength and difficulty of breaking, and ease of purchase. Additionally Si-based microelectronics can be integrated with InGaN-based optoelectronics.

In this chapter, we divided our chapter into three sections. Section “[Improvement of Emission Efficiency and Droop by MQWs and EBL Design](#)” is related to the studies on the improvement of emission efficiency and droop by MQWs and EBL design, while section “[Improvements of the GaN-Based LEDs Main Material Qualities](#)” is concerned with the studies on the improvements of the GaN-based LEDs’ main material qualities. Finally, section “[Low-Cost and High-Efficiency GaN-Based LEDs on Large-Area Si Substrate](#)” is the study on the epitaxial growth of GaN-based LEDs on Si substrate for the development of low-cost and high-efficiency GaN-based LED devices.

Growth and Characterization Methods of Gallium Nitride (GaN)-Related Materials

The epitaxial structures of InGaN-based LEDs were grown by using a low-pressure metal-organic chemical vapor deposition (LP-MOCVD) system with a water-cooled vertical (or horizontal) reactor. The substrate susceptor is made of graphite, coated

with a SiC film on the top surface by the CVD technique. Additionally, the metal-organic sources are stored in bubblers and brought out by carrier gas (typically H_2 or N_2). The composition and growth rate of GaN-based epilayer were precisely controlled by mass flow controller (MFC). The vapor pressure of all metal-organic sources can be precisely controlled by modulating the temperature of bubbler. All samples were grown on a 2-in (0001) sapphire and 6-in (111) Si substrate. The substrates used were placed on a susceptor, which were thermal-resistently heated or radiantly heated by a strip heater or heated by a radio-frequency (RF) coil. The metal-organic compounds of trimethylgallium (TMGa), trimethylaluminum (TMAI), and trimethylindium (TMIIn) were employed as the reactant source materials for Ga, Al, and In, respectively. Group V sources used are most commonly gaseous hydrides; for example, ammonia (NH_3) is used for the growth of nitrides. Besides, Silane (SiH_4) and bis-cyclopentadienyl magnesium (Cp_2Mg) were used as the sources for n-type and p-type dopants, respectively.

Following the epitaxial growth, the surface morphology of specimens was observed by atomic force microscopy (AFM) with a scanning area of $10 \times 10 \mu\text{m}^2$. The crystalline quality and interface of the epitaxial structures herein were evaluated by high-resolution double crystal X-ray diffraction (HRDCXD D8) with Cu $\text{K}\alpha$ radiation as the X-ray source ($\lambda = 1.54056 \text{ \AA}$). The distribution and threading behaviors of dislocations in the epilayer were studied by transmission electron microscopy (TEM). The interfacial microstructures of the epilayer were observed by high-resolution TEM (HRTEM). Finally, the light-current-voltage (L-I-V) characteristics of all packaged LED chips were measured at room temperature in continuous-wave (CW) mode. APSYS software, which was developed by Crosslight Software Inc., was used to determine the physical origin of the improvement in the efficiency of the InGaN-based LEDs on sapphire and Si.

Improvement of Emission Efficiency and Droop by MQWs and EBL Design

Wider InGaN Quantum Well

The InGaN-based LEDs were grown on *c*-plane sapphire substrates by low-pressure metal-organic chemical vapor deposition (MOCVD). On top of the sapphire substrate, a 20-nm-thick GaN nucleation layer was grown at low temperature followed by a 4- μm n-type GaN buffer layer. After the growth of the buffer layer, a ten-period InGaN/GaN multiple quantum well (MQW) was grown. A 10-nm p-type AlInGaN layer was grown on top of the MQWs and followed by two sets of p-type AlGaIn/GaN (10/2 nm) electron-blocking layers (EBLs). Finally, a heavily Mg-doped p-type GaN contact layer (~120 nm) was grown. Subsequently, the LED mesa with an area of the $350 \times 350 \mu\text{m}^2$ was defined by using standard photolithography and dry etching. In addition, a transparent conduction

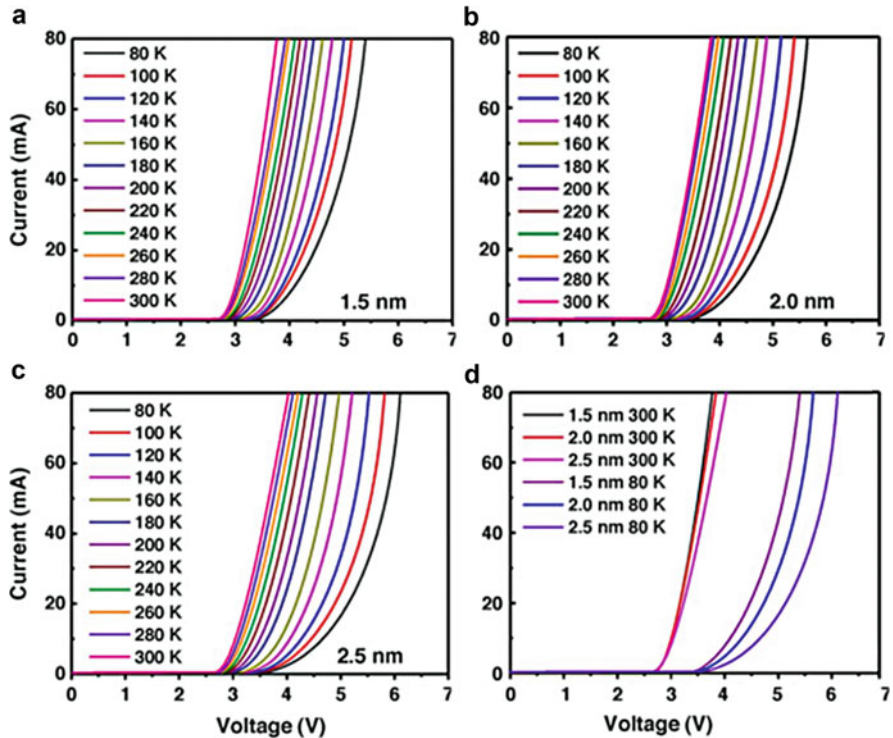


Fig. 1 Measured I - V characteristics of the InGaN LEDs with different well widths of (a) 1.5, (b) 2.0, and (c) 2.5 nm under different temperatures. The I - V curves at high and low temperature are further summarized in (d)

indium tin oxide (ITO) layer was employed to be the p-type ohmic contact layer. Ni/Au and Ti/Al/Ti/Au metallizations are deposited by e-beam evaporation to serve as p-type and n-type ohmic contacts, respectively. The QW thicknesses of the InGaN LEDs were controlled by changing the growth time of 1 min (1.5 nm), 1.33 min (2.0 nm), and 1.67 min (2.5 nm). The LED chips were mounted onto TO-46 lead frames without epoxy encapsulation.

Figure 1 shows the measured I - V characteristics of the InGaN-based LEDs with different well widths of (a) 1.5, (b) 2.0, and (c) 2.5 nm under different temperatures. The forward voltage of the InGaN-based LEDs increases with decreasing temperature, which mainly originates from the reduction of hole mobility and the difficulty in activating hole concentration due to the high activation energy of Mg dopants. For clearly comparing the I - V curves of the LEDs with different well widths at high and low temperatures, we further summarized the results at 80 K and 300 K, respectively, in Fig. 1d. It is noteworthy that the necessary forward voltage significantly increases

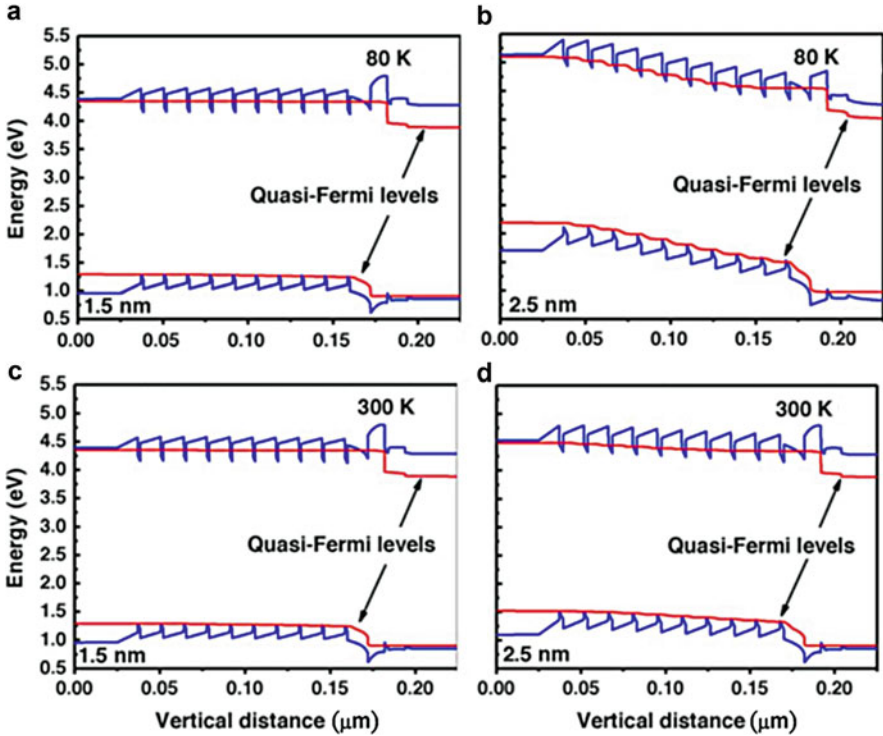


Fig. 2 Calculated band diagrams under an injection current of 80 mA for the LEDs with 1.5- and 2.5-nm wells at 80 K and 300 K

from 300 to 80 K by nearly 1.5–2.0 V with increasing quantum well (QW) thickness when the injection current is 80 mA. Previously, Cao and LeBoeuf demonstrated that when the temperature decreases from room temperature (300 K) to low temperature (80 K), the electron concentration and mobility of n-type GaN would be decreased by several times. However, the hole concentration and mobility of p-type GaN would be decreased by two orders and by about one order, respectively (Cao and LeBoeuf 2007). This dramatically reduced hole concentration and mobility at low temperature will result in significantly nonuniform carrier distribution within the ten-pair MQW-active region. Therefore, we deduced that the increased forward voltage with well width is resulted from the increased undoped region and the difficulty of hole transport between each QW at low temperature.

To investigate the carrier transport within MQW region, simulations of the band diagrams and carrier distributions were performed by employing the APSYS modeling software. Figure 2 shows the calculated band diagrams under an injection current of 80 mA for the LEDs with 1.5- and 2.5-nm wells at 80 K and 300 K. Since the LED with thicker well (2.5 nm) requires higher forward voltage than that with thinner well (1.5 nm) under the same injection current at 80 K, the conduction band on the n-side

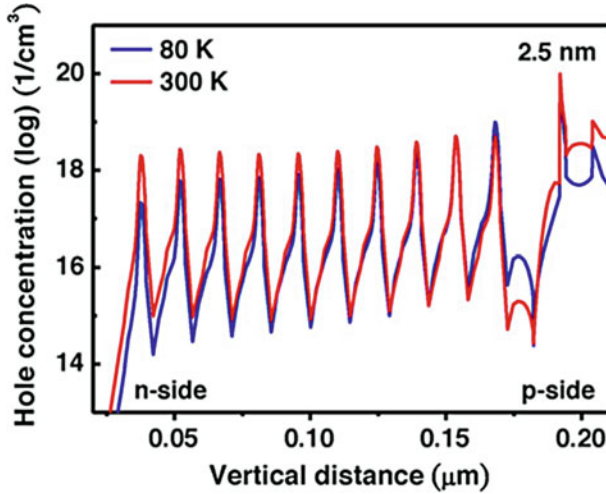


Fig. 3 Calculated hole concentration distributions under the injection current of 80 mA at 80 K and 300 K for the LED with 2.5-nm wells

of the LED is higher than that on the p-side, which favors the electrons to escape to the p-side of the LEDs from the observation of electron quasi-Fermi level in the EBL region, as shown in Fig. 2b. On the contrary, when the temperature increases to 300 K, the thermally activated hole concentration and thermally enhanced hole transport lead to the relatively minor difference in forward voltage between the LEDs with 1.5- and 2.5-nm wells, resulting in the relatively flat conduction band on the n-side and p-side, as shown in Fig. 2c, d. Figure 3 shows the calculated hole concentration distributions under the injection current of 80 mA at 80 K and 300 K for the LED with 2.5-nm well width. As can be seen from Fig. 3, the hole concentration in the QW nearest the p-side is about two orders of magnitude higher than that in the QW nearest the n-side when the temperature is 80 K. This nonuniform hole distribution will strongly enhance the electron overflow from the QWs as well. When the temperature is increased to 300 K, the relatively uniform hole distribution within the MQW region can effectively enhance the radiative recombination in each QW. Consequently, the energy band diagram is strongly influenced by temperature-dependent forward voltage, which will play an important role in the temperature dependence of EL efficiency as a function of input current.

Figure 4 shows the measured results of external quantum efficiency (EQE) versus logarithmic input current at temperature between 80 and 300 K for the LEDs with different well widths. The maximum EQE is normalized in Fig. 4. It can be found that the maximum EL efficiency shifts toward lower injection current with decreasing temperature. When the temperature is 80 K, the rapid efficiency droop is induced from the severely nonuniform hole distribution, as shown in Fig. 3. Under the condition of low temperature and low current injection (~ 0.1 mA), holes are confined in the QW nearest the p-side, which results in the effective recombination

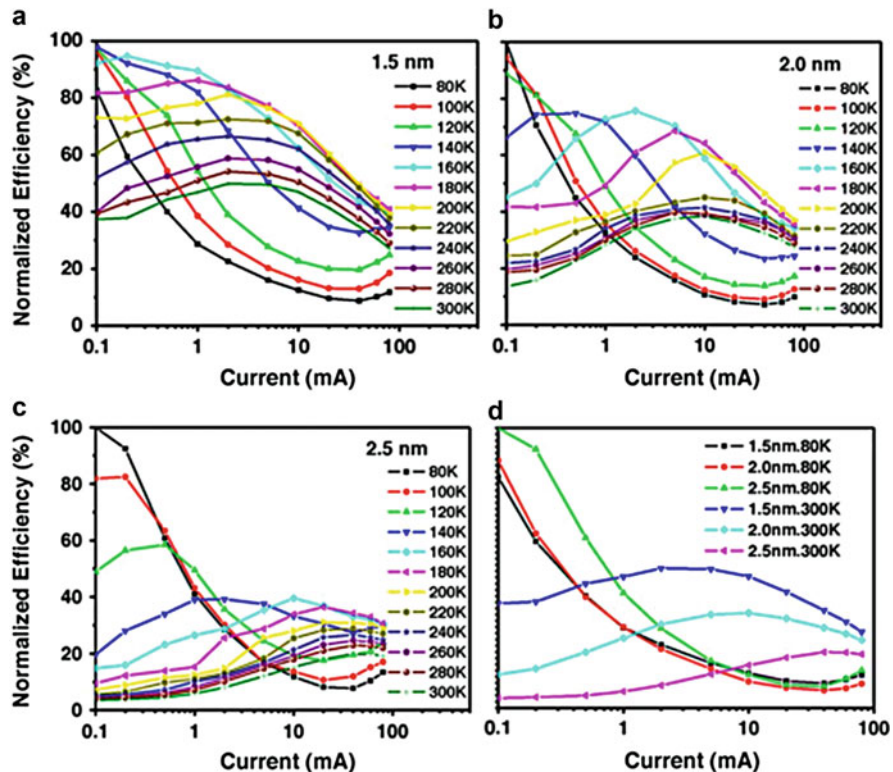


Fig. 4 Measured normalized EQE of the blue InGaN LEDs with different well widths as a function of input current at temperature between 80 and 300 K

in this QW due to the relatively easy electron transport through the MQW region. Nevertheless, with increased injection current, holes are still confined in this QW due to the low mobility at low temperature, and electrons seriously overflow into the p-side since the increase of injection current requires large forward voltage at low temperature, leading to the corresponding change in energy band structure, as shown in Fig. 2b. The significant nonuniform hole distribution and electron overflow cause the rapid EL efficiency droop with increasing input current at low temperature. Furthermore, in the case of high temperature (300 K), the EL efficiency rises at first and then decreases with the increasing injection current. This is quite different as compared with the case at low temperature. As we know, the efficiency would be increased with decreasing temperature because of the suppression of nonradiative recombination centers. However, from simulation results of Fig. 3, more severe nonuniformity of hole distribution at MQWs at low temperature reduced the luminescence efficiency, especially at higher injection level. Suppression of nonradiative centers and poorer hole distribution phenomenon will compete with each other, and therefore, a maximum efficiency would be observed at each temperature.

To further study the temperature-dependent carrier transport in the LEDs with different well widths, we summarized the EL efficiency at 80 K and 300 K in Fig. 4d. It is obvious that the EL efficiency curves nearly follow the same droop behavior at low temperature for the three LED structures. The EL efficiency of the LEDs with different well widths has a minimum value at 80 K as the input current is about 40 mA. It means that the droop behavior is dominated by the low hole mobility and is independent on the QW thickness. This phenomenon also implies that the efficiency droop is less sensitive to the Auger recombination since the droop behavior is obviously observed at low injection current. The distinct EL efficiency as a function of input current is found at 300 K for the three LEDs. A much lower efficiency droop is observed for the LED with 2.5-nm QW with increasing injection current. However, the LED with thicker QW gives rise to lower EL efficiency, as shown in Fig. 4d. It is attributed to spatial separation of electrons and holes induced by the stronger internal polarization field. Furthermore, to achieve maximum EL efficiency, the required input current values are very different for the three LED structures. The input current values achieving maximum EL efficiency are around 4, 10, and 60 mA for the LEDs with 1.5-, 2.0-, and 2.5-nm QWs, respectively. The efficiency droop from the maximum value is dominated by the effect of electron overflow at 300 K. The LED with 2.5-nm QWs can effectively confine electrons within the MQWs and suppress the electron overflow as compared with the LED with thinner QWs, which causes the maximum EL efficiency to shift toward higher injection current with the increased QW thickness.

In summary, we have presented the temperature-dependent efficiency droop in blue InGaN LEDs with different well widths. At low temperature, the droop behavior is dominated by the low hole mobility and is nearly independent on the QW thickness. However, at high temperature, the input current values achieving maximum EL efficiency are around 4, 10, and 60 mA for the LEDs with 1.5-, 2.0-, and 2.5-nm QWs, respectively. This difference of the efficiency droop at high and low temperature is closely relevant to the effects of carrier transport and escape processes from the active region and is less sensitive to the Auger recombination.

Graded-Thickness Multiple Quantum Wells (GQWs)

InGaN-based LEDs with graded-thickness multiple quantum wells (GQWs) were designed and grown on c-plan sapphire by metal-organic chemical vapor deposition. A 20-nm-thick low-temperature GaN nucleation layer followed by a 4 μm n-type GaN buffer layer and a ten-pair InGaN/GaN superlattice were grown on the top of sapphire, respectively. After that, six-pair MQWs were grown with 10-nm-thick GaN barriers. For our designed experiment, the thicknesses of In_{0.15}Ga_{0.85}N quantum wells for GQW LED structure, controlled by growth time, are 1.5, 1.8, 2.1, 2.4, 2.7, and 3 nm along the [0001] direction, while the reference LED structure has a unique well thickness of 2.25 nm. It's worth noting here that the total volumes of active region for the two samples are the same. Finally, a 20-nm-thick electron-blocking layer with Al_{0.15}Ga_{0.85}N and a 200-nm-thick p-GaN layer were grown to

complete the epi-structure. For EL measurements, the LED chips were fabricated by regular chip process with ITO current spreading layer and Ni/Au contact metal, and the size of mesa is $300 \times 300 \mu\text{m}^2$.

It has been reported that, with the same indium content, wider well has longer radiative recombination lifetime (Sun et al. 1997a; Charash et al. 2009). In our designed GQWs, the well thickness gradually increases along the [0001] direction. Therefore, one can expect that the holes in wider well tend to escape to the next narrower well before they radiatively recombine with electrons, leading to the hole concentrations decrease in the wider well, but increase in the narrower wells. In other words, the hole distribution will be improved. To prove the above hypothesis, we investigate the carrier distribution of both GQW and reference LED structures mentioned above by APSYS simulation.

Based on our experimental structures, we built up the model of the reference and GQW LED structures. The typical LED structure was composed of 4- μm -thick n-type GaN layer (n-doping = $2 \times 10^{18} \text{ cm}^{-3}$), six pairs of $\text{In}_{0.15}\text{Ga}_{0.85}\text{N}/\text{GaN}$ MQWs with 10-nm-thick GaN barriers, 20-nm-thick p-Al_{0.15}Ga_{0.85}N electron-blocking layer (p-doping = $5 \times 10^{17} \text{ cm}^{-3}$), and 200-nm-thick p-type GaN layer (p-doping = $1 \times 10^{18} \text{ cm}^{-3}$). Other material parameters of the semiconductors used in the simulation can be found in Bernardini (2007). Commonly accepted Shockley–Read–Hall recombination lifetime (several nanoseconds) and Auger recombination coefficient ($2 \times 10^{-30} \text{ cm}^6/\text{s}$) are used in the simulations.

Figure 5 shows the simulated hole distribution and radiative recombination distribution along MQWs at $100 \text{ A}/\text{cm}^2$. For reference LED structure, it can clearly be seen that holes mostly concentrate in the QW nearest p-side (denoted as the first QW), so does the radiative recombination. This phenomenon coincides with the optical measurement result in David et al. (2008), which is mainly due to poor transportation of holes. While in the case of GQW LED structure, the hole concentration decreases in the first QW by about 16 %, but increases in the second, third, and fourth QWs by 7 %, 94 %, and 175 %, respectively, as compared with reference LED. It indicates that the holes are more capable of transporting across the first QW, consisted with our hypothesis. On the other hand, electrons are relatively not being affected due to their high mobility. Therefore, more wells will participate in the recombination process, as illustrated by the radiative recombination distribution in Fig. 5b. Accordingly, the simulated EL spectrum of GQW LED at current density of $100 \text{ A}/\text{cm}^2$ exhibits larger full width at half maximum (FWHM) than that of reference LED, as shown in the inset of Fig. 5b. Moreover, due to the relative low carrier densities in the first QW and more uniform of carrier distribution, the possibility of Auger scattering and carrier overflow can be lower. And the alleviation of efficiency droop can be expected.

Figure 6 shows the power-dependent EL average wavelength and FWHM of reference and GQW LED at room temperature. The EL measurement was performed by on-wafer probing with a spectrometer. The emission wavelength (457.7 nm at $1 \text{ A}/\text{cm}^2$) and FWHM (21.9 nm at $1 \text{ A}/\text{cm}^2$) for GQW LED are larger than those for reference LED (448.5 nm and 17.9 nm), respectively. It could be due to the graded-thickness and wider wells near to p-side in GQW. Besides, as increasing the injection

Fig. 5 Simulated (a) hole distribution and (b) radiative recombination distribution in reference and GQW LEDs

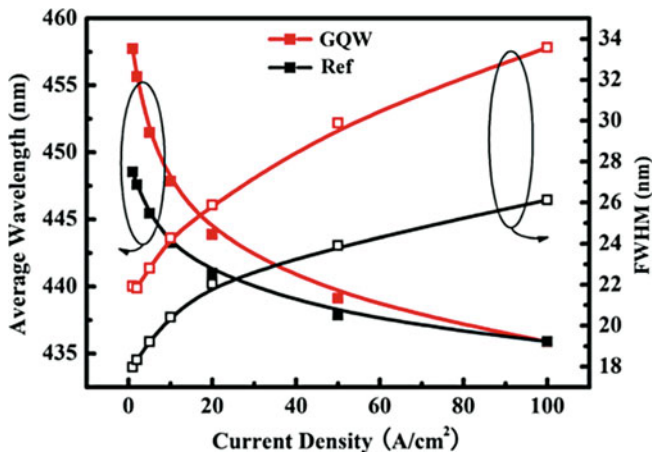
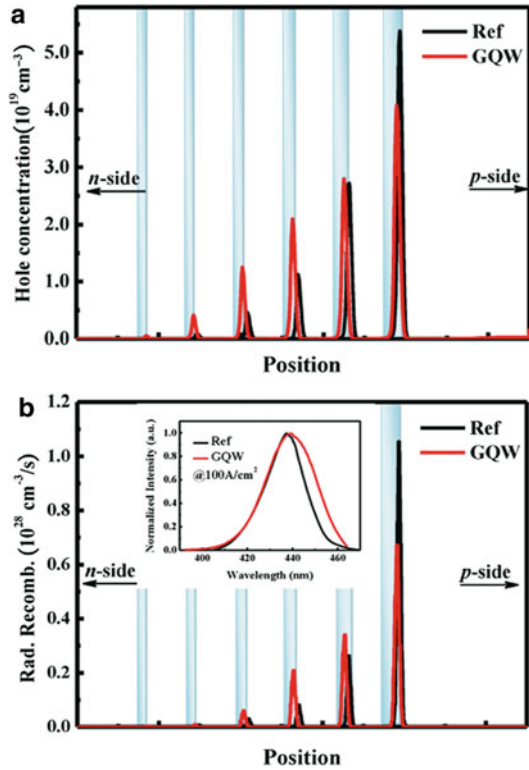


Fig. 6 Average wavelength and FWHM as a function of current density for reference and GQW LEDs

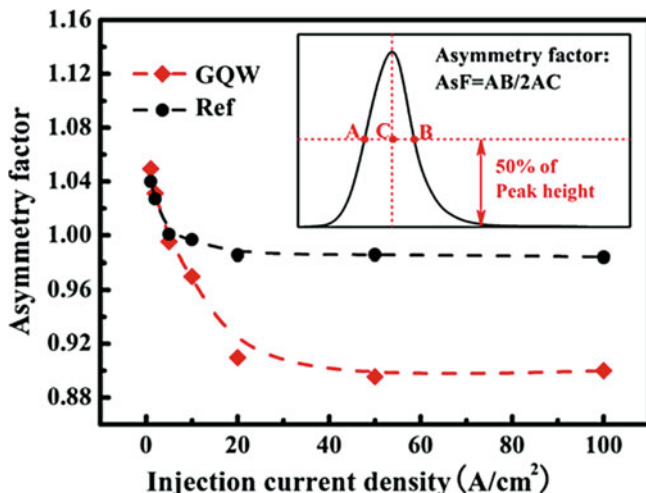


Fig. 7 Current-dependent asymmetry factor of EL spectra of reference and GQW LEDs

current from 1 to 100 A/cm², EL spectrum for GQW LED exhibits significant blueshift of 21.8 nm and broadening of about 11.6 nm, compared with 12.6 nm and 8.2 nm, respectively, for reference ED. Generally, the blueshift of the GaN-based LED can be attributed to the band-filling effect in localized states and the charge screening effect of quantum-confined Stark effect (QCSE) (Takeuchi et al. 1997). And the broadening of FWHM is mainly due to the band-filling effect and self-heating effect. In GQW and reference LEDs, the band-filling effect and self-heating effect can be considered to be equivalent because they have the same indium content and total volumes of active region. Thus, there must be other reasons for such significant blueshift and broadening of EL spectra in GQW LED.

According to the simulated results mentioned above, more holes distribute in the narrower wells in GQW LED structure. Once more carriers radiatively recombine in narrower wells, the intensity of shorter-wavelength part in emission spectrum will rise. Thus, changes in symmetry of spectrum could be expected. To investigate the symmetry of EL spectrum in detail, the asymmetry factor (AsF) was calculated. As illustrated in the inset of Fig. 7, it can be defined as the distance from the center line of the peak to the back slope (AB) divided by twice the distance from the center line of the peak to the front slope (2 AC), with all measurements made at the positions determined by 50 % of the peak height. The calculated AsF under each injection level for both samples is summarized in Fig. 7. It can be clearly seen that the AsF of the reference LED decreases slightly from 1.04 to about 0.98 when injection current increases from 1 to 100 A/cm². While GQW LED shows a larger variation, the AsF starts at 1.05 (0.1 A/cm²) and saturates at about 0.89 (after 20 A/cm²). According to the definition of AsF, if the bluer light emits from the narrower well, the symmetry of spectrum would be interrupted to make AsF become smaller than 1. Therefore, one can conclude that GQW does have superior radiative recombination distribution,

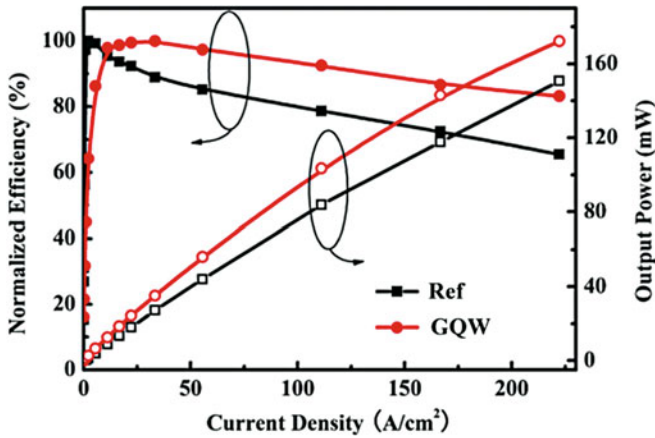


Fig. 8 Comparison of normalized EL efficiency and L-I curves

which leads to significant blueshift and broadening for the EL spectrum with increasing injection current. These enormous changes in wavelength and line width might make the design of the GQW concept impractical for lighting applications. In the future, we will optimize the GQW structure, such as appropriately reducing the indium content of the wider wells, to alleviate these effects.

Finally, we investigated the efficiency droop behaviors in both LEDs. The output powers measured with a calibrated integrating sphere and the normalized efficiency (η) of reference and GQW LED are plotted in Fig. 8 as a function of injection current density. The light output power of GQW LED is found to be enhanced by 35 % at 20 A/cm², as compared with the reference LED (24.3 vs. 18.0 mW). This indicates that even with wider wells (worse wave function overlap for electrons and holes) near p-side, the overall efficiency for GQW LED is still higher than reference, and the utilization rate of MQWs is improved. More importantly, the maximum efficiency (η_{peak}) of GQW LED appears at injection current density of 30 A/cm², which is much higher than that for reference LED (at 2 A/cm²). And the efficiency droop, defined as $\left(\eta_{\text{peak}} - \eta_{200 \text{ A/cm}^2} / \eta_{\text{peak}}\right)$ is alleviated from 32 % in reference LED to 16 % in GQW LED. This improvement could be mainly attributed to the superior hole distribution and radiative recombination distribution and also the reduction of Auger scattering resulting from the lower carrier concentration in QW nearest p-side.

In conclusion, InGaN/GaN LEDs with graded-thickness multiple quantum wells were investigated both experimentally and numerically. The APSYS simulations indicate that superior hole distribution can be achieved in the GQW-designed MQWs, in which the well thickness increases along the [0001] direction. It might be attributed to the longer radiative recombination lifetime in the wider well nearest to p-type layer. Moreover, by analyzing the EL spectra in detail, the additional emission from the narrower wells was demonstrated. This indicates that more carriers distribute in the previous wells, which agrees well with the simulated results.

As a result, the efficiency droop behavior was alleviated from 32 % in reference LED to 16 % in GQW LED. In addition, the light output power was enhanced from 18.0 to 24.3 mW at 20 A/cm² compared to reference LED with the same active volume. This work implies that with suitable active region design, carrier transportation behavior could be modified, which is very useful for alleviating efficiency droop.

Graded-Composition Multiple Quantum Barriers (GQB)

In this study, we report a new design of the barrier layers in MQWs by grading the composition of barriers from In_xGa_{1-x}N to GaN along the [0001] direction, to form a graded-composition multiple quantum barriers (GQB) and show the improvement in transport of holes in active region and substantial reduction in efficiency droop behavior. The injected holes do not accumulate at the well closest to p-GaN, denoted as the last well, and uniformly spread in the active region. The droop behavior predicted by our simulation agrees well with the experiments.

For conventional LEDs operated under forward bias, the band diagram of multiple quantum barriers (MQBs) shows triangular shape due to the internal polarization field and forward bias (Ling et al. 2010), as shown in Fig. 9. The valance band of MQBs shows an upward slope from the n-GaN side toward p-GaN side, which retards the holes to travel across the triangular barrier. But if the composition of indium in barriers decreases from the n-GaN side toward p-GaN side, the bandgap broadens gradually. As a result, the barrier in valance band could be leveled down and even overturned, while the slope of conduction band could be enhanced. This could enhance the hole transporting across the barriers.

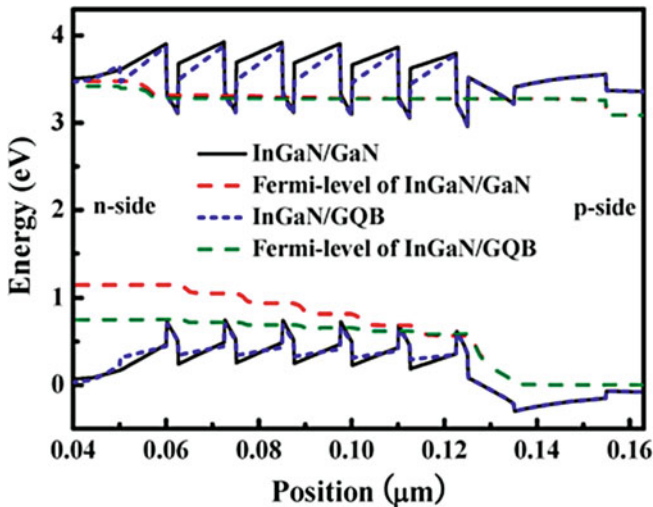


Fig. 9 Calculated band diagrams of the conventional and GQB LEDs

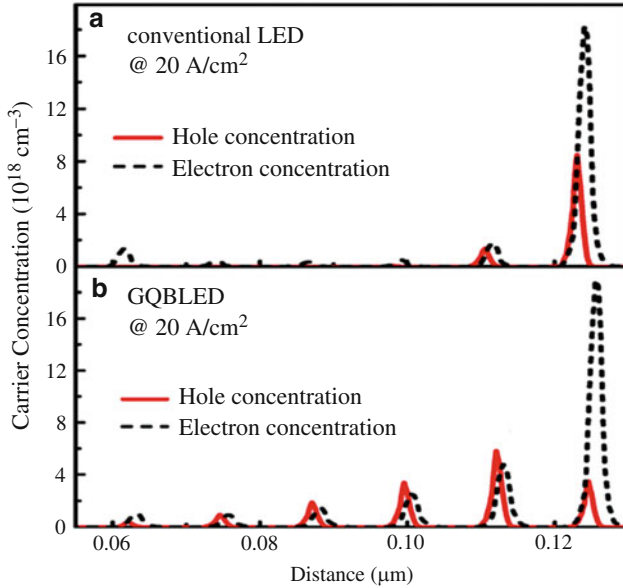


Fig. 10 Calculated carrier concentrations at current density of 20 A/cm² for (a) conventional and (b) QQB LEDs

We then simulated the band diagrams and carrier distributions in QQB LED using APSYS simulation software. The simulation LED structures were composed of 4- μm -thick n-type GaN layer (n-doping = $2 \times 10^{18} \text{ cm}^{-3}$), six pairs of In_{0.15}Ga_{0.85}N/GaN MQWs with 2.5-nm-thick wells and 10-nm-thick barriers, 20-nm-thick p-Al_{0.15}Ga_{0.85}N EBL (p-doping = $5 \times 10^{17} \text{ cm}^{-3}$), and 200-nm-thick p-type GaN layer (p-doping = $1 \times 10^{18} \text{ cm}^{-3}$). For the QQB LED, the composition of indium was graded from 5 % to 0 % along the [0001] direction and compared with the conventional LED with GaN barriers. Commonly accepted physical parameters were adopted to perform the simulations, the percentage of screening effect of 50 %, the Shockley–Read–Hall recombination lifetime of 1 ns, and the Auger recombination coefficient in quantum wells with order of $10^{-31} \text{ cm}^6/\text{s}$, respectively (Piprek 2007). Other material parameters used in the simulation can be referred to Vurgaftman and Meyer (2003). The calculated band diagrams of conventional and QQB LEDs at current density of 100 A/cm² are shown in Fig. 1. The triangular barriers in valance band are partially leveled in QQB LED. Unexpectedly, the quasi-Fermi level for holes in QQB LED is lower than that of conventional LED. This phenomenon could further favor the transport of holes in active region.

Figures 10 and 11 show the calculated carrier concentration in the active region for conventional and QQB LEDs at current density of 20 and 200 A/cm², respectively. At low current density of 20 A/cm², holes appear to hop out of the last well and spread to the others for QQB LED, as shown in Fig. 10b. The hole concentrations in the last well are about 8.5×10^{28} and $3.5 \times 10^{28} \text{ cm}^{-3}$ for conventional and

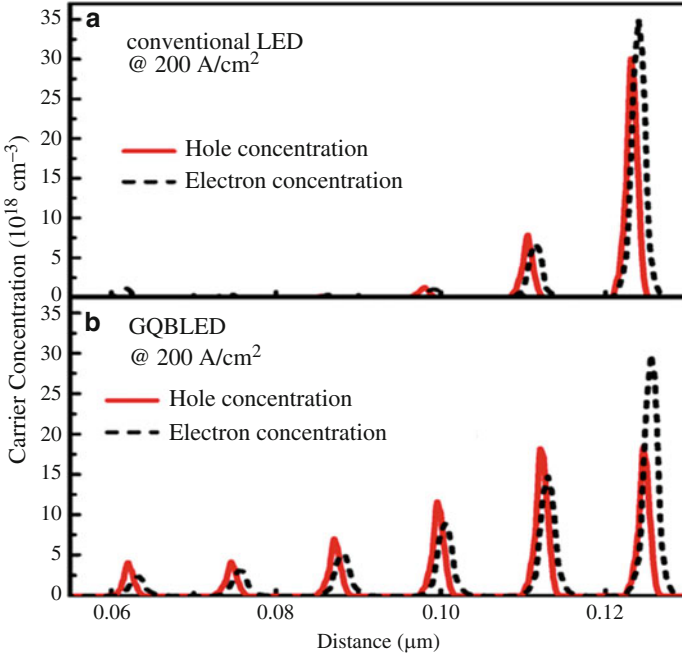


Fig. 11 Calculated carrier concentrations at current density of 200 A/cm^2 for (a) conventional and (b) QQB LEDs

GQB LEDs, respectively, while they are about 1.3×10^{28} and $5.8 \times 10^{28} \text{ cm}^{-3}$ in the fifth well for conventional and GQB LEDs, respectively. These results indicate that GQB LED has better hole transport even at low current density, lowering the hole concentration at the last well. At high current density of 200 A/cm^2 , hole distribution in GQB LED is more uniform than that at low current density. Such improved hole distribution is useful for droop reduction (Ni et al. 2008). Meanwhile, the calculated electron concentration in the active region for conventional and GQB LEDs was shown in Figs. 10 and 11. At either low or high current density, electron concentration in GQB LED is higher than that in conventional LED, which could be attributed to lower electron overflow for GQB LED. The distribution of electrons in GQB LED seems to be more uniform than that in conventional LED, which might correspond to better transport of holes. From the simulation results, hole and electron distributions in GQB LEDs, either at low or high current density, are favorable for reduction of droop behavior.

In light of simulation results, we have grown the LED structures with GaN barriers and GQB on c-plane sapphire substrates by metal-organic chemical vapor deposition (MOCVD). After depositing a low-temperature GaN nucleation layer, a $4\text{-}\mu\text{m}$ n-type GaN layer, and a ten-pair InGaN/GaN superlattice prestrain layer, the rest of the LED structures were grown based on our simulation design. The graded-composition barriers were grown using In/Ga ratio ramping to prevent the change in

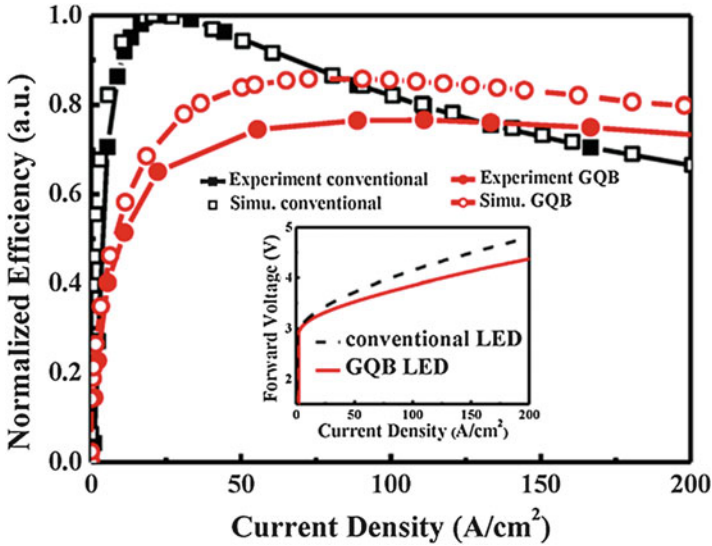


Fig. 12 Experiment and simulation normalized efficiency for conventional and QQB LEDs. The inset figure shows I–V characteristics of conventional and QQB LEDs

the growth rate. Finally, the LED chips were fabricated by regular chip process with ITO current spreading layer and Ni/Au contact metal. The LED has a typical chip size of $300 \times 300 \mu\text{m}^2$. The emission wavelengths of both types of LEDs were around 450 nm at 22 A/cm^2 .

The I–V characteristics of conventional and QQB LEDs are shown in the inset of Fig. 12. The series resistance is reduced from 8.2Ω in conventional LED to 6.5Ω in QQB LED, which indicates a certain degree of the improvement in hole transport. As a result, the forward voltage at 22 A/cm^2 is reduced from 3.4 V in conventional LED to 3.27 V in QQB LED. The normalized efficiency of conventional and QQB LEDs as a function of current density was investigated by experiment and simulation, as illustrated in Fig. 12. It can be seen that the experimental data have similar droop behavior to simulation results. However, the efficiency of QQB LED shows slightly lower value in experiment. This can be attributed to nonoptimized epitaxial parameters for graded-composition barriers. The most important result is the reduction of efficiency droop, defined as $(\eta_{\text{peak}} - \eta_{200 \text{ A/cm}^2})/\eta_{\text{peak}}$, which is reduced from 34 % for conventional LED to 6 % for QQB LED. This result confirms that the graded-barrier design did contribute to the reduction of efficiency droop.

Although the efficiency droop has been reduced, the efficiency of QQB LED at 20 A/cm^2 , which is the typical operation current density for most LEDs, is only 70 % of that of conventional LED. This phenomenon has also been observed in other droop-reduction methods related to the layer design of MQBs (Ni et al. 2008) or low polarization field structures (Schubert et al. 2008; Ling et al. 2010). To understand this, the radiative recombination distribution in the active region for both LEDs was

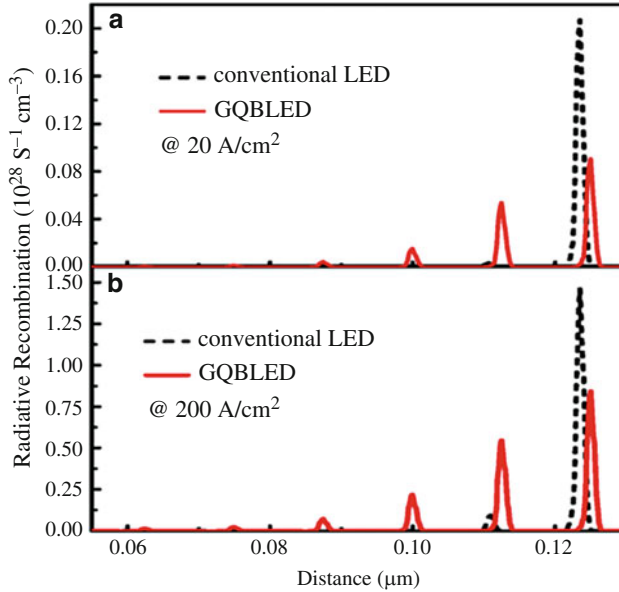


Fig. 13 Calculated radiative recombination of conventional and GQB LEDs at current density of (a) 20 A/cm² and (b) 200 A/cm²

calculated and illustrated in Fig. 13. The results show that the radiative recombination distribution in GQB LED is more uniform than that in conventional LED, at current density of 20 A/cm²; however, the total amount of radiative recombination in GQW LED is only 70 % of that in conventional LED. At 200 A/cm², the total amount of radiative recombination in GQW LED is 119 % of that in conventional LED. The reason could be referred to the poor spatial distribution overlap between holes and electrons. For good hole transport as GQB LED, the spatial distributions of holes and electrons are quite different. As shown in Figs. 10 and 11, at current density of 20 A/cm², most of the electrons still concentrate in the last well, while the hole concentration in the last well is less than that in the fifth well. That is, the spatial overlap between electrons and holes in GQB LED is not the optimal situation at low current density. While at 200 A/cm², such mismatch is alleviated due to more holes are injected. However, in conventional LED, both holes and electrons concentrate at the wells near p-GaN, so the radiative recombination is quite effective at that location. These results indicate that to reduce droop behavior without deteriorating the total recombination, one should pay more attention to the spatial distribution between holes and electrons.

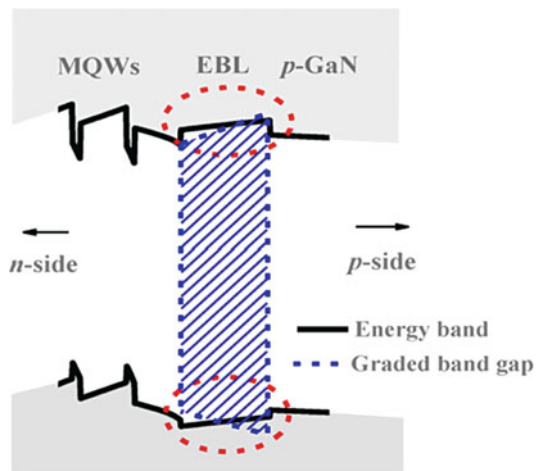
In summary, we have designed a graded-composition multiple quantum barriers for InGaN/GaN LED to improve the hole transport in active region. The simulation results showed that the triangular barrier of multiple quantum barriers at valance band could be balanced by increasing the bandgap of In_xGa_{1-x}N along the [0001] direction. As a result, the hole transport in MQWs was significantly enhanced at

either low or high current density, which is beneficial for droop reduction. The GQB LED was realized by MOCVD, and the I–V curves showed that GQB LED has lower series resistance than the conventional one, and the efficiency droop was reduced from 34 % in conventional LED to only 6 % in GQB LED, which is in agreement with our simulation results. Beyond these, the efficiency of GQB LED at 20 A/cm² was found to be only 70 % of that in conventional LED. The reason for such low recombination could be attributed to the poor spatial distribution overlap between holes and electrons. These results indicate that although the improvement in hole transport facilitates the reduction of efficiency droop, spatial distribution between electrons and holes should be taken into consideration.

Graded-Composition Electron-Blocking Layer (GEBL)

In this study, we designed a graded-composition electron-blocking layer (GEBL) for InGaN/GaN LEDs employing the concept of band engineering, which not only suppressed the electron overflow out of active region but also enhanced the hole injection. The improvements in electron confinement and hole injection of LED with GEBL were demonstrated in simulation. Then it was realized by using metal-organic chemical vapor deposition (MOCVD), and the efficiency droop in LED with GEBL was found to be much smaller than that in conventional LED with constant-composition Al_xGa_{1-x}N EBL. The concept of band engineering started from the observation on the band diagram of InGaN-based LEDs. For conventional LEDs operated under forward bias, the band diagram of EBL shows a triangular shape due to the internal polarization field and forward bias (Arif et al. 2007), as shown in Fig. 14. The valence band of electron-blocking layer (EBL) slopes upward from the n-GaN side toward the p-GaN side, which retards the holes to transport across the triangular barrier. But if the composition of aluminum in EBL increases from the

Fig. 14 Schematic diagram of the concept of band engineering at EBL



n-GaN side toward the p-GaN side, the bandgap broadens gradually. As a result, the barrier in the valence band could be leveled down and even overturned, while the slope of the conduction band could be enhanced. Then, the improvement in the capability of hole transportation across the EBL as well as the electron confinement could be expected.

To prove the feasibility of the hypothesis above, the band diagrams and carrier distributions in LED with GEBL were investigated first by APSYS simulation program. The simulation LED structures were composed of 4- μm -thick n-type GaN layer (n-doping = $2 \times 10^{18} \text{ cm}^{-3}$), six pairs of $\text{In}_{0.15}\text{Ga}_{0.85}\text{N}/\text{GaN}$ multiple quantum wells (MQWs) with 2.5-nm-thick wells and 10-nm-thick barriers, 20-nm-thick p- $\text{Al}_x\text{Ga}_{1-x}\text{N}$ EBL or GEBL (p-doping = $5 \times 10^{17} \text{ cm}^{-3}$), and 200-nm-thick p-type GaN layer (p-doping = $1 \times 10^{18} \text{ cm}^{-3}$). For the LEDs with GEBL, three types of GEBLs with compositions of aluminum graded along the [0001] direction from 0 % to 15 %, 25 %, and 35 %, respectively, were simulated and denoted as LEDs A, B, and C. Furthermore for the conventional LED, the composition of aluminum was a constant of 15 %. Commonly accepted physical parameters were adopted to perform the simulations, the percentage of screening effect of 50 %, the Shockley–Read–Hall recombination lifetime of 1 ns, and the Auger recombination coefficient in quantum wells of $2 \times 10^{-30} \text{ cm}^6/\text{s}$, respectively (Piprek 2007). Other material parameters used in the simulation can be referred to Vurgaftman and Meyer (2003).

Figure 15 shows the energy band diagrams of LEDs A, B, and C at current density of $100 \text{ A}/\text{cm}^2$. According to our concept of band engineering, the degree of gradation had the decisive influence on the capability of hole injection. Even with small degree

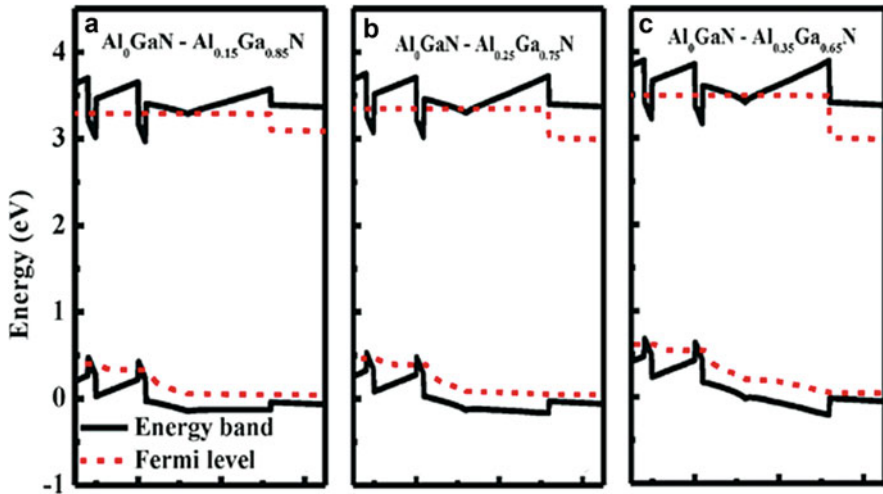


Fig. 15 Calculated energy band diagrams of (a) $\text{Al}_0\text{Ga}_1\text{N}$ to $\text{Al}_{0.15}\text{Ga}_{0.85}\text{N}$, (b) $\text{Al}_0\text{Ga}_1\text{N}$ to $\text{Al}_{0.25}\text{Ga}_{0.75}\text{N}$, and (c) $\text{Al}_0\text{Ga}_1\text{N}$ to $\text{Al}_{0.35}\text{Ga}_{0.65}\text{N}$ graded-composition EBLs at a current density of $100 \text{ A}/\text{cm}^2$

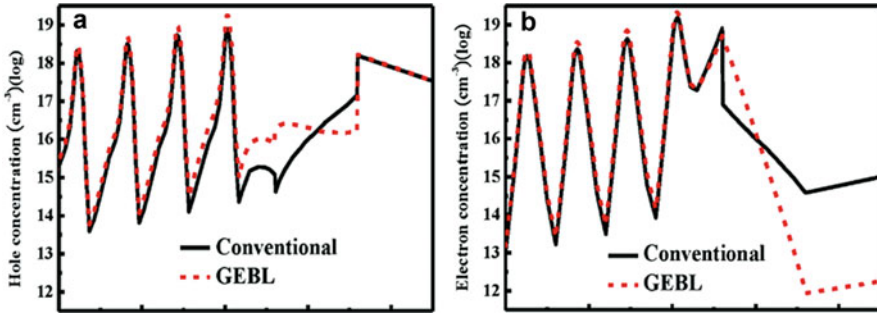


Fig. 16 Calculated (a) hole concentration distribution and (b) electron concentration distribution of conventional and GEBL LEDs at a current density of 100 A/cm^2

of gradation as LED A, the slope of the valence band can be leveled. Then the slope starts to overturn when the composition of aluminum at the p-side increases up to 25 %. Moreover, it is worth noting that the (ΔE_v) between the last GaN barrier and the EBL is diminished in all three LEDs with GEBL. Therefore, the hole injection can be improved effectively by using the GEBL. In the meantime, as the degree of gradation increased, the conduction band offset at the interface of p-GaN and EBL increases as well, so does the confinement capability of electrons. However, a corresponding increase in the (ΔE_v) between EBL and p-GaN with the composition of aluminum might retard the transportation of holes. In addition, high aluminum-composition EBL is not practical for actual application due to the low acceptor-activation efficiency and the low crystal quality in epitaxy (Katsuragawa et al. 1991). Consequently, only LED B with the composition of aluminum grading from 0 % to 25 % is discussed in the following paragraph.

The profiles of hole and electron concentration distribution at a current density of 100 A/cm^2 are illustrated in Fig. 16a, b, respectively. It can clearly be seen that with GEBL, injected holes uniformly distribute along the EBL region compared to conventional one, demonstrating that the flat valence band indeed favored the hole transportation across EBL. Meanwhile, the hole concentration in MQWs is significantly increased as expected. Moreover, the electron concentration in MQWs is also enhanced, while the electron distribution within the GEBL region and p-GaN is enormously decreased over two orders. This result indicates that GEBL can suppress the electron overflow out of active region more effectively than conventional EBL, even though the conduction band offset between the last GaN barrier and the GEBL is diminished.

Then, the LED structures with EBL and GEBL were grown on c-plane sapphire substrates by MOCVD. After depositing a low-temperature GaN nucleation layer, a $4\text{-}\mu\text{m}$ n-type GaN layer, and a ten-pair InGaN/GaN superlattice prestrain layer, the rest of the LED structures were grown based on our simulation design. The epitaxial recipe for the GEBL is worth noting. Generally, the composition-graded ternary III-nitride semiconductors can be grown by two methods: growth temperature ramping and III/III ratio ramping (Sun et al. 1997b; Kim et al. 2001a). Here we

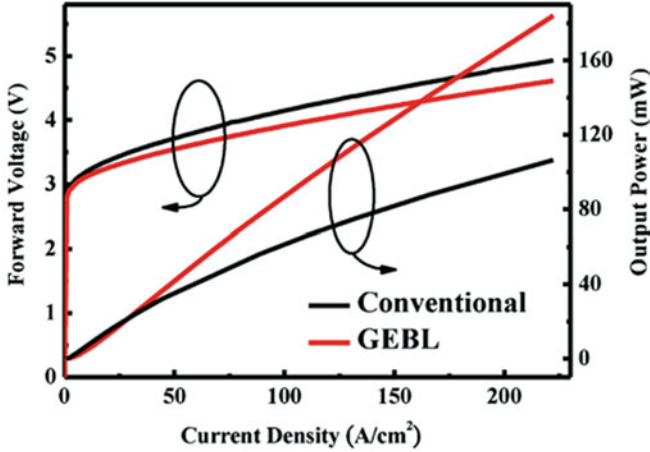


Fig. 17 Forward voltage and output power as a function of current density for conventional and GEBL LEDs

adopted the Al/Ga ratio ramping because the temperature ramping would change the growth rate, and the higher temperature might damage the quality of QWs. The growth temperature of conventional EBL and GEBL was the same (870 °C), and the aluminum-composition profile of the GEBL was approximately graded from 0 % to 25 %. Finally, the LED chips were fabricated by regular chip process with ITO current spreading layer and Ni/Au contact metal, and the size of mesa is $300 \times 300 \mu\text{m}^2$. The emission wavelengths of both LEDs were around 450 nm at 22 A/cm^2 .

Figure 17 shows the L-I-V curves of the conventional and GEBL LEDs. The output powers were measured with a calibrated integrating sphere. The forward voltages (V_f) at 22 A/cm^2 and series resistances (R_s) of GEBL LED are 3.28 V and 7Ω , respectively, which are lower than that of 3.4 V and 8Ω for conventional LED. The reduced (V_f) and (R_s) can be attributed to the improvement in hole injection and the higher p-type-doping efficiency in GEBL (Simon et al. 2010). In the case of L-I curves in Fig. 17, although the output power of GEBL LED is a little lower at low current density (below 30 A/cm^2), it increases more rapidly as the injection current increases as compared to the conventional one. The output powers were enhanced by 40 % and 69 % at 100 and 200 A/cm^2 , respectively. This phenomenon can be explained as follows. At low current density, it is more difficult for holes to tunnel across the barrier at the interface of p-GaN and EBL in GEBL LED because the ΔE_v is larger than that in conventional LED. While, at high current density, the tunneling process of holes can be negligible, and the diffusion process is dominated for the hole transportation into the MQW (Han et al. 2009). As discussed above, the diffusion process in GEBL is much easier than that in conventional one due to the flat valence band and much lower ΔE_v at the interface of the last GaN barrier and EBL. In conjunction with the superior electron confinement, much stronger light output was achieved in GEBL LED at high current density.

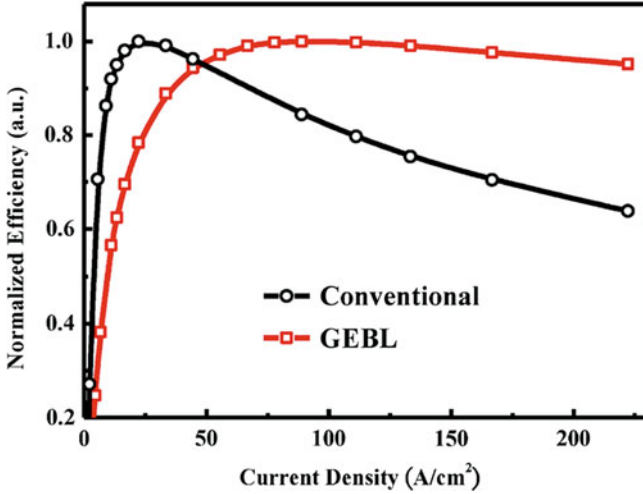


Fig. 18 Normalized efficiency as a function of current density for conventional and GEBL LEDs

Finally, the normalized efficiency of conventional and GEBL LEDs as a function of current density was investigated, as shown in Fig. 18. The maximum efficiency (η_{peak}) of GEBL LED appears at an injection current density of 80 A/cm^2 , which was much higher than that for conventional LED (at 20 A/cm^2). More interestingly, the efficiency droop, defined as $\left(\eta_{\text{peak}} - \eta_{200 \text{ A/cm}^2} / \eta_{\text{peak}}\right)$, was reduced from 34 % in conventional LED to only 4 % in GEBL LED. This significant improvement in efficiency can be mainly attributed to the enhancement of hole injection as well as electron confinement, especially at high current density.

In conclusion, we have designed a graded-composition electron-blocking layer for InGaN/GaN LED by employing the band engineering. The simulation results showed that the triangular barrier of conventional EBL at the valence band could be balanced, while the slope of the conduction band could be increased by increasing the bandgap of $\text{Al}_x\text{Ga}_{1-x}\text{N}$ along the [0001] direction. As a result, the hole concentration in MQWs was significantly increased, while the electron distribution within the GEBL region and p-GaN was enormously decreased over two orders, indicating that the GEBL can effectively improve the capability of hole transportation across the EBL as well as the electron confinement. Furthermore, the LED structure with GEBL was realized by MOCVD. The L-I-V characteristics of GEBL LED showed the smaller (V_f) and (R_s) due to the improvement in hole injection and a more effective p-type doping in GEBL as compared to the conventional LED. More importantly, the efficiency droop was reduced from 34 % in conventional LED to only 4 % in GEBL LED. This work implies that carrier transportation behavior could be appropriately modified by employing the concept of band engineering.

Improvements of the GaN-Based LEDs' Main Material Qualities

Freestanding GaN Substrate (FS-GaN)

The recent availability of freestanding GaN (FS-GaN) substrate with low defect density (about $10^6/\text{cm}^2$) maybe could facilitate this development, which can enhance the light output power, IQE and EQE of InGaN-based UV-LEDs (Fang et al. 2012; Cao et al. 2004b; Kim et al. 2007b). Hence, in order to further improve the performance of UV-LEDs with InGaN/InAlGaN MQW devices, in the present study, we will introduce FS-GaN substrate for the homo-epitaxial growth of high-quality InGaN-based UV-LED devices with InAlGaN quaternary barrier. Besides, the effects of the FS-GaN substrate on the InGaN-based UV-LEDs grown by atmospheric pressure metal-organic chemical vapor deposition (AP-MOCVD) are systematically analyzed. Detailed analyses of the grown InGaN-based UV-LEDs will be demonstrated, and electro-optical properties of UV-LEDs based on such FS-GaN substrate will also be discussed.

All UV-LEDs with InGaN/InAlGaN MQW epitaxial structure were grown by a commercial AP-MOCVD system (model: Nippon SR4000) with a horizontal reactor in the same run. The substrates employed herein were 2-in freestanding GaN (FS-GaN) substrate with 300 μm in thickness, which was provided from Sumitomo Electric Industries, Ltd in Japan and was manufactured from GaAs substrate (Nakamura and Motoki 2013). The epitaxial structure of the UV-LEDs investigated is depicted in Fig. 19, comprising a 2.5- μm -thick n-GaN epilayer grown at 1150 $^\circ\text{C}$, a ten-period InGaN/InAlGaN MQWs active layer grown at 830 $^\circ\text{C}$, a 15-nm-thick Mg-doped $\text{Al}_{0.3}\text{Ga}_{0.7}\text{N}$ and a 10-nm-thick Mg-doped $\text{Al}_{0.1}\text{Ga}_{0.9}\text{N}$ electron-blocking layers (EBLs) grown at 1030 $^\circ\text{C}$, a 55-nm-thick p-GaN layer grown at 1030 $^\circ\text{C}$, and a 5-nm-thick p-InGaN contact layer. In addition, the growth of UV-LEDs with InGaN/InAlGaN MQW epitaxial structure on undoped GaN/sapphire template was also conducted for comparison. After epitaxial growth, the indium tin oxide (ITO) film (180 nm) was first deposited on the UV-LEDs as a transparent contact layer (TCL). Then partially etching the surfaces of the UV-LEDs until 1.5 μm depth of the n-GaN layers were exposed. We subsequently deposited Cr/Pt/Au (50/50/150 nm)

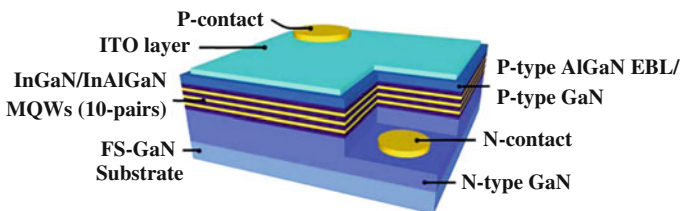
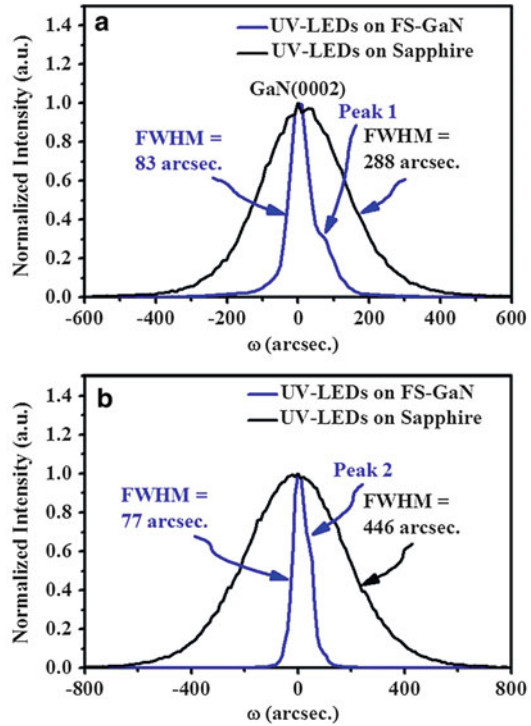


Fig. 19 Schematics of UV-LEDs with InGaN/InAlGaN MQW structures grown on FS-GaN substrate

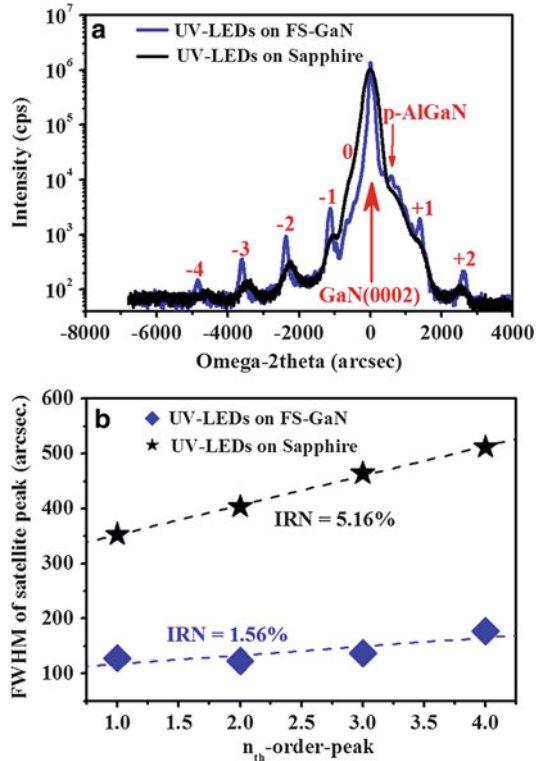
Fig. 20 HRDCXD rocking curves of UV-LEDs grown on FS-GaN and sapphire substrate in the (a) (0002) and (b) ($10\bar{1}2$) reflection



onto the exposed n-GaN and p-GaN layer to serve as the n-type and p-type electrode, respectively. Finally, UV-LEDs were cut into square pieces with a dimension of $300 \times 300 \mu\text{m}^2$.

Figure 20 shows the HRDCXD rocking curves in the (0002) and ($10\bar{1}2$) reflections, obtained from the UV-LEDs grown on FS-GaN and sapphire substrate. The full widths at half maximum (FWHM) of UV-LEDs on FS-GaN substrate are only 83 arcsec for (0002) and 77 arcsec for ($10\bar{1}2$), respectively. Both FWHM values of UV-LEDs on FS-GaN substrate are narrower than that of UV-LEDs on sapphire substrate by a factor of 4–5. In wurtzite GaN-based films, the FWHM of the (0002) rocking curve is associated with the density of screw or mixed dislocations, while the FWHM of the ($10\bar{1}2$) rocking curve is affected by all dislocations (Heying et al. 1996). Therefore, we can calculate the total defect density by the formula of Lee et al. (2005). The calculated total defect density is about $(6\text{--}7) \times 10^8 \text{ cm}^{-2}$ for UV-LEDs on sapphire substrate and $(7\text{--}10) \times 10^6 \text{ cm}^{-2}$ for UV-LEDs on FS-GaN substrate. This result shows that the defect density of UV-LEDs can be reduced dramatically by introducing the FS-GaN substrate as a substrate for epitaxy of InGaN/InAlGaN MQW UV-LEDs. In addition, it is clear that the rocking curves in the (0002) and ($10\bar{1}2$) reflection contain the additional peaks by Gauss function

Fig. 21 (a) Shows (0002) reflection HRDCXD $\omega/2\theta$ curves of UV-LEDs grown on FS-GaN and sapphire and (b) displays the FWHM of satellite peak in InGaN/InAlGaN MQWs as a function of the order of the satellite peak



which peak 1 is 72.01 arcsec for (0002) and peak 2 is 48.44 arcsec for $(10\bar{1}2)$ away from the main peak of FS-GaN substrate, respectively. These additional peaks, which peak 1 and peak 2 in the (0002) and $(10\bar{1}2)$ reflection, respectively, correspond to internal structural grain boundary (Bhagavannarayana et al. 2005) with very low angle (tilt angle ≤ 1 arcmin) whose tilt angle of peak 1 is 72.01 arcsec for (0002) and peak 2 is 48.44 arcsec for $(10\bar{1}2)$ by Gauss function, from the adjoining regions, respectively. The HRDCXD results were attributed to the hexagonal inverse pyramidal pits which are constructed by $\{11\bar{2}2\}$ facets of FS-GaN substrate (Nakamura and Motoki 2013).

Figure 21a shows the HRDCXD $\omega/2\theta$ scan for the UV-LED epitaxial structures grown on the FS-GaN and sapphire substrates. As can be seen, all the HRDCXD patterns demonstrate periodical structures, which can be attributed to the InGaN/InAlGaN MQWs grown. The strongest peak is due to the GaN layer, and evidently the spectra for UV-LEDs on FS-GaN substrate clearly show high-order InGaN/InAlGaN MQWs diffraction peaks with the fourth-order satellite peak still being observable indicating good layer periodicity. Besides, the InGaN/InAlGaN MQW period can be determined from the positions of the InGaN/InAlGaN MQW satellite peaks. It suggests that the specimen of UV-LEDs on FS-GaN substrate also has

abrupt interfaces between InAlGa_nN barrier and InGa_nN well. The interface roughness (IRN) of InGa_nN/InAlGa_nN MQW structures grown on FS-GaN substrate was further analyzed by using the following Eq. 1 (Zhang et al. 2005; Li et al. 2009):

$$W_n = W_0 + (\ln 2)^{\frac{1}{2}n} \Delta\theta_M \cdot \frac{\sigma}{\Lambda}, \quad (1)$$

where n is the order of the satellite; Λ and σ/Λ are the period of satellite peak and IRN, respectively; $\Delta\theta_M$ is the angle distance between adjacent satellite peaks; W_0 and W_n are the full width at half maximum of the zeroth- and n th-order peaks, respectively.

The Fig. 21b shows the FWHM of satellite peak in InGa_nN/InAlGa_nN MQWs as a function of the order of the satellite peak. As can be seen, the IRN of UV-LEDs on sapphire substrate is about 5.16 %. However, the IRN can further be reduced to 1.55 % as the UV-LEDs grown on the FS-GaN substrate. As indicated in Zhang et al. (2005), the IRN of MQWs is affected by the defects, microstructure, and phase separation in MQWs. Therefore, our X-ray analysis results might manifest that the crystalline quality of the InGa_nN/InAlGa_nN MQW epitaxial structure grown on the FS-GaN substrate is superior to those grown on sapphire substrates.

It was well known that the traditional InGa_nN-based UV-LEDs existed large biaxial strain due to large lattice mismatch and thermal expansion coefficient incompatibility between GaN epilayer and sapphire, resulting in a large piezoelectric field along the c -plane orientation, thus forming the quantum-confined Stark effect (QCSE) in MQWs, leading to the reduction in light output power and recombination rate of electron and hole. As the c -plane FS-GaN substrate manufacturing from GaAs substrate which either less lattice mismatch or less difference of thermal expansion coefficient (Nakamura and Motoki 2013) than FS-GaN substrate manufacturing from sapphire (Ishikawa et al. 2012; Liu et al. 2014), the less residual stress in the FS-GaN substrate manufacturing from GaAs substrate may be obtained. Thus in this paper using a complementary method for the residual stress measurement is Raman spectroscopy. The residual stress measurement using the Raman spectroscopy is based on the peak position shift of observable Raman modes, which is directly proportional to the lattice strain. Figure 22a, b show the Raman spectrum for UV-LEDs grown on sapphire and FS-GaN substrate. The Raman spectra show two Raman shift peaks: one can be attributed to the E_2 (high) mode of GaN epilayer and another can be attributed to the A_1 (LO) mode of GaN epilayer. According to Harima (2002) report, the E_2 (high) mode gives atomic displacement perpendicular to the c -axis, while the other A_1 (LO) mode gives atomic displacement along the c -axis. Although two optical photon modes which E_2 (high) and A_1 (LO) are observed in c -plane GaN epilayer backscattering according to the selection rules, our study selected the nonpolar E_2 (high) photon mode to estimate the residual stress due to be sensitive to the strain in the basal plane (c -plane) and directly to the in-plane residual stress in the (0001)-oriented GaN epilayers. The Raman shift peak of E_2 (high) for UV-LEDs grown on FS-GaN and sapphire substrates shown in Fig. 22c was located at around 567.2 and

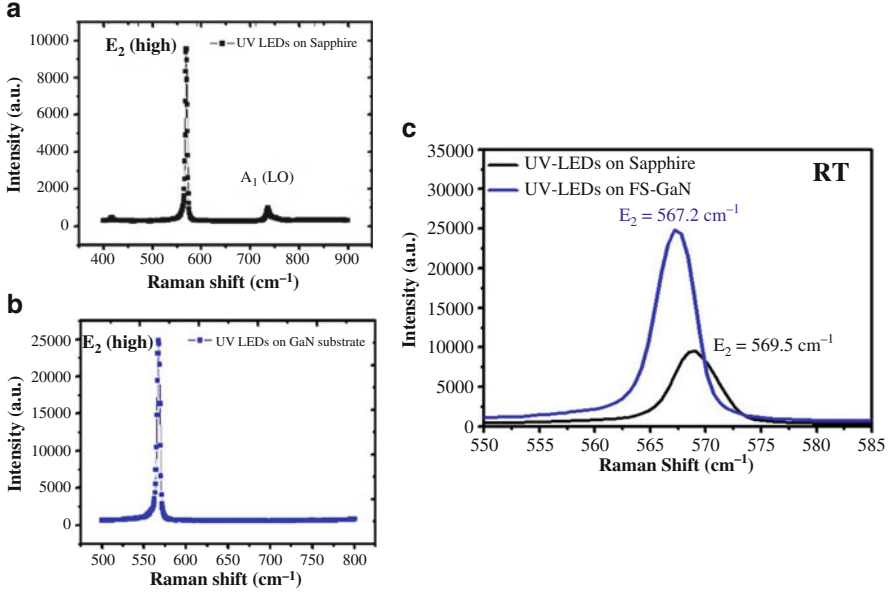


Fig. 22 Room temperature (*RT*) Raman spectrum for UV-LEDs grown on (a) sapphire and (b) FS-GaN substrates; (c) the residual stress based on the peak position shift of observable Raman E_2 (high) modes

569.7 cm^{-1} , respectively. We can calculate the strain value of InGaN/InAlGaN MQWs by following Eq. 2 (Puech et al. 2004):

$$\Delta\omega_{E_2}(\text{cm}^{-1}) = -2.25\sigma(\text{GPa}) \quad (2)$$

where $\Delta\omega$ is the Raman shift peak difference between the strained GaN epilayer and the strain-free GaN epilayer. The calculated in-plane compressive stress (σ) was about 0.31 and 1.12 GPa for UV-LEDs grown on FS-GaN and sapphire substrate, respectively. In other words, the InGaN-based UV-LED was grown on FS-GaN substrate by homo-epitaxial technique, i.e., FS-GaN substrate can help the recombination rate of electron and hole, and thus the light output power is expected to rise. Hence, the results from the Raman analyses are in good agreement with those presented by the studies of electric properties described below.

On the other hand, we used TEM to investigate crystalline quality of UV-LEDs grown on FS-GaN and sapphire substrate. In other words, the two-beam TEM images will be used for quantifying the nature and density of defect. Heying et al. pointed out that the pure edge and mixed defects can be visible under $g = (10\bar{1}0)$ two-beam condition; the pure screw and mixed defects are visible under $g = (0002)$ two-beam condition (Heying et al. 1996). In our case, Fig. 23a, b show bright-field scanning TEM images for UV-LEDs grown on sapphire and FS-GaN substrate, respectively. Besides, the two-beam TEM images were taken, as shown in

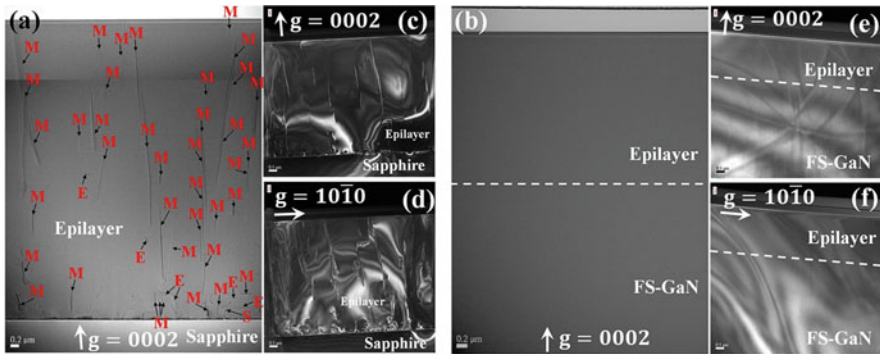


Fig. 23 Bright-field cross-sectional TEM images from UV-LEDs on (a) sapphire and (b) FS-GaN substrate. Two-beam TEM images were taken from UV-LEDs on sapphire (c, d) and FS-GaN (e, f). Where (c) and (e) dark-field cross section, $g = 0002$, (d) and (f) dark-field cross section, $g = 10\bar{1}0$

Fig. 23c, d for UV-LEDs on sapphire and (e)-f for UV-LEDs on FS-GaN. When UV-LEDs were grown on sapphire, we can clearly find that many edge, screw, and mixed defects appear in the epilayer of UV-LEDs grown on sapphire. Besides, the defects can be seen radiating vertically from the interface between GaN-based epilayer and sapphire into the InGaN/InAlGaIn MQW region and p-AlGaIn layer, as shown in Fig. 23c, d. Therefore, quite a large number of defects were presented in the whole film on sapphire substrate, as shown in Fig. 23a. From the image of Fig. 23a, the edge, screw, and mixed defects were estimated to be 1.0×10^8 , 3.6×10^6 , and $4.5 \times 10^8 \text{ cm}^{-2}$, respectively, leading to a total defect density of about $5.5 \times 10^8 \text{ cm}^{-2}$. However, when the substrate for the epitaxy of UV-LEDs was changed from sapphire to FS-GaN, it can be clearly found that the crystallography of UV-LED epilayer was drastically different from that of UV-LED epilayer on sapphire substrate. No edge, screw, and mixed defects were observed throughout the observed area. As shown in Fig. 23b, e, f, the total defect density including edge, screw, and mixed type was considered to be less than $3.6 \times 10^6 \text{ cm}^{-2}$ or less, which agrees well with our HRDCXD rocking curve data, further proving that homo-epitaxial is an effective measure to improve the crystal quality of UV-LEDs.

To better understand the effect of defect on the formation of V-shape pits (V-pits) or microstructures or quantum dot (QD) structures of InGaN/InAlGaIn MQWs, the high-resolution TEM images of InGaN/InAlGaIn MQW region was performed, as shown in Fig. 24. As can be seen, the difference of InGaN/InAlGaIn MQWs between the two specimens was relatively large. For UV-LEDs on sapphire substrate, there are many V-pits created at MQW region, as shown in Fig. 24a. Besides, the microstructures or QD structures can be clearly observed, and the spacing between these microstructures or QD structures was estimated to be 1–2 nm, as shown in Fig. 24b. However, when InGaN/InAlGaIn MQWs was grown on FS-GaN substrate, the InGaN/InAlGaIn MQWs exhibited relatively perfect crystalline structure and without any V-pits or microstructures or QD structures to be observed, as shown in Fig. 24c, d. The nonuniform distribution of indium and phase separation in the

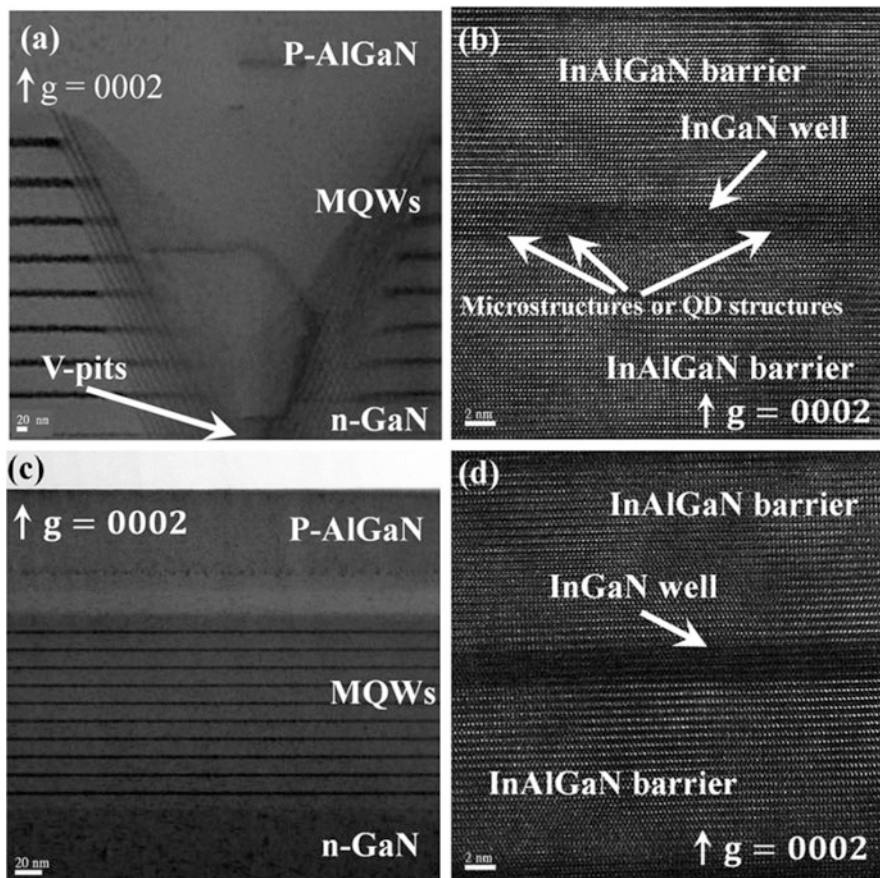
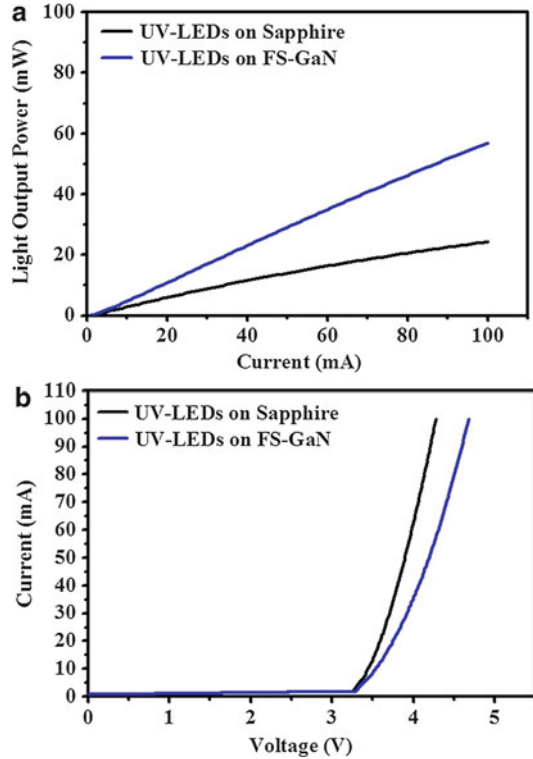


Fig. 24 HRTEM images for UV-LEDs on sapphire substrate (a, b) and FS-GaN (c, d). The diffraction condition is $g = 0002$

InGaN well was not the main reason for resulting in the difference of InGaN/InAlGaN MQWs between the two specimens since they were grown in the same run. The difference of substrate for epitaxy of UV-LEDs could be responsible for such a great difference in InGaN/InAlGaN MQWs between the two specimens since the homo-epitaxial growth of UV-LEDs on FS-GaN substrate can reduce the defects in MQWs. Besides, the V-pits or microstructures or QD structures can result in rougher interfaces in InGaN/InAlGaN MQWs. Therefore, compared the results in Fig. 24 with those in Fig. 23, we can demonstrate that the UV-LEDs with InGaN/InAlGaN MQWs grown on FS-GaN substrate exhibited more order structure, better uniformity, and sharp interface in InGaN/InAlGaN MQWs due to reduction in the defects including edge, screw, and mixed type, radiating from the interface between GaN-based epilayer and substrate.

Fig. 25 (a) Light output as a function of forward current and (b) shows the I–V characteristics for UV-LEDs grown FS-GaN and sapphire substrate. All data in Fig. 7 were obtained under CW operation



The LED devices fabricated on the basis of the material structure of UV-LEDs with InGaN/InAlGaIn MQWs grown on the FS-GaN and sapphire substrates were characterized. Figure 25a gives the light output power as a function of the forward current for all the UV-LEDs fabricated. Generally, the commercialized UV-LEDs operate at a forward current of 20 mA for $300 \times 300 \mu\text{m}^2$ in chip size. However, as exhibited in Fig. 25b, the I–V characteristics of the UV-LEDs grown on sapphire substrate differ somehow from those of the UV-LEDs grown on FS-GaN substrates. Hence, in our case, the light output power of UV-LEDs was obtained as follows. For the UV-LEDs grown on FS-GaN and sapphire substrate, the light output power was obtained under the forward current of 20 and 100 mA. The light output powers obtained from UV-LEDs grown on sapphire substrate are 6.0 mW at 20 mA and 24.3 mW at 100 mA. However, the light output power can go up to 10.8 mW at 20 mA and 57 mW at 100 mA as the UV-LEDs were grown on FS-GaN substrate. In other words, compared with UV-LEDs grown on sapphire, the light output power for UV-LEDs grown on FS-GaN substrate was improved by 80 % at 20 mA and 90 % at 100 mA. Evidently, the light output of UV-LEDs grown on FS-GaN substrate is superior to sapphire case, which helps to further recognize the merit of using the FS-GaN substrate for the fabrication of UV-LEDs.

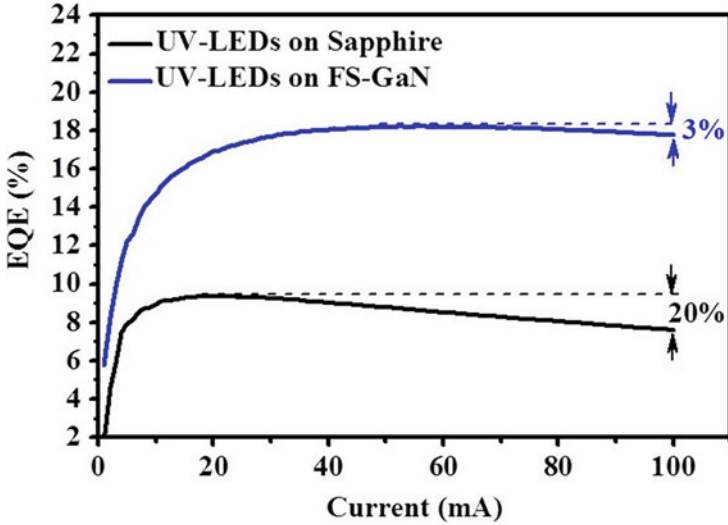


Fig. 26 The external quantum efficiency as a function of forward current under CW operation for UV-LEDs on sapphire and FS-GaN substrate

Finally, Fig. 26 shows the external quantum efficiency (EQE) as a function of forward current for UV-LEDs grown on FS-GaN and on sapphire substrate. The EQE of these two types of UV-LEDs are under CW operations. The maximum EQE (η_{peak}) of UV-LEDs grown on FS-GaN substrate appear at an injection current of 53 mA, which was much higher than that for UV-LEDs grown on sapphire substrate (at 20 mA). More interestingly, the efficiency droop, defined as $(\eta_{\text{peak}} - \eta_{100 \text{ mA}}) / \eta_{\text{peak}}$, was reduced from 20 % in UV-LEDs grown on sapphire substrate to 3 % in UV-LEDs grown on FS-GaN substrate. Recently, the results of Yu et al. pointed out that the lower carrier localization and better uniformity in MQWs will show lower droop behavior (Yu et al. 2012). Hence, this significant improvement in efficiency can be mainly attributed to the improvement of crystal quality and uniformity in MQWs and better heat dissipation in the case of UV-LEDs grown on FS-GaN substrate. Therefore, the FS-GaN substrates should be a better choice for high-power UV-LED application.

In summary, we have grown UV-LED epitaxial structure on FS-GaN substrates manufactured from GaAs substrate by AP-MOCVD. HRDCXD $\omega/2\theta$ measurements showed that the UV-LED structure grown on FS-GaN substrate exhibited sharp interfaces between InAlGaN barrier and InGaN well and a uniform QW period of InGaN/InAlGaN. Besides, the IRN of InGaN/InAlGaN MQW structure grown on FS-GaN substrate was estimated to be 1.56 %. For crystal quality, the FWHM of the HRDCXD rocking curve in the (0002) and $(10\bar{1}2)$ reflections was about 83 and 77 arcsec, respectively. Moreover, cross-sectional TEM observations revealed that the UV-LEDs grown on FS-GaN substrate surpass those grown on sapphire substrates in material quality. Hence, the total defect density including edge, screw, and

mixed type was considered to be less than $3.6 \times 10^6 \text{ cm}^{-2}$ or less. The device of UV-LEDs grown on FS-GaN substrate shows a light output power of 10.8 mW at 20 mA and 57 mW at 100 mA in $300 \times 300 \mu\text{m}^2$ in chip size. In other words, the light output power of UV-LEDs on FS-GaN substrate is higher than that of UV-LEDs on sapphire by 80 % at 20 mA and 90 % at 100 mA. Besides, the efficiency droop was reduced from 20 % in UV-LEDs grown on sapphire substrate to 3 % in UV-LEDs grown on FS-GaN substrate. Conclusively, the use of c-plane FS-GaN substrate suggests an effective technique to fabricate thereon high-power UV-LEDs.

Low-Cost and High-Efficiency GaN-Based LEDs on Large-Area Si Substrate

Growing a device-quality InGaN-based epilayer on a Si substrate is difficult owing to large differences between both their lattice constants and their thermal expansion coefficients. GaN cannot be grown as a buffer layer directly on Si substrates owing to the poor nucleation of GaN on Si (Zamir et al. 2000). Therefore, many other growth techniques have been used; they include a GaN-free buffer layer (Kobayashi et al. 1998; Strittmatter et al. 1999), a strained layer of AlN/GaN superlattices (SLs) (Eric et al. 2001), a single AlN buffer layer or multiple layers of AlN/GaN (Marchand et al. 2001; Jang et al. 2002; Kim et al. 2001b), an AlGaIn buffer layer with an Al gradient (Able et al. 2005), a patterned Si substrate (Chiu et al. 2011a, b), selective growth (Strittmatter et al. 2001; Haffouz et al. 2003), a porous Si substrate (Missaoui et al. 2002), and Si on an insulator as a compliant substrate (Zamir et al. 2002). Although the successful growth of crack-free and high-quality GaN epilayer on Si has been reported, the white light LED performance of InGaN-based LEDs on Si remains very poor, and no major breakthrough in their emission efficiency or light output power has been made (Lau et al. 2011; Kim et al. 2011a, 2012; Osram 2012; Pinos et al. 2013).

In this research, a composite buffer layer structure (CBLS) with multiple AlGaIn layers and grading of Al composition/u-GaN1/(AlN/GaN) SLs/u-GaN2 and InAlGaIn/AlGaIn quaternary superlattices electron-blocking layers (QSLs-EBLs) for use in InGaN-based LEDs on Si substrate is designed. The design exploits the concept of compensation for tensile strain and strain engineering. A buffer layer structure and QSLs-EBL that supports InGaN-based LEDs on Si with high emission efficiency are identified by comparing the structural and optical properties of specimens that are fabricated on Si substrates. A white LEDs emitter with an emission efficiency of over 100 lm/W that fabricated from the epi-wafer of InGaN-based LEDs on Si was demonstrated, opening up the possibility of fabricating InGaN-based LEDs on Si for SSL applications.

The epitaxial structure of an InGaN-based LED was grown on Si substrate using a commercial low-pressure MOCVD system (model: Veeco K465i) with a vertical reactor. The liquid/solid MO compounds of trimethylgallium (TMGa), trimethylindium (TMIn), trimethylaluminum (TMAI), and gaseous NH_3 were used

as the sources of the reactants Ga, In, Al, and N, respectively. The carrier gas was a mixture of gaseous N_2 and H_2 . The substrates employed herein were 6-in just (111)-oriented Si substrates. These substrates exhibited n-type conductivity with a carrier concentration of approximately 10^{18} cm^{-3} . Prior to growth, the Si substrate was etched by boiling it in $H_2SO_4:H_2O_2 = 3:1$ for 15 min and then dipped in HF solution (HF; $H_2O = 1:10$) for 15 s to remove the native oxide that formed on the surface of Si substrate. The Si substrate after loading was firstly heated to 1020–1050 °C under a H_2 ambient for 5–10 min to remove the surface-passivated layer. Following thermal cleaning, the CBLs was grown; it comprised a multiple $Al_xGa_{1-x}N$ layers with an Al step gradient from 1 to 0.17, a 0.5 μm -thick u-GaN epilayer (u-GaN1), a 60-period AlN (1 nm)/GaN (1 nm) SLs, and a 1.5 μm -thick u-GaN epilayer (u-GaN2). Here, the multiple $Al_xGa_{1-x}N$ layers comprised a 76 nm-thick AlN layer, a 86 nm-thick $Al_{0.75}Ga_{0.25}N$ layer, a 133 nm-thick $Al_{0.56}Ga_{0.44}N$ layer, a 123 nm-thick $Al_{0.43}Ga_{0.57}N$ layer, a 133 nm-thick $Al_{0.34}Ga_{0.66}N$ layer, and a 143 nm-thick $Al_{0.17}Ga_{0.83}N$ layer. Finally, the epitaxial structure of the InGaN-based LEDs was grown on top of the CBLs; it comprised a 3 μm -thick n-GaN epilayer, a 30-period $In_{0.07}Ga_{0.93}N$ (2 nm)/GaN (2 nm) prestrained layer (PSL), a nine-period $In_{0.2}Ga_{0.8}N$ (3.5 nm)/GaN (12 nm) multiple quantum wells (MQWs), a 60 nm-thick p-GaN last barrier, an eight-period p- $In_{0.02}Al_{0.23}Ga_{0.75}N$ (2 nm)/p- $Al_{0.1}Ga_{0.9}N$ (2 nm) QSLs-EBL, and a 0.15 μm -thick p-GaN epilayer. Figure 27 presents (a) a photograph of the as-grown epi-wafer of an InGaN-based LED on 6 in Si(111) and a schematic view of the epitaxial structure on (b) Si substrate and (c) sapphire.

Following epitaxial growth, the blue InGaN-based LED chip is fabricated. The Si substrate is well known to have a small energy bandgap of approximately 1.12 eV, so it absorbs light with a wavelength of less than 1.1 μm . To eliminate this problem, the Si substrate must be removed. Therefore, a vertical blue LED chip is developed here; its process flowchart is as shown in Fig. 28. The epi-wafer of InGaN-based LEDs on Si was initially degreased in acetone (ACE) and isopropanol (IPA). After cleaning, it was coated with negative photoresist to a thickness of typically approximately 3.6 μm ; then a pattern was photolithographically formed with p-metal/bonding metal. Before the metal films were deposited, the specimen was dipped in HCl: H_2O (1:1) for 3 min to remove native oxide from the p-GaN epilayer surface. Then, the multilayered metal system of Ni(10 Å)/Ag(3000 Å) and Ti(500 Å)/Pt(2000 Å)/Ti(500 Å)/Au(500 Å)/AuSn(2 μm) that they were fabricated by electron beam evaporation at a pressure of 1×10^{-6} torr to provide p-metal layers and bonding metal layers, respectively, as presented in Fig. 28b. Following the deposition of the metals, the specimen was bonded to the permanent substrate (Si) via Ti(500 Å)/Pt(2000 Å)/Ti(500 Å)/Au(500 Å)/AuSn(3 μm)-bonding metal layers, as presented in Fig. 28c. The Si substrate for use in epitaxy was then removed by dipping it in $HNO_3:HF:NaClO_2 = 5:1:1$ at 80 °C for 90 min. After the Si substrate for use in epitaxy had been removed, inductively coupled plasma (ICP) is used to etch the CBLs layer, as presented in Fig. 28d. Additionally, to improve the extraction of light, a 40–50 % KOH solution was used to complete the surface roughening of the n-GaN epilayer at 80 °C for 3 min, as presented in Fig. 28e. Then, the multilayered metal system Al

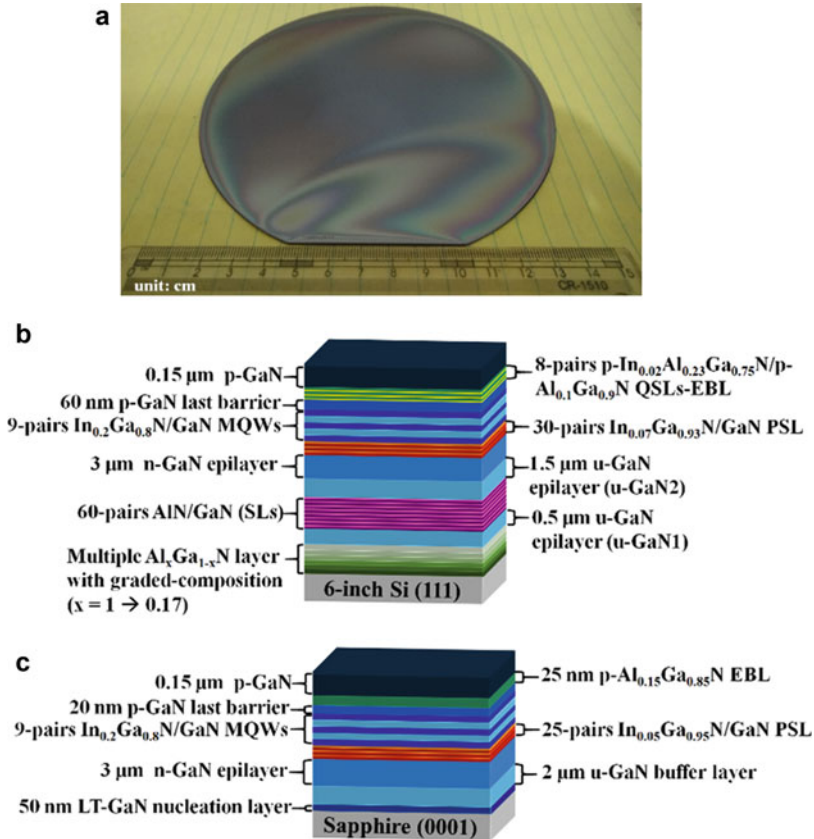


Fig. 27 (a) Photograph of as-grown epi-wafer of InGaN-based LEDs on 6-in Si (111) and schematic epitaxial structure for InGaN-based LEDs on (b) Si and (c) sapphire

(2000 Å)/Ti(500 Å)/Au(2.6 μm) was deposited on the surface of the roughened n-GaN epilayer to provide the n-metal layers. The blue LED chip was cut into squares with dimensions of $1 \times 1 \text{ mm}^2$, as presented in Fig. 28f. Finally, the blue LED chip was packaged by a lead frame 5070 (LF-5070) to produce the blue and white LED emitters, as presented in Fig. 28g, h, respectively.

In addition, a commercialized vertical LED emitter was also fabricated from the epi-wafer of InGaN-based LEDs epitaxial structure on sapphire for comparison. Due to many natural differences between sapphire and Si, such as thermal conductivity and lattice-mismatch-induced strain, the optimized growth conditions, layer structure, and chip process procedure were different from those used for the InGaN-based LEDs on a Si substrate. In other words, the epitaxial structure of the InGaN-based LEDs on sapphire substrate investigated is depicted in Fig. 27c, comprising a 50 nm-thick low-temperature GaN (LT-GaN) nucleation layer, a 2 μm-thick undoped GaN (u-GaN) buffer layer, a 3 μm-thick n-GaN epilayer, a 25-period $\text{In}_{0.05}\text{Ga}_{0.95}\text{N}$

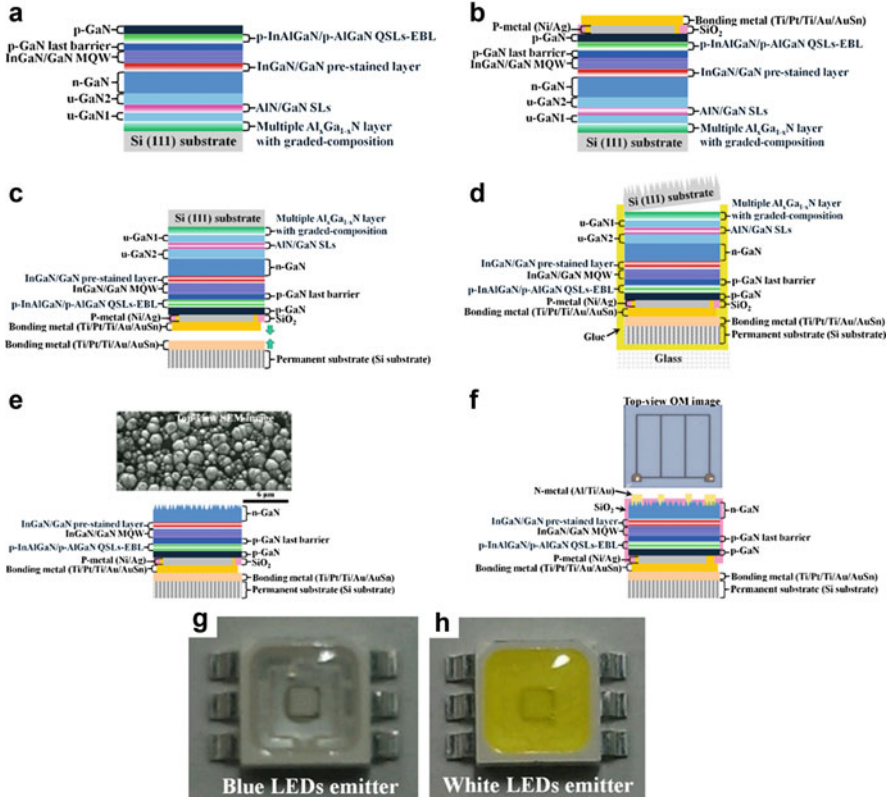


Fig. 28 Process flow for fabrication of blue and white LED emitters: (a) epitaxial growth of InGaN-based LEDs on Si substrate by MOCVD, (b) deposition of p-metal layers and bonding metal layers using E-gun evaporator, (c) wafer bonding, (d) removal of Si substrate for epitaxial growth of InGaN-based LEDs by dipping into $\text{HNO}_3\text{:HF:NaClO}_2$ solution and then CBLs etching by ICP, (e) roughening of n-GaN surface by dipping in KOH, (f) deposition of n-metal layer using E-gun evaporator and then cutting it into squares with dimensions of $1 \times 1 \text{ mm}^2$, (g) blue LED emitter that comprised blue LEDs chip and clear lens, (h) white LED emitter that comprised blue LEDs chip and yellow phosphor

(2 nm)/GaN (2 nm) PSL, a nine-period $\text{In}_{0.2}\text{Ga}_{0.8}\text{N}$ (3.5 nm)/GaN (12 nm) MQWs, a 20 nm-thick p-GaN last barrier, an 25 nm-thick $\text{p-Al}_{0.15}\text{Ga}_{0.85}\text{N}$ EBL, and a $0.15 \mu\text{m}$ -thick p-GaN epilayer. Following the epitaxial growth, the vertical-type LEDs were fabricated by wafer bonding and laser lift-off (LLO) process, and the detailed chip process procedure is described in Chiu et al. (2008b).

The surface morphology of specimens was observed by atomic force microscopy (AFM) with a scanning area of $10 \times 10 \mu\text{m}^2$. The crystalline quality and interface of the epitaxial structures herein were evaluated by high-resolution double crystal X-ray diffraction (HRDCXD D8) with $\text{Cu K}\alpha$ radiation as the X-ray source ($\lambda = 1.54056 \text{ \AA}$). The distribution and threading behaviors of dislocations in the

Table 1 Material parameters used in the simulation

Parameter	GaN	InN
m_e/m_0	0.2	0.07
m_h/m_0	1.25	0.6
γ_e	1.0	1.0
$N_{g,e}(\text{cm}^{-3})$	2×10^{17}	8×10^{18}
$\mu_{\max,e}(\text{cm}^2\text{V}^{-1}\text{s}^{-1})$	1000	1100
$\mu_{\min,e}(\text{cm}^2\text{V}^{-1}\text{s}^{-1})$	55	30
γ_h	2.0	2.0
$N_{g,h}(\text{cm}^{-3})$	3×10^{17}	3×10^{17}
$\mu_{\max,h}(\text{cm}^2\text{V}^{-1}\text{s}^{-1})$	170	340
$\mu_{\min,h}(\text{cm}^2\text{V}^{-1}\text{s}^{-1})$	3	3

epilayer were studied by transmission electron microscopy (TEM). The interfacial microstructures of the epilayer were observed by high-resolution TEM (HRTEM). Finally, the light-current-voltage (L-I-V) characteristics of all packaged LED chips were measured at room temperature in continuous-wave (CW) mode. APSYS software, which was developed by Crosslight Software Inc., was used to determine the physical origin of the improvement in the efficiency of the InGaN-based LEDs on Si. The simulated structures, such as layer thicknesses, doping concentrations, and aluminum composition, are the same as the actual devices. The commonly accepted Shockley–Read–Hall recombination lifetime approximately 6 ns, the percentage of screening of 50 %, and Auger recombination coefficient approximately $10^{-30} \text{ cm}^6\text{s}^{-1}$ are used in the simulations. Besides, other detailed material parameters used in the simulation are shown in Table 1 (Bernardini 2007).

Generally, the large difference in thermal expansion coefficients between GaN ($5.59 \times 10^{-6} \text{ K}^{-1}$) and Si ($2.59 \times 10^{-6} \text{ K}^{-1}$) produces tensile stress in GaN and causes the formation of cracks during the cooling down from the epitaxial growth temperature. On the other hand, the circular defects can be observed by optical microscope if the Ga melt-back etching (Ga-MBE) was formed on the top surface of epi-wafer. Some studies confirmed that the circular defects caused by the reaction of Ga and Si atoms out diffuse from Si substrate (Wei et al. 2011). This phenomenon was reported as Ga-MBE (Dadgar et al. 2003b). In the experiment in this work, no crack was observed, and no Ga melt-back etching (Ga-MBE) was identified in this scan, as presented in Fig. 29a. Figure 29b presents the AFM images of surface morphologies of GaN on Si. As can be seen, a very small root-mean-square (RMS) value of the surface roughness of 0.82 nm was achieved.

Figure 30a, b plot the (0002) and $(10\bar{1}2)$ reflection peaks, respectively, in the HRDCXD rocking curve of the InGaN-based LEDs that were grown on the 6-in Si substrate. The full width at half maximum (FWHM) of the (0002) peak is only 330 arcsec and that of the $(10\bar{1}2)$ peak is only 450 arcsec. The FWHM of the (0002) rocking curve of wurtzite GaN-based films is related to the density of screw or mixed dislocations, while that of the $(10\bar{1}2)$ rocking curve is related to all dislocations

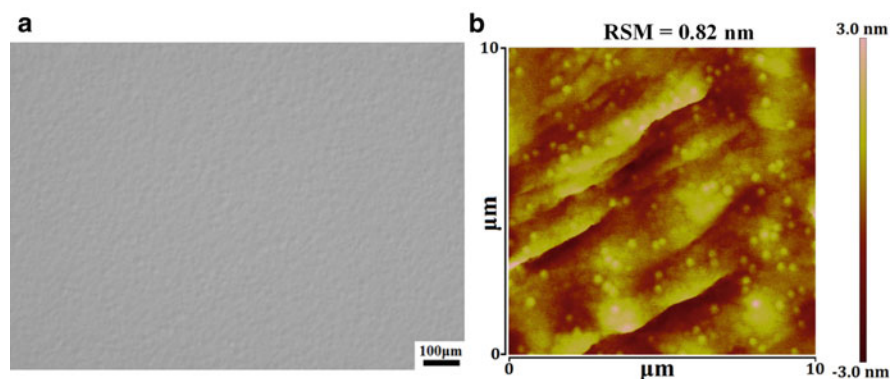
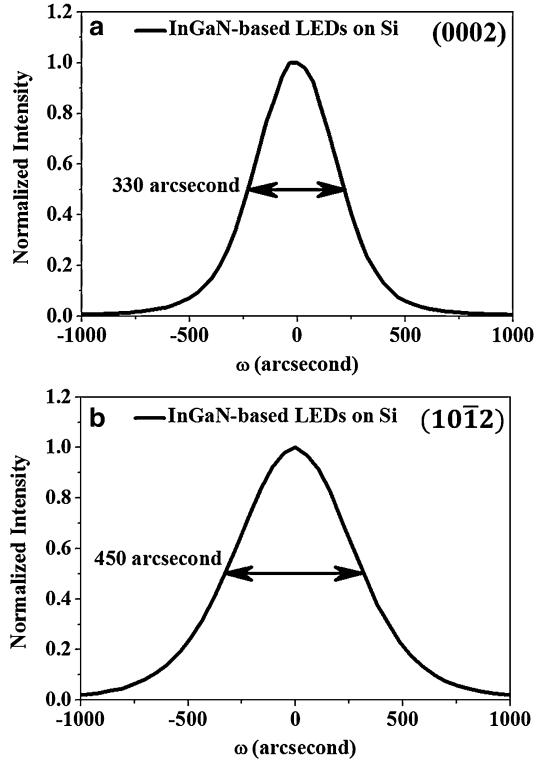


Fig. 29 (a) Optical microscopic and (b) AFM top view of surface morphology of InGaN-based LEDs on Si with CBLs and QSLs-EBL

(Heying et al. 1996). The total threading dislocation density (TDD) can be calculated using the published formula (Lee et al. 2005). The calculated total TDDs, including edge, screw, and mixed dislocations, are approximately $(6\text{--}7) \times 10^8 \text{ cm}^{-2}$. Although the FWHM of (0002) and $(10\bar{1}2)$ reflections of the InGaN-based LEDs on Si substrate do not reach the level of 250–300 arcsec, which was the value obtained from InGaN-based LEDs on sapphire, in the experiment herein, it is lower than what has been reported elsewhere, such as the value of 385 arcsec for (0002) and 795 arcsec for $(10\bar{1}2)$ that was presented in the work of Jong Ock Kim et al. (2011b).

To estimate the crystalline quality of InGaN-based LEDs on Si, TEM is utilized. Two-beam TEM images are captured to determine the nature and density of defects. Heying et al. pointed out that pure edge and mixed defects are visible under the $g = (10\bar{1}0)$ two-beam condition; pure screw and mixed defects are visible under the $g = (0002)$ two-beam condition (Heying et al. 1996). Figure 31a, d present bright-field scanning TEM images of the InGaN-based LEDs on Si herein. Figure 31b, c, e, and f present the two-beam TEM images. No screw-type dislocation is observed in the scanned area. In Fig. 31a–c, many threading dislocations (TDs) are observed at the interface between the multiple AlGaIn layers and the Si substrate; their number decreases gradually with the layer number in the multiple AlGaIn layers. Therefore, the total TDDs, including edge, screw, and mixed dislocations, were estimated to be $\sim 10^{10} \text{ cm}^{-2}$ at the interface between the multiple AlGaIn layers and the Si substrate, falling to $\sim 10^9 \text{ cm}^{-2}$ in the u-GaN1 region. When the AlN/GaN SL was introduced, fewer TDs were found in the u-GaN2 region. The densities of the edge, mixed, and screw dislocations in the u-GaN2 region were estimated to be 1.5×10^8 , 4.0×10^8 and less than $2.6 \times 10^7 \text{ cm}^{-2}$, respectively. Restated, the total TDD in the u-GaN2 region is further reduced to approximately $6.0 \times 10^8 \text{ cm}^{-2}$. The n-GaN, MQWs, and p-GaN grown on Si substrate with CBLs have many fewer TDs, which radiated vertically from the interface between the multiple AlGaIn layers and the Si substrate within the visible

Fig. 30 (a) (0002) and (b) $(10\bar{1}2)$ reflections in HRDCXD rocking curves of InGaN-based LEDs grown on Si substrate



range in view. As presented in Fig. 31d–f, the densities of edge, mixed, and screw dislocations in the n-GaN region were estimated to be 2.5×10^8 , 3.6×10^8 and less than $2.0 \times 10^7 \text{ cm}^{-2}$, respectively, yielding a total TDD of approximately $6.3 \times 10^8 \text{ cm}^{-2}$. The densities of edge, mixed, and screw dislocations in the p-GaN region were estimated to be less than 4.0×10^7 , 1.6×10^8 and less than $2.0 \times 10^7 \text{ cm}^{-2}$, respectively, yielding a total TDD of approximately $2.2 \times 10^8 \text{ cm}^{-2}$ (typically, $1 \sim 5 \times 10^8 \text{ cm}^{-2}$ for commercialized InGaN-based LEDs grown on sapphire case). Therefore, quite a large number of TDs can be blocked and bent within the multiple AlGaIn layers with the grading of Al composition and AlN/GaN SLs, as presented in Fig. 31a. The TEM agrees closely with the HRDCXD rocking curve data, further proving that the CBLS is effective in improving the crystalline quality of InGaN-based LEDs on Si substrates. To determine the microstructures in the InGaN/GaN MQW region, HRTEM images, presented in Fig. 31g, were obtained. Clearly, the MQW region in the InGaN-based LEDs on Si exhibited a more ordered structure, a greater uniformity, and a sharper interface between the InGaN well and the GaN barrier.

Finally, to understand the performance of devices, the LED emitters fabricated on the basis of the material structure of InGaN-based LEDs grown on Si substrates were characterized. In the experiment herein, measurements and two types of LED

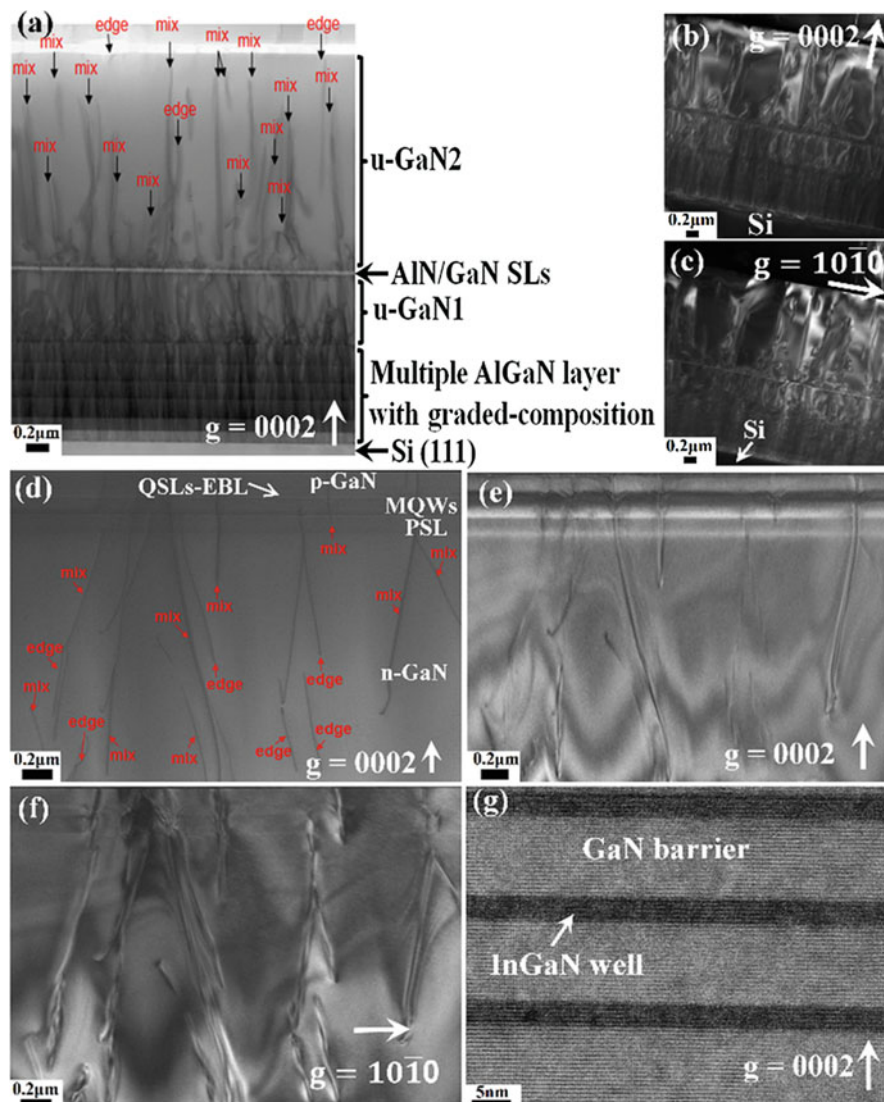
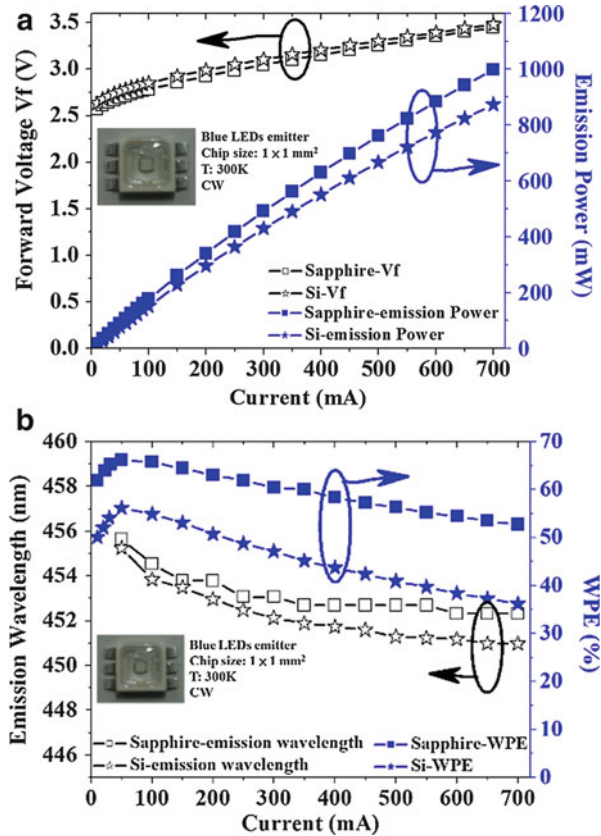


Fig. 31 (a) and (d) Bright-field cross-sectional TEM images of InGaN-based LEDs on Si. Two-beam TEM images of InGaN-based LEDs on Si (b, c, e, f); (b) and (e) dark-field cross section, $g = 0002$; (c) and (f) dark-field cross section, $g = 10\bar{1}0$. (g) The HRTEM image of InGaN-based LEDs on Si

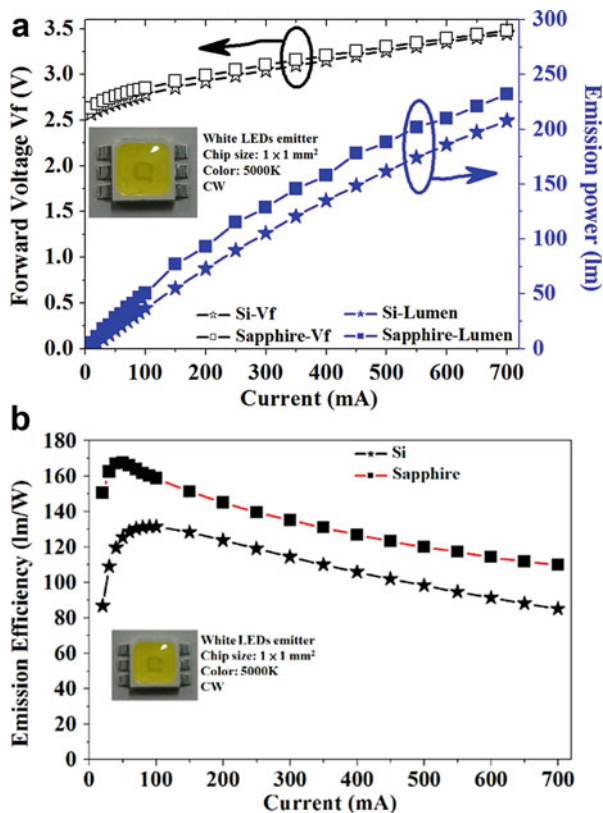
emitters were made and analyses performed: one was the blue LED emitter that comprised blue LED chip and clear lenses; the other was the white LED emitter that comprised blue LED chip and yellow phosphor. Most commercialized power chip LEDs are operated at a forward current of 350 mA per $1 \times 1 \text{ mm}^2$ of chip area. The V_f , emission power, and wall-plug efficiency (WPE, η) of the blue LED emitter,

Fig. 32 The electrical characteristic of blue LED emitter that comprised blue LED chip and clear lens for all the vertical InGaN-based LEDs fabricated. (a) Light-current–voltage (L-I-V) curve, (b) emission wavelength versus injection current and droop behavior



which was fabricated from the epi-wafer of the InGaN-based LEDs on a Si substrate, were approximately 3.1 V, 490 mW, and 45 %, respectively, at a forward current of 350 mA, as presented in Fig. 32a, b. The strain-induced shift in emission wavelength was approximately 5 nm (from 50 to 700 mA), as presented in Fig. 32b. Additionally, the efficiency droop, defined as $(\eta_{350 \text{ mA}} - \eta_{700 \text{ mA}}) / \eta_{350 \text{ mA}}$, was approximately 80 %, as presented in Fig. 32b. From Fig. 33a, the Vf and emission power of the white LED emitter, which was fabricated from the epi-wafer of InGaN-based LEDs on a Si substrate, were approximately 3.1 V and 120 lm, respectively, at a forward current of 350 mA, yielding an emission efficiency of as high as 110 lm/W. The efficiency droop was estimated to be approximately 78 %, as presented in Fig. 33b. Therefore, a comparison of the results in Fig. 33 with those in Fig. 32 reveals that the performance of the blue and white LED emitter that was fabricated from the epi-wafer of InGaN-based LEDs on a Si substrate was comparable to the performance of the blue and white LED emitter that was fabricated from the epi-wafer of InGaN-based LEDs on sapphire because the crystalline quality, stress, and recombination rate of electron and hole in the InGaN-based LEDs on the Si substrate were drastically improved by introducing a CBLS and QSLs-EBL.

Fig. 33 The electrical characteristic of white LED emitter that comprised blue LED chip and yellow phosphor for all the vertical InGaN-based LEDs fabricated. (a) Light-current-voltage (L-I-V) curve, (b) droop behavior



More interestingly, the emission efficiency of the white LED emitter, which was fabricated from the epi-wafer of InGaN-based LEDs on a Si substrate, was successfully increased to 110 lm/W. To the best of the authors' knowledge, this white LED emitter with InGaN-based LEDs on a Si substrate is the first to reach an efficiency of 110 lm/W (Lau et al. 2011; Kim et al. 2011a, 2012; Osram 2012; Pinos et al. 2013). We believe that the efficiency of InGaN-based LEDs on Si substrate can be further increased by improving the technique of epitaxial growth and the chip process.

From the above analyses, the possible mechanisms by which the CBLS and QSLs-EBL yield superior InGaN-based LEDs on Si substrate are as follows:

- (a) First, the Al content in multiple AlGaIn layers with a graded composition gradually decreased from the Si substrate to the u-GaN1 side, causing the lattice constant of the final AlGaIn layer in the multiple AlGaIn layers to equal that of u-GaN1, according to Vegard's law (Denton and Ashcroft 1991). Therefore, the multiple AlGaIn layers may generate a graded lattice constant between the upper u-GaN1 layer and the lower Si substrate, yielding a lower density of TDDs in the u-GaN1 region, as presented in Fig. 34a density.

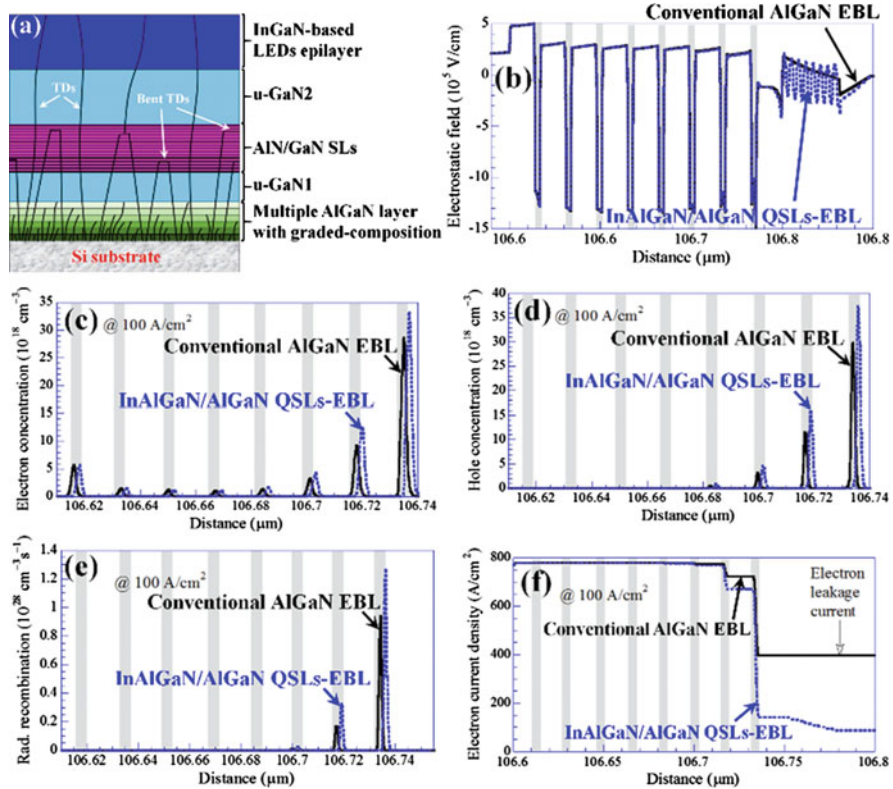


Fig. 34 (a) Potential mechanisms of reduction of TDDs in InGaN-based LEDs on Si with CBLs. Results of APSYS simulation of InGaN-based LEDs on Si with InAlGaIn/AlGaIn QSLs-EBL under a high forward current density of 100 A/cm^2 ; distribution of (b) electrostatic field, (c) electrons, (d) holes, (e) radiative recombination rate, and (f) electron leakage current

- (b) SLs with strained layers are reportedly effective for reducing the TDDs of epilayer owing to the highly coherent strain energy in the region of the SLs, which arises from the interface misfit strain in the system of SLs. The strain in the region of SLs exerts a net force on the dislocations causing them to be bent or terminated at the strained epilayer edge without threading through the epilayer to the top surface (Yamaguchi et al. 1989; Whelan et al. 1990). In our work, the thickness of each layer in the SL structure is approximately 1 nm. The Peach–Koehler force exceeds the line tension of the TDDs, pulling the TDDs along the layer interface of the SL structure (Sharan et al. 1987). The high strain in the SL structure is concluded effectively to bend and suppress the TDDs, reducing the TDDs of the epilayer and improving the surface morphology of the specimen, as presented in Figs. 34a and 29b.
- (c) To improve the performance of InGaN-based LEDs on Si, an InAlGaIn/AlGaIn QSLs-EBL is designed to replace the conventional AlGaIn EBL. For a general

InGaN-based LED structure, the interface between the AlGaN EBL and the last GaN barrier is in a strong piezoelectric polarization field (Kim et al. 2007a), which promotes the leakage of carriers. Some research has shown that the lattice constant of InAlGaN with a particular In content almost matches the lattice constant of the GaN epilayer (Schubert et al. 2008). Some research has also verified that the band offset between the GaN epilayer and the InAlGaN epilayer is higher than the band offset between the GaN epilayer and the AlGaN epilayer (Tu et al. 2011). Restated, when the InAlGaN/AlGaN QSLs-EBL is used, the higher band offset and weaker lattice-mismatch-induced piezoelectric polarization field can favor the electron confinement and distribution of holes in the MQW region. The simulation results, as presented in Fig. 34b–f, in this study reveal that the sample with the InAlGaN/AlGaN QSLs-EBL exhibited a weaker electrostatic field than the sample with conventional AlGaN EBL, better carrier transport/distribution and lower carrier leakage owing to the weaker polarization effects, and improved carrier confinement. A direct consequence is the increase in the radiative recombination rate in the MQW region and, therefore, a rise in light output power.

These improvements that are produced by the use of a CBLs and QSLs-EBL are responsible for the high performance of InGaN-based LEDs on Si with reduced TDDs and strain and increased light output power and emission efficiency.

High performance InGaN-based LEDs on Si substrate were successfully grown without cracking and Ga-MBE, using a CBLs of multiple AlGaN layers with the grading of Al composition/u-GaN1/(AlN/GaN) SLs/u-GaN2 and an InAlGaN/AlGaN QSLs-EBL. The RMS value of surface roughness scanned by AFM was as low as about 0.82 nm. With respect to crystalline quality, the FWHM of the (0002) and (10 $\bar{1}$ 2) reflection peaks in the HRDCXD rocking curve were approximately 330 and 450 arcsec, respectively. Additionally, cross-sectional TEM observations revealed that the total TDDs, including edge, screw, and mixed dislocations, were estimated to be approximately $(3.0\text{--}6.0) \times 10^8 \text{ cm}^{-2}$. The results of TEM indicated that the InGaN-based LEDs on the Si substrate were of almost the same material quality as those grown on sapphire substrates (typically, $1 \sim 5 \times 10^8 \text{ cm}^{-2}$). Based on the results obtained above, the InGaN-based LEDs on a Si substrate had an emission power of 490 mW at 350 mA in the blue LED emitter and an emission efficiency of 110 lm/W at 350 mA in the white LED emitter. The efficiency droops were estimated to be 80 % and 78 %, respectively.

Reference:

- Able A, Wegscheider W, Engl K, Zweck J (2005) Growth of crack-free GaN on Si(1 1 1) with graded AlGaN buffer layers. *J Cryst Growth* 276:415
- Akasaka T, Gotoh H, Saito T, Makimoto T (2004) High luminescent efficiency of InGaN multiple quantum wells grown on InGaN underlying layers. *Appl Phys Lett* 85:3089

- Arif RA, Ee YK, Tansu N (2007) Polarization engineering via staggered InGa_N quantum wells for radiative efficiency enhancement of light emitting diodes. *Appl Phys Lett* 91:091110
- Bernardini F (2007) Chapter 3, Spontaneous and piezoelectric polarization: basic theory vs. practical recipes. In: Piprek J (ed) Nitride semiconductor devices: principles and simulation. Wiley, New York, pp 49–67
- Bhagavannarayana G, Ananthamurthy RV, Budakoti GC, Kumar B, Bartwal KS (2005) A study of the effect of annealing on Fe-doped LiNbO₃ by HRXRD, XRT and FT-IR. *J Appl Crystallogr* 38:768–771
- Butter E, Fitzl G, Hirsch D, Leonhardt G, Seifert W (1979) The deposition of group III nitrides on silicon substrates. *Thin Solid Films* 59:25
- Cao XA, LeBoeuf SF (2007) Current and temperature dependent characteristics of deep-ultraviolet light-emitting diodes. *IEEE Trans Electron Devices* 54(12):3414–3417
- Cao XA, Teetsov JA, Shahedipour-Sandvik F, Arthur SD (2004a) Microstructural origin of leakage current in GaN/InGa_N light-emitting diodes. *J Cryst Growth* 264:172
- Cao XA, LeBoeuf SF, D'Evelyn MP, Arthur SD, Kretchmer J, Yan CH, Yang ZH (2004b) Blue and near-ultraviolet light-emitting diodes on free-standing GaN substrate. *Appl Phys Lett* 84:4313–4315
- Chao C-L, Xuan R, Yen H-H, Chiu C-H, Fang Y-H, Li Z-Y, Chen B-C, Lin C-C, Chiu C-H, Guo Y-D, Kuo H-C, Chen J-F, Cheng S-J (2011) Reduction of efficiency droop in InGa_N light-emitting diode grown on self-separated freestanding GaN substrates. *IEEE Photon Technol Lett* 23:798
- Charash R, Maaskant PP, Lewis L, McAleese C, Kappers MJ, Humphreys CJ, Corbett B (2009) Carrier distribution in InGa_N/Ga_N tricolor multiple quantum well light emitting diodes. *Appl Phys Lett* 95:151103
- Chiu CH, Yen HH, Chao CL, Li ZY, Peichen Y, Kuo HC, Lu TC, Wang SC, Lau KM, Cheng SJ (2008a) Nanoscale epitaxial lateral overgrowth of GaN-based light-emitting diodes on a SiO₂ nanorod-array patterned sapphire template. *Appl Phys Lett* 93:081108
- Chiu CH, Lee CE, Chao CL, Cheng BS, Huang HW, Kuo HC, Lu TC, Wang SC, Kuo WL, Hsiao CS, Chen SY (2008b) Enhancement of light output intensity by integrating ZnO nanorod arrays on GaN-based LLO vertical LEDs. *Electrochem Solid-State Lett* 11:H84
- Chiu C-H, Lin D-W, Lin C-C, Li Z-Y, Chen Y-C, Ling S-C, Kuo H-C, Lu T-C, Wang S-C, Liao W-T, Tanikawa T, Honda Y, Yamaguchi M, Sawaki N (2011a) Optical properties of (11 $\bar{1}$ 01) semi-polar InGa_N/Ga_N multiple quantum wells grown on patterned silicon substrates. *J Cryst Growth* 318:500
- Chiu C-H, Lin C-C, Deng D-M, Lin D-W, Li J-C, Li Z-Y, Shu G-W, Lu T-C, Shen J-L, Kuo H-C, Lau K-M (2011b) Optical and electrical properties of GaN-based light emitting diodes grown on micro- and nano-scale patterned Si substrate. *IEEE J Quantum Electron* 47:899
- Dadgar A, Bläsing J, Diez A, Alam A, Heuken M, Krost A (2000) Metalorganic chemical vapor phase epitaxy of crack-free GaN on Si (111) exceeding 1 μm in thickness. *Jpn J Appl Phys* 39:1183
- Dadgar A, Poschenrieder M, Bläsing J, Contreras O, Bertram F, Riemann T, Reiher A, Kunze M, Daumiller I, Krtschil A, Diez A, Kaluza A, Modlich A, Kamp M, Christen J, Ponce FA, Kohn E, Krost A (2003a) MOVPE growth of GaN on Si(1 1 1) substrates. *J Cryst Growth* 248:556
- Dadgar A, Strittmatter A, Bläsing J, Poschenrieder M, Contreras O, Veit P, Riemann T, Bertram F, Reiher A, Krtschil A, Diez A, Hempel T, Finger T, Kasic A, Schubert M, Bimberg D, Ponce FA, Christen J, Krost A (2003b) Metalorganic chemical vapor phase epitaxy of gallium-nitride on silicon. *Phys Status Solidi (C)* 0:1583
- David A, Grundmann MJ (2010) Droop in InGa_N light-emitting diodes: a differential carrier lifetime analysis. *Appl Phys Lett* 96:103504
- David A, Grundmann MJ, Kaeding JF, Gardner NF, Mihopoulos TG, Krames MR (2008) Carrier distribution in (0001)InGa_N/Ga_N(0001)InGa_N/Ga_N multiple quantum well light-emitting diodes. *Appl Phys Lett* 92:053502
- Davis RF, Bishop SM, Mita S, Collazo R, Reitmeier ZJ, Sitar Z (2007) Epitaxial growth of gallium nitride. *AIP Conf Proc* 916:520

- Denton AR, Ashcroft NW (1991) Vegard's law. *Phys Rev A* 43:3161
- Ding K, Zeng YP, Wei XC, Li ZC, Wang JX, Lu HX, Cong PP, Yi XY, Wang GH, Li JM (2009) A wide-narrow well design for understanding the efficiency droop in InGaN/GaN light-emitting diodes. *Appl Phys B: Lasers Opt* 97:465
- Eric F, Beaumont B, Laügt M, de Mierry P, Vennéguès P, Lahrèche H, Leroux M, Gibart P (2001) Stress control in GaN grown on silicon (111) by metalorganic vapor phase epitaxy. *Appl Phys Lett* 79:3230
- Fang YH, Fu YK, Xuan R (2012) High efficiency and output power of near-ultraviolet light-emitting diodes grown on GaN substrate with back-side etching. *Phys Scr* 85:045703
- Feng ZC, Zhang X, Chua SJ, Yang TR, Deng JC, Xu G (2002) Optical and structural properties of GaN materials and structures grown on Si by metalorganic chemical vapor deposition. *Thin Solid Films* 409:15
- Goldenberg B, Zook JD, Ulmer RJ (1993) Ultraviolet and violet light-emitting GaN diodes grown by low-pressure metalorganic chemical vapor deposition. *Appl Phys Lett* 62:381
- Gong JR, Yeh MF, Wang CL (2003) Growth and characterization of GaN and AlN films on (1 1 1) and (0 0 1) Si substrates. *J Cryst Growth* 247:261
- Haffouz S, Grzegorzczak A, Hageman PR, Vennéguès P, Van der Drift EWJM, Larsen PK (2003) Structural properties of maskless epitaxial lateral overgrown MOCVD GaN layers on Si (1 1 1) substrates. *J Cryst Growth* 248:568
- Han S-H, Lee D-Y, Lee S-J, Cho C-Y, Kwon M-K, Lee SP, Noh DY, Kim D-J, Kim YC, Park S-J (2009) Effect of electron blocking layer on efficiency droop in InGaN/GaN multiple quantum well light-emitting diodes. *Appl Phys Lett* 94:231123
- Harima H (2002) Properties of GaN and related compounds studied by means of Raman scattering. *J Phys Condens Matter* 14:R967–R993
- Heying B, Wu XH, Keller S, Li Y, Kopolnek D, Keller BP, DenBaar SP, Speck JS (1996) Role of threading dislocation structure on the x-ray diffraction peak widths in epitaxial GaN films. *Appl Phys Lett* 68(5):643–645
- Ishikawa Y, Tashiro M, Hazu K, Furusawa K, Namita H, Nagao S, Fujito K, Chichibu SF (2012) Local lifetime and luminescence efficiency for the near-band-edge emission of free-standing GaN substrates determined using spatio-time-resolved cathodoluminescence. *Appl Phys Lett* 101(21):212106
- Jang S-H, Lee S-J, Seo I-S, Ahn H-K, Lee O-Y, Leem J-Y, Lee C-R (2002) Characteristics of GaN/Si(1 1 1) epitaxy grown using $\text{Al}_{0.1}\text{Ga}_{0.9}\text{N}/\text{AlN}$ composite nucleation layers having different thicknesses of AlN. *J Cryst Growth* 241:289
- Jeong S-M, Kissinger S, Kim D-W, Lee SJ, Kim J-S, Ahn H-K, Lee C-R (2010) Characteristic enhancement of the blue LED chip by the growth and fabrication on patterned sapphire (0 0 0 1) substrate. *J Cryst Growth* 312(258)
- Katsuragawa M, Sota S, Komori M, Anbe C, Takeuchi T, Sakai H, Amano H, Akasaki I (1998) Thermal ionization energy of Si and Mg in AlGaIn. *J Cryst Growth* 189–190:528
- Kim M-H, Do Y-G, Kang HC, Noh DY, Park S-J (2001a) Effects of step-graded Al_xGa_{1-x}N interlayer on properties of GaN grown on Si (111) using ultrahigh vacuum chemical vapor deposition. *Appl Phys Lett* 79:2713
- Kim M-H, Do Y-G, Kang HC, Noh DY, Park S-J (2001b) Effects of step-graded Al_xGa_{1-x}N interlayer on properties of GaN grown on Si (111) using ultrahigh vacuum chemical vapor deposition. *Appl Phys Lett* 79(2713)
- Kim MH, Schubert MF, Dai Q, Kim JK, Schubert EF, Piprek J, Park Y (2007a) Origin of efficiency droop in GaN-based light-emitting diodes. *Appl Phys Lett* 91:183507
- Kim KC, Schmidt MC, Sato H, Wu F, Fellows N, Jia Z, Satio M, Nakamura S, DenBaars SP, Speck JS, Fujito K (2007b) Study of nonpolar m-plane InGaIn/GaN multiquantum well light emitting diodes grown by homoepitaxial metal-organic chemical vapor deposition. *Appl Phys Lett* 91(18):181120
- Kim J-Y, Tak Y, Hong H-G, Chae S, Lee JW, Choi H, Kim JK, Min B, Park Y, Chung U-I, Kim M, Lee S, Cha N, Shin Y, Sone C, Kim J-R, Shim J-I (2011a) Highly efficient InGaIn/GaN blue

- LEDs on large diameter Si(111) substrates comparable to those on sapphire. In: Proceedings of the SPIE 8123, eleventh international conference on solid state lighting, 81230A, San Diego, 23 Sept 2011
- Kim JO, Hong SK, Kim H, Moon KW, Choi CJ, Lim KY (2011b) *J Korean Phys Soc* 58:1374
- Kim J-Y, Tak Y, Kim J, Hong H-G, Chae S, Lee JW, Choi H, Park Y, Chung U-I, Kim J-R, Shim J-I (2012) Highly efficient InGaN/GaN blue LED on 8-inch Si (111) substrate. In: Proceedings of the SPIE 8262, gallium nitride materials and devices VII, 82621D, San Francisco, 9 Feb 2012
- Kobayashi NP, Kobayashi JT, Choi W-J, Dapkus PD, Zhang X, Rich DH (1998) Growth of single crystal GaN on a Si substrate using oxidized AlAs as an intermediate layer. *J Cryst Growth* 189/190:172
- Krost A, Dadgar A (2002) GaN-based optoelectronics on silicon substrates, *Mater Sci Eng B* 93:77
- Kuo YK, Chang JY, Tsai MC, Yen SH (2009) Advantages of blue InGaN multiple-quantum well light-emitting diodes with InGaN barriers. *Appl Phys Lett* 95:011116
- Lau KM, Wong KM, Zou X, Chen P (2011) Performance improvement of GaN-based light-emitting diodes grown on patterned Si substrate transferred to copper. *Opt Express* 19:A956
- Lee SR, West AM, Allerman AA, Waldrip KE, Follstaedt DM, Provencio PP, Koleske DD, Abernathy CR (2005) Effect of threading dislocation on the Bragg peakwidths of GaN, AlGaN, and AlN heterolayers. *Appl Phys Lett* 86(24):241904
- Li ZY, Uen WY, Lo MH, Chiu CH, Lin PC, Hung CT, Lu TC, Kuo HC, Wang SC, Huang YC (2009) Enhancing the emission efficiency of In_{0.2}Ga_{0.8}N/GaN MQW blue LED by using appropriately misoriented sapphire substrates. *J Electrochem Soc* 156(2):H129–H133
- Ling SC, Lu TC, Chang SP, Chen JR, Kuo HC, Wang SC (2010) Low efficiency droop in blue-green m-plane InGaN/GaN light emitting diodes. *Appl Phys Lett* 96:231101
- Liu BL, Lachab M, Jia A, Yoshikawa A, Takahashi K (2002) MOCVD growth of device-quality GaN on sapphire using a three-step approach. *J Cryst Growth* 234:637
- Liu N, Wu J, Li W, Luo R, Tong Y, Zhang G (2014) Highly uniform growth of 2-inch GaN wafers with a multi-wafer HVPE system. *J Cryst Growth* 388:132–136
- Lu Y, Liu X, Lu D-C, Yuan H, Hu G, Wang X, Wang Z, Duan X (2003) The growth morphologies of GaN layer on Si(1 1 1) substrate. *J Cryst Growth* 247:91
- Marchand H, Zhao L, Zhang N, Moran B, Coffie R, Mishra UK, Speck JS, DenBaars SP, Freitas JA (2001) Metalorganic chemical vapor deposition of GaN on Si(111): Stress control and application to field-effect transistors. *J Appl Phys* 89:7846
- Missaoui A, Ezzaouia H, Bessaïs B, Boufaden T, Matoussi A, Bouaïcha M, El Jani B (2002) Morphological study of GaN layers grown on porous silicon. *Mater Sci Eng B* 93:102
- Monemar B, Sernelius BE (2007) Defect related issues in the “current roll-off” in InGaN based light emitting diodes. *Appl Phys Lett* 91:181103
- Nakamura S (1998) The roles of structural imperfections in InGaN-based blue light-emitting diodes and laser diodes. *Science* 281:956
- Nakamura T, Motoki K (2013) GaN substrate technologies for optical devices. *Proc IEEE* 101 (10):2221–2228
- Nakamura S, Mukai T, Senoh M (1994) Candela-class high-brightness InGaN/AlGaN double-heterostructure blue-light-emitting diodes. *Appl Phys Lett* 64:1687
- Ni X, Fan Q, Shimada R, Özgür Ü, Morkoç H (2008) Reduction of efficiency droop in InGaN light emitting diodes by coupled quantum wells. *Appl Phys Lett* 93:171113
- Osram (2012) Osram Opto unveils R&D results from GaN LEDs grown on silicon. <http://ledsmagazine.com/news/9/1/19>
- Pinos A, TanW-S, Chitnis A, Nishikawa A, Groh L, Hu C-Y, Murad S, Lutgen S (2013) Highly uniform electroluminescence from 150 and 200 mm GaN-on-Si-based blue light-emitting diode wafers. *Appl Phys Express* 6:095502
- Piprek J (2007) Nitride semiconductor devices: principles and simulation. Wiley, Berlin, p 279
- Puech P, Demangeot F, Frandon J, Pinquier C, Kuball M, Domnich V, Gogotsi Y (2004) GaN nanoindentation: a micro-Raman spectroscopy study of local strain fields. *J Appl Phys* 96 (5):2853–2856

- Schubert MF, Xu J, Kim JK, Schubert EF, Kim MH, Yoon S, Lee SM, Sone C, Sakong T, Park Y (2008) Polarization-matched GaInN/AlGaInNGaInN/AlGaInN multi-quantum-well light-emitting diodes with reduced efficiency droop. *Appl Phys Lett* 93:041102
- Sharan S, Jagannadham K, Narayan J (1987) Stress distribution and critical thicknesses of thin epitaxial films. *Mat Res Soc Symp Proc* 91:311
- Simon J, Protasenko V, Lian C, Xing H, Jena D (2010) Polarization-induced hole doping in wide-band-gap uniaxial semiconductor heterostructures. *Science* 327:60
- Son JK, Lee SN, Sakong T, Paek HS, Nam O, Park Y, Hwang JS, Kim JY, Cho YH (2006) Enhanced optical properties of InGaN MQWs with InGaN underlying layers. *J Cryst Growth* 287:558
- Strittmatter A, Krost A, T€urck V, Straßburg M, Bimberg D, Bl€asing J, Hempel T, Christen J, Neubauer B, Gerthsen D, Christmann T, Meyer BK (1999) LP-MOCVD growth of GaN on silicon substrates – comparison between AlAs and ZnO nucleation layers. *Mater Sci Eng B* 59:29
- Strittmatter A, Rodt S, Reißmann L, Bimberg D, Schröder H, Obermeier E, Riemann T, Christen J, Krost A (2001) Maskless epitaxial lateral overgrowth of GaN layers on structured Si(111) substrates. *Appl Phys Lett* 78:727
- Sun CK, Keller S, Chiu TL, Wang G, Minsky MS, Bowers JE, DenBaars SP (1997a) Well-width dependent studies of InGaN-GaN single-quantum wells using time-resolved photoluminescence techniques. *IEEE J Sel Top Quantum Electron* 3:731
- Sun CK, Chiu TL, Keller S, Wang G, Minsky MS, DenBaars SP, Bowers JE (1997b) Time-resolved photoluminescence studies of InGaN/GaN single-quantum-wells at room temperature. *Appl Phys Lett* 71:425
- Takeuchi T, Sota S, Katsuragawa M, Komori M, Takeuchi H, Amano H, Akasaki I (1997) Quantum-confined stark effect due to piezoelectric fields in GaInN strained quantum wells. *Jpn J Appl Phys* 36:L382
- Tu PM, Chang CY, Huang SC, Chiu CH, Chang JR, Chang WT, Wu DS, Zan HW, Lin CC, Kuo HC, Hsu CP (2011) Investigation of efficiency droop for InGaN-based UV light-emitting diodes with InAlGaN barrier. *Appl Phys Lett* 98:211107
- Uen W-Y, Li Z-Y, Lan S-M, Liao S-M (2005) Epitaxial growth of high-quality GaN on appropriately nitridated Si substrate by metal organic chemical vapor deposition. *J Cryst Growth* 280:335
- Vampola KJ, Iza M, Keller S, DenBaars SP, Nakamura S (2009) Measurement of electron overflow in 450 nm InGaN light-emitting diode structures. *Appl Phys Lett* 94:061116
- Vurgaftman I, Meyer JR (2003) Band parameters for nitrogen-containing semiconductors. *J Appl Phys* 94:3675
- Wang CH, Chen JR, Chiu CH, Kuo HC, Li YL, Lu TC, Wang SC (2010) Temperature-dependent electroluminescence efficiency in blue InGaN-GaN light-emitting diodes with different well widths. *IEEE Photon Technol Lett* 22:236
- Wei M, Wang X, Pan X, Xiao H, Wang CM, Hou Q, Wang Z (2011) Effect of AlN buffer thickness on GaN epilayer grown on Si(1 1 1). *Mater Sci Semicond Process* 14(97)
- Wetzel C, Volm D, Meyer BK, Pressel K, Nilsson S, Mikhov EN, Baranov PG (1994) GaN epitaxial layers grown on 6H-SiC by the sublimation sandwich technique. *Appl Phys Lett* 65:1033
- Whelan JS, George T, Weber ER, Nozaki S, Wu AT, Umeno M (1990) Transmission electron microscopy investigation of dislocation bending by GaAsP/GaAs strained-layer superlattices on heteroepitaxial GaAs/Si. *J Appl Phys* 68:5115
- Xie J, Ni X, Fan Q, Shimada R, Özgür Ü, Morkoç H (2008) On the efficiency droop in InGaN multiple quantum well blue light emitting diodes and its reduction with p-doped quantum well barriers. *Appl Phys Lett* 93:121107
- Yamaguchi M, Nishioka T, Sugo M (1989) Analysis of strained-layer superlattice effects on dislocation density reduction in GaAs on Si substrates. *Appl Phys Lett* 54:24

- Yu SF, Lin RM, Chang SJ, Chu FC (2012) Efficiency droop characteristics in InGaN-based near ultraviolet-to-blue light-emitting diodes. *Appl Phys Express* 5:022102
- Zamir S, Meyler B, Zolotoyabko E, Salzman J (2000) The effect of AlN buffer layer on GaN grown on (1 1 1)-oriented Si substrates by MOCVD. *J Cryst Growth* 218:181
- Zamir S, Meyler B, Salzman J (2002) Reduction of cracks in GaN films grown on Si-on-insulator by lateral confined epitaxy. *J Cryst Growth* 243:375
- Zehnder U, Weimar A, Strauss U, Fehrer M, Hahn B, Lugauer H-J, H€arle V (2001) Industrial production of GaN and InGaN-light emitting diodes on SiC-substrates. *J Cryst Growth* 230:497
- Zhang H, Ye Z, Zhao B (2000) Investigation of preparation and properties of epitaxial growth GaN film on Si(1 1 1) substrate. *J Cryst Growth* 210:511
- Zhang JC, Jiang DS, Sun Q, Wang JF, Wang YT, Liu JP, Chen J, Jin RQ, Zhu JJ, Ying H (2005) Influence of dislocations on photoluminescence of InGaN/GaN multiple quantum wells. *Appl Phys Lett* 87(7):071908

LED Materials: GaN on Si

Armin Dadgar and Alois Krost

Contents

Introduction	124
Development of GaN on Si LEDs	125
Light Propagation and Extraction	126
GaN on Si LED Growth	131
Strain Engineering	132
Semi-polar LED Growth	137
Processing of GaN-on-Si LEDs	137
Thermal and Electrical Considerations	138
On Wafer LEDs	139
Thin Film LEDs	141
Closing Remarks	143
References	144

Abstract

LED materials for incandescent lighting are based on thin Gallium-Nitride layers. Due to the lack of Gallium-Nitride substrates such layers are usually grown as thin crystal layers on sapphire or silicon-carbide substrates. Gallium-Nitride grown on silicon is a material platform which offers a huge benefit as low substrate cost, large substrate diameter, and also opens a route for manufacturing in depreciated Si wafer fabs. But long GaN on Si was believed to be a niche and not suited for high performance devices. This is because material growth requires processes with temperatures above 1000 °C and thermal stress leads to cracking

A. Dadgar (✉)

Fakultaet fuer Naturwissenschaften, Abteilung Halbleiterepitaxie, Otto-von-Guericke Universitaet Magdeburg, Magdeburg, Germany
e-mail: armin.dadgar@ovgu.de

A. Krost

Otto-von-Guericke Universität Magdeburg, Magdeburg, Germany
e-mail: alois.krost@physik.uni-magdeburg.de

of layers even below device relevant thicknesses. In the last 15 years these problems have been solved and today GaN on Si based LEDs are competitive to GaN on sapphire based devices. This chapter describes the development of GaN on Si LEDs, the differences to GaN on sapphire based structures and different routes for achieving a high output power although these layers are originally grown on a light absorbing substrate.

Introduction

Because of their unique properties group-III-nitrides are ideally suited as efficient light emitters in the visible spectral range. Therefore GaN based LEDs already replace the classical light bulb in many application fields. Lab results demonstrate high external quantum efficiencies in the blue spectral region of 84.3 % and wall plug efficiencies of up to 81.3 % for white light LEDs (Narukawa et al. 2010) and with it the unbeatable potential of GaN based LEDs for incandescent lighting. Still some problems have to be solved, as efficient longer wavelength emission which is hampered by the quantum confined Stark effect (Waltereit et al. 2000) as well as in-homogenous InGaN alloys and difficulties in semi- and non-polar material growth as strain relaxation (Fischer et al. 2009). Also substrate choice is a long debated topic. While GaN growth on sapphire and SiC is well established since the early 1990s, GaN on Si growth took longer until the mid 2000s to be accepted as a potential alternative. Now even GaN substrates come into focus for LED manufacturing with Soraa, a Californian company founded by UCSB professors and LED pioneer Shuji Nakamura, Steven DenBaars and James Speck offering GaN on GaN based devices.

In contrast to all other substrates the benefit of Si substrates is their low price, large diameter and established processing while sapphire or even GaN substrates benefit from their high transparency and stiffness and GaN substrates also from their high conductivity. GaN substrates are, however, expensive and still not available in large quantities and diameters.

Nowadays Si substrates have been established for GaN LED production by only few bigger companies as Toshiba (technology from Bridgelux), Sanken Electric, and Latticepower but a significant growth of this market segment is expected and some companies are known to develop the technology as Samsung and OSRAM OS (status early 2014).

The reasons why GaN on Si has not been widely established until today are mostly difficulties in epitaxial growth and with it often a lower yield and lower output power than for processes on sapphire. Epitaxial growth is usually performed by metalorganic vapor phase epitaxy (MOVPE) at temperatures of up to 1100 °C. The high growth temperature leads to the first difficulty because of a large thermal mismatch between GaN and Si leading to cracking of thicker layers (above ~1 μm) in contrast to the growth on sapphire where it leads to compression of the GaN layer and no cracks. Further difficulties lie in the achievement of low dislocation densities and the prevention of melt-back etching, a destructive alloying reaction between Ga and Si (Ishikawa et al. 1998).

Today high power LEDs grown on sapphire are often manufactured as thin film LEDs where the III-nitride film has been mounted p-side down onto a suited carrier. But not all high power LEDs and especially not all LEDs in general are fabricated this way and other routes as sapphire structuring are also applied. High brightness LEDs on Si substrates are unlikely to be successfully applied when an LED is directly fabricated and kept on the substrate. This is because Si is not transparent and only reflective to 30–40 % in the visible wavelength range. Thus light absorption is high and overall extraction efficiency is poor as will be discussed later. This can be best solved by thin film LEDs similar to the processing of such structures grown on sapphire.

This chapter gives an overview over the development of GaN based LEDs on Si, their difference in layer structure compared to GaN based LEDs on sapphire, and the main challenges to achieve high performance devices.

Development of GaN on Si LEDs

In 1998, 6 years after the commercialization of GaN based LEDs by Nichia, the first GaN based LED on Si was demonstrated by Guha and Bojarczuk (1998a, b). In this case the LED structure was grown by molecular beam epitaxy (MBE) and yielded in low power light emission. Nevertheless, this was an important step to demonstrate that p-type doping is possible because it was commonly believed that Si, either diffusing from the substrate or by evaporation from the substrate backside, will auto-dope the III-N layers n-type and hinder successful p-type doping. For MOVPE it was soon after also demonstrated by Strittmatter et al. that Si diffusion is not a concern and only the first few nanometers, if at all, are contaminated by Si (Strittmatter et al. 1999). Typically ammonia and Si form a SiN passivation layer in a self-limiting process (de Almeida and Baumvol 2000). This layer is also present on the backside of the substrate after III-N growth in MOVPE.

LEDs were then demonstrated by several authors but they were typically cracked, thus the real device performance could not be determined. This is because cracks lead to non-uniform current distribution and, more severe, to enhanced but non-reproducible light extraction at cracks. The first such MOVPE grown and 4 μm thick but cracked LED was presented already in 1999 (Tran et al. 1999). Until 2001 MOVPE grown LEDs were presented which were either relatively thin (Umeno et al. 2001; Egawa et al. 2002) or thicker and crack-reduced by different methods (Dadgar et al. 2001a; Feltin et al. 2001). Soon after this the first crack-free GaN on Si LEDs (Dadgar et al. 2001a, 2002a, b) with a device relevant thickness well above one micron were demonstrated. This marks an important step forward. Firstly thicker layers are required for current spreading but also to reduce the dislocation density. Secondly, crack-free layers are prerequisite for device production and by demonstrating this GaN on Si LEDs lost their academic research attribute. In 2003 a blue LED with an output power close to 1 mW (0.89 mW @ 20 mA, 477 nm) was demonstrated with the GaN on sapphire LED layers grown on the same equipment yielding only a factor of 4–5 higher output power (Dadgar

et al. 2004). Taking into account light propagation in these devices the internal quantum efficiency for the device on Si was comparable to that on sapphire (for light propagation see next section). Then not much happened in the following years in regard of output power. GaN on Si on 150 mm wafers was demonstrated (Dadgar et al. 2006; Li et al. 2006), and with it also demonstrating the possibility of achieving a homogenous wavelength distribution on 150 mm substrates. Soon after LEDs on Si(001) in contrast to the commonly used Si(111) orientation were achieved (Schulze et al. 2008). But no higher power values were published. In the meantime GaN on sapphire LEDs were further improved and achieved output powers above 30 mW (@ 20 mA). The next big step came when companies like OSRAM OS, Samsung, Latticepower, Bridgelux (now the technology is with Toshiba), and also Phillips Lumileds started their GaN on Si programs. This was mostly motivated by the increase in wafer diameter, difficulties with sapphire in growth and processing on large diameters and the high price of large diameter sapphire wafers.

This industry effort finally demonstrated that an output power very close to that of state-of-the-art GaN on sapphire LEDs (OSRAM 2012) is possible. To achieve high output powers for LEDs grown on sapphire two routes can be chosen: Firstly, structuring of sapphire which leads to epitaxial lateral overgrowth (ELO, ELOG) and an improved material quality and improved light scattering at non-planar interfaces within the structure. This results in an enhancement in light extraction efficiency (Narukawa et al. 2010). Secondly, by a thin film approach where the LED layer is bonded to a new carrier with a highly reflective p-contact metal and removing the old substrate (Schnitzer et al. 1993; Haerle et al. 2004). Surface roughening then leads to an enhanced light extraction. This processing method can be also applied to LED structures on Si and led to the results presented by Latticepower and OSRAM OS (CS 2010; OSRAM 2012).

Light Propagation and Extraction

Understanding light propagation and absorption in LEDs is most important to estimate the potential and perform necessary steps to achieve high brightness LEDs, especially when grown on Si. Already a simple view on light absorption for laterally propagating light in such an LED explains why light output is not only a factor of two lower than for LEDs on sapphire but a factor of 4–5. Figure 1 shows the remaining intensity for multiple reflections for differently reflecting surfaces. Even for highly reflecting surfaces (90 %) multiple reflections significantly reduce the amount of light extracted at the side facet (<60 % after 5 reflections). In reality these values are even slightly lower when taking into account Fresnel losses for each reflection.

Directly extracted light amounts to only a small portion of total light generated. Typically, within the vertical light cone within an opening angle of $\sim 76^\circ$ ($2 \times 38^\circ$ for $n_{\text{GaN}} \sim 2.4$, $n_{\text{plastic}} \sim 1.5$) only about 10 % of all light generated internally is extracted. To this amount the backward emitted and reflected light is added. All other light undergoes total reflection at the top surface and propagates laterally. Of this

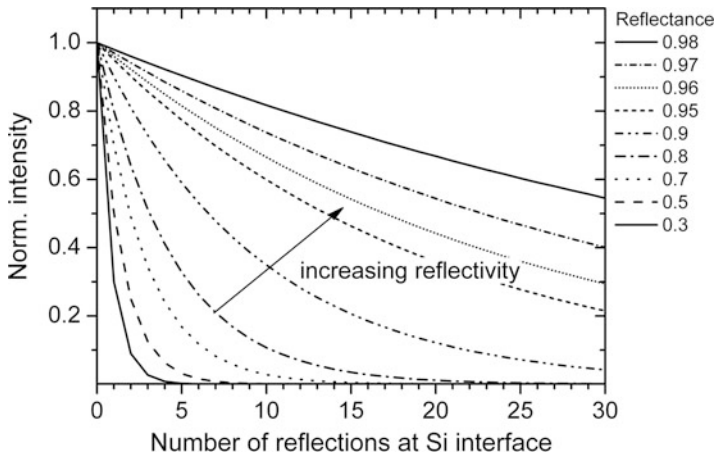


Fig. 1 Damping of light intensity after multiple reflections for different interface reflectivities. Even for a high reflectivity of 0.9 five reflections lead to a ~40 % reduction in intensity. The only way to circumvent this is by immediate extraction of light from the LED structure

light, except for a very small fraction directly hitting the side facets, all light is reflected multiple times with the losses described. In sum for a simple LED structure on Si without internal or surface texturing light output is about a factor of 4–5 lower than for a similar LED on sapphire.

Therefore, to achieve a high extraction efficiency, either a very high reflectivity at the III-nitride / carrier interface and low absorption which can be caused by, e.g., a Ni/Au p-type top contact, is required, or as few reflections as possible. As even a reflectivity above 90 % leads to a substantial damping in light intensity it is difficult to find a proper metal as few metals have higher reflectance values. E. g. silver, also well suited as p-type contact metal has a reflectivity in the visible spectrum range of around 95 %. When applying this to an LED its reflectance is usually lower due to processing and alloying with other metals. In conclusion when multiple reflections occur, even with a silver mirror on one side substantial absorption losses will occur.

Several methods exist to improve light extraction. One approach is to increase the reflectivity of the layers above the Si substrate and below the active region, e.g., by introducing a Bragg mirror. This has been tested by several authors and compared for a different number of Bragg mirror $\lambda/4$ pairs by Ishikawa et al. (2004a, b; Fig. 2). Here cracks were observed for the fivefold bragg mirror but already a threefold mirror structure lead to an approximately doubled light output power. The main difficulty of this method is the high period number necessary to achieve a high reflectivity because the usually applied layer scheme of AlGaIn/GaN does bear the risk of cracking and only offers a low refractive index change. Therefore lattice matched AlInN/GaN with a higher refractive index change was also tested by Ishikawa et al. (Ishikawa et al. 2008). With a 20-fold AlInN/GaN stack they obtained a 3.6-fold increase in light intensity.

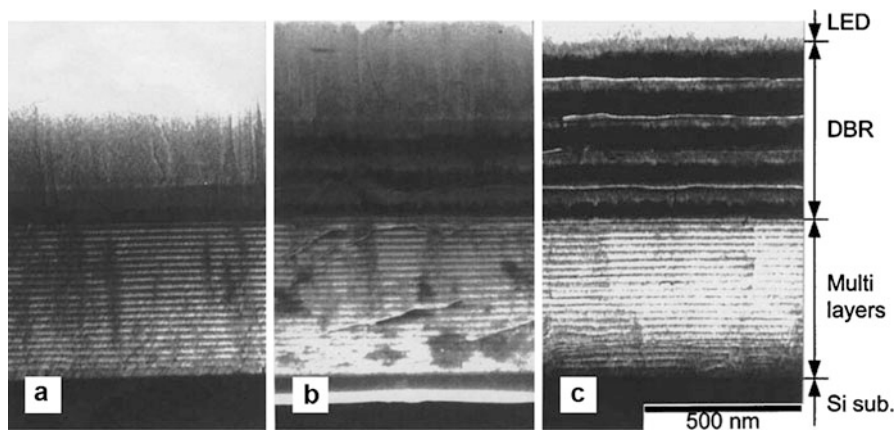


Fig. 2 TEM cross sectional images of single (a), threefold (b), and fivefold (c) DBR structures on a strain engineering AlN/GaN buffer layer (Reprinted from *Journal of Crystal Growth*, 272, Hiroyasu Ishikawa, Baijun Zhang, Kenta Asano, Takashi Egawa, Takashi Jimbo, Characterization of GaInN light-emitting diodes with distributed Bragg reflector grown on Si, 322–326, Copyright (2004), with permission from Elsevier)

But Bragg-mirror layers do only work well for vertically backward emitted light and all light emitted laterally does only partially benefit from these layers. Nevertheless with a high reflectivity and higher pumping currents in principal the vertically emitted light intensity can be boosted when the multi quantum well (MQW) is placed in a resonant cavity structure.

Another way to increase light extraction is by photonic bandgap structures (Bykov 1972; Yablonovitch 1987). They can be either etched into the surface after the structure has been grown (Tripathy et al. 2009; Lin et al. 2010), put on top of it, e. g., as structured metal or, e.g., structured Indium-Tin-Oxide (ITO) transparent contact. Alternatively they can be etched structures into the Si substrate which can also then be transferred to the top by a thin film process (Orita et al. 2008). Such structures directly on or within the substrate can also promote epitaxial overgrowth and potentially improve layer quality (Fig. 3).

It is thought that light emission can be improved by suited structures when they, e. g., suppress lateral light propagating. Nevertheless there is no substantial benefit if compared to roughened interfaces and surfaces in achieved light output power.

An alternative to such etched or structured photonic bandgap structures are nano- or microcolumn LEDs (Kikuchi et al. 2004). Their growth has been demonstrated by several groups but for mass application still issues as reliable contacting have to be solved. Also extracted light can be reabsorbed at neighboring columns if they are densely packed or if they have no inclined side facets. Indeed until now such structures have not proven to yield in enhanced light emission in LEDs and it is questionable if they can be manufactured at lower or competitive cost than conventional LED structures.

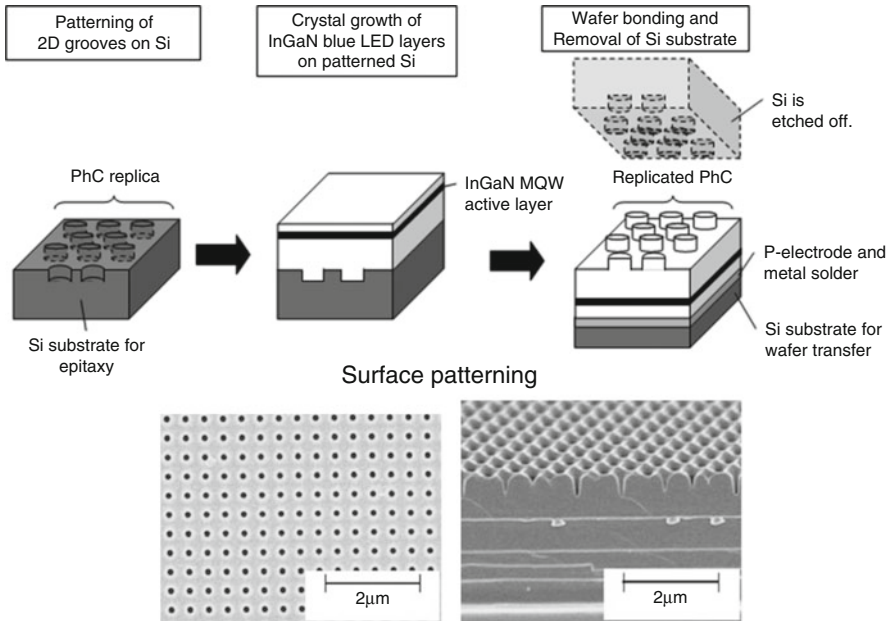


Fig. 3 Photonic bandgap structures by structuring the substrate prior to growth (*top*) (Reprinted from Quantum Electronics, IEEE Journal of, 44, Orita, K., Takase, Y., Fukushima, Y., Usuda, M., Ueda, T., Takigawa, S., Tanaka, T., Ueda, D., Egawa, T., Integration of Photonic Crystals on GaN-Based Blue LEDs Using Silicon Mold Substrates, Copyright (2008), with permission from IEEE) or the p-GaN surface after growth (*bottom*) (Reproduced by permission of The Electrochemical Society) (Images from Orita et al. (2008) and Lin et al. (2010))

The most versatile solution for light extraction is substrate removal and a thin film LED process (Zhang et al. 2005; Zou et al. 2010; Egawa and Bin Abu Bakar Ahmad 2010; Alarcón-Lladó et al. 2010; Shaohua et al. 2013). For thin film LEDs the substrate is removed and most light is extracted directly or (ideally) after only one reflection at the backside metallization. Prerequisite for this is surface structuring to avoid multiple reflections of the light. Otherwise all laterally propagating light will be extracted at the side facets after a significant reduction in intensity by absorption losses at the metal layer (see Figs. 1 and 4). In any case minimizing absorption losses requires a proper metallization scheme. For the backside metallization, ideally a highly reflective Ag coating ($R > 90\%$) is applied. It does not only reflect most of the light but also acts as p-type contact and is also part of the metal bond of the LED layer structure to a new carrier. Apart from this commonly applied method one can think of transferring the layer to a carrier with sputtered Bragg mirrors, e. g., for resonant cavity LEDs.

While in extracting light after as few as possible reflections at interior interfaces is the best method to achieve highest efficiencies also the LED shape can have a huge impact on light extraction.

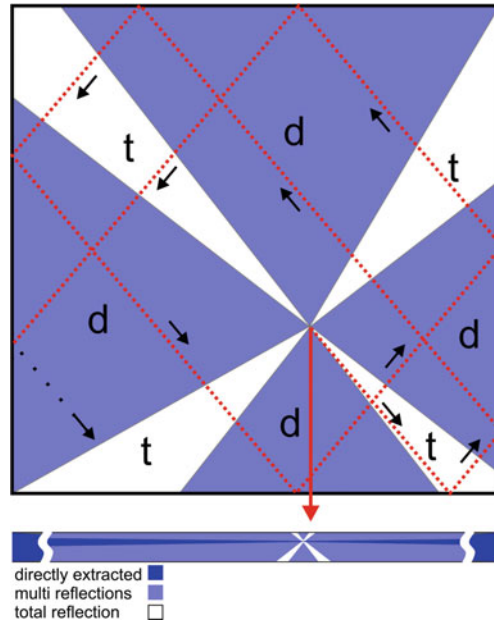


Fig. 4 Light paths in a simplified square LED chip for an arbitrary light emission position. The lighter blue areas (d) are the areas (light cones) of potentially directly extracted light. In sum in each direction this amounts to about 10 % of all light generated within the structure. In total 63 % of all light is within these areas. Light in between (t) is trapped within the structure as shown in the upper drawing (dashed line). In cross section (bottom) one finds that, except for regions close to side facets, only a very small portion of laterally emitted light is directly extracted (less than 1 %, darker blue area) and that all other light undergoes multiple reflections before extraction (lighter blue area) or is trapped within the structure (white areas in between) (Reprinted from Journal of Crystal Growth, 298, Baoshun Zhang, Hu Liang, Yong Wang, Zhihong Feng, Kar Wei Ng, Kei May Lau, High-performance III-nitride blue LEDs grown and fabricated on patterned Si substrates, 725–730, Copyright (2007), with permission from Elsevier)

Let's assume a LED with square or rectangular shape: For such a structure many light paths exist where the light wave is trapped inside the LED (Fig. 4). In a square or rectangular LED structure one also has to take into account that six light cones with an angle of $\sim 38^\circ$ from the surface normal exist (critical angle of total reflection). Thus only 64 % of light generated at the active region can be extracted directly assuming 100 % reflecting back-metallization and no internal absorption losses. Typically surface roughening or special surface coatings are applied to minimize losses by multiple reflections.

For an LED with a triangular shape as applied, e. g., by Sora for GaN on GaN substrate LEDs the situation is different. While only five light cones and thus only around 53 % of light can be directly extracted (only if the backward light cone is fully reflected) all other light is not trapped by the side facets but usually extracted after one reflection at a side facet (Fig. 5). Still for laterally propagating light the

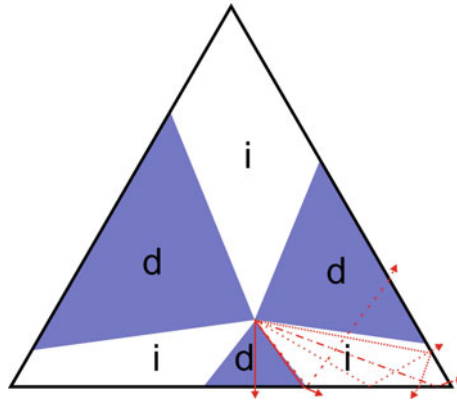


Fig. 5 Light paths in a triangular shaped chip in plan-view for an arbitrary light emission position. There are in total five light cones perpendicular to the surfaces, three of them shown in this sketch in plan-view assuming the chip being placed in epoxy resin ($n \sim 1.5$). From them light is directly (d) extracted (e.g., for the examples as full red lines). All other laterally propagating light is indirectly extracted (i). These light rays are first totally reflected at the outer GaN/plastic interface. After this reflection, almost all of the photons hit an outer surface in an angle below the angle of total reflection as demonstrated for some light paths (*dashed lines*). Although there are the same problems with laterally propagating light when the *top* and *bottom* interfaces are not highly reflective the amount of light potentially being extracted is increased due to the absence of trapped light paths

problem of multiple reflections exists. But with a thick substrate and a small chip size multi reflections at the upper and lower interface and the absorption losses within the structure are minimized.

GaN on Si LED Growth

Epitaxial growth of commercial GaN LEDs is usually performed by MOVPE. This thin film growth method requires high growth temperatures above $1000\text{ }^{\circ}\text{C}$ to achieve high quality GaN layers. Because of the large differences in thermal expansion coefficients between GaN and Si of around 54 % and the stiffness of the group-III-nitrides cracking of GaN layers typically occurs at layer thicknesses already below $1\text{ }\mu\text{m}$. But in sum the functional layers of a typical LED structure require more than $2\text{ }\mu\text{m}$ in thickness. For contacting and current spreading in the n-type layer ($>2\text{ }\mu\text{m}$), the MQW active region ($50\text{--}100\text{ nm}$), and the p-type layer ($100\text{--}200\text{ nm}$). To this the buffer layers and further functional layers, e.g., for surface structure etching, must be added. In sum a typical device grown on sapphire is around $4\text{--}5\text{ }\mu\text{m}$ in thickness and the same thickness is targeted for layers on Si leading to heavy crack formation if no countermeasures are taken.

In addition a high number of defects is present in the buffer layer due to a high lattice mismatch of 19 % between the AlN(0001) nucleation layer and the Si(111)

substrate (Krost and Dadgar 2002; Liu et al. 2003). This leads to a poor in-plane (twist) alignment of the AlN layer while the tilt usually is rather low. Twisted and tilted islands then lead to edge as well as screw type dislocation formation upon coalescence, respectively. As a consequence the dislocation density is high, often around 10^{10} cm^{-2} within the first hundred nanometers of GaN growth. To achieve a dislocation density below 10^9 cm^{-2} dislocation reduction schemes (stress gradients, interfaces, masking layers, epitaxial lateral overgrowth) or at least thicker layers are required.

For mass production and to achieve device grade material quality (dislocation density in the low 10^8 cm^{-2}) a high reproducibility of the growth process is of utmost importance. Here it has been observed that the reproducibility of the GaN growth process on silicon is strongly depending on the growth system used. While for some it is easy to maintain a high throughput and achieve a high reproducibility in others after each growth process a cleaning procedure must be applied. This is mostly due to deposits and the high reactivity of the Si surface. Upon heating deposits can evaporate and alter the Si surface. This deteriorates a proper nucleation of the first AlN layer. Meltback etching is also a problem when particles are present in the MOVPE reactor chamber from growth deposits. They must be avoided in any case.

Prerequisite for thicker layers is strain engineering to avoid cracking of GaN. Because strain engineering does impact the layer structure and with it device processing it is briefly described in the following section.

Further details on MOVPE growth of GaN on Si have been described, e.g., in (Li et al. 2010; Dadgar and Krost 2013).

Strain Engineering

Today several strain engineering methods are available. They are based on two different principles. First of all by growth in areas small enough that tensile stress does not lead to cracking (section “[Selective Epitaxy](#)”). Secondly, by growing compressed layers with sufficient compression to counterbalance all or the biggest part of tensile stress generated upon cooling. While for the latter there are several varieties in layer schemes, compositions, etc. only the two with the highest significance are briefly explained in sections “[AlGaIn Buffer](#)” and “[AlN Interlayers](#).”

In all cases thermal stress will bend the substrate after cooling from MOVPE growth temperature. This can be counterbalanced by well balanced strain engineering layers and also minimized by using Si substrates thicker than SEMI standard. The latter is also advised to postpone plastic substrate deformation. To prevent this substrates should be also highly doped (Dadgar et al. 2013). For a 4–5 μm thick LED structure on a Si substrate which remains free from plastic deformation while sufficient pre-stress during the growth process was applied to yield in a flat wafer free after cooling, the wafer thickness should be at least 1000 μm , depending on the growth temperature and purity of the substrate.

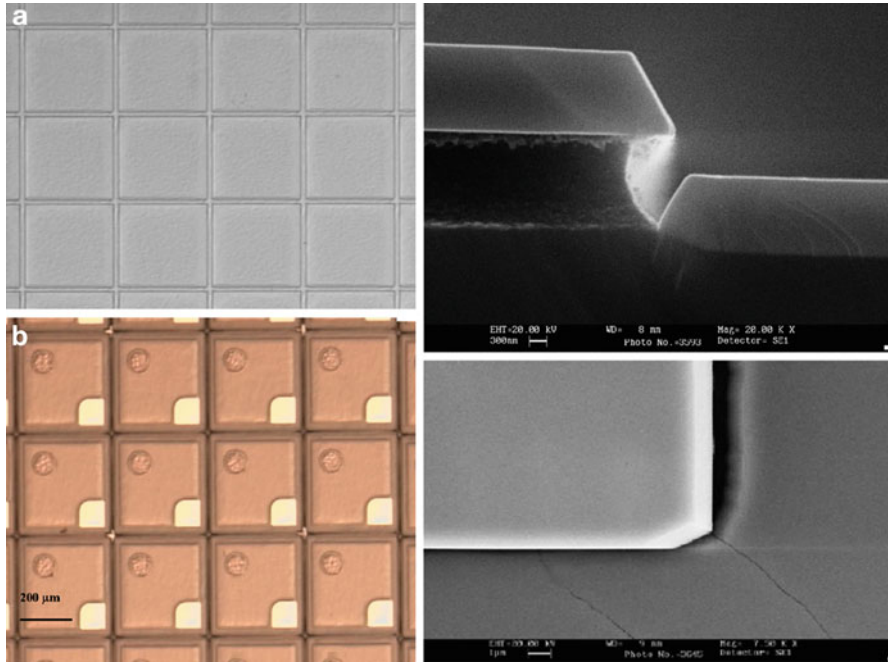


Fig. 6 Left: Top view on a selectively grown LED structure with an AlN nucleation layer/in-situ SiN masking/ $1\ \mu\text{m}$ GaN/LT-AlN/ $1\ \mu\text{m}$ GaN/MQW/p-GaN cap before (a) and after (b) processing. Right: cross-sectional (a) and plan-view (b) SEM images showing the trench etched Si substrate, the LED structure and its lateral overgrowth (a). Cracks appear in the deeper region in between the square fields (From Zhang et al. (2007))

For many steps during processing as, e. g., lithography a bow value of less than $30\ \mu\text{m}$ is often required. When a thin film LED process is performed also wafer bonding requires a low bow, ideally a flat wafer to ensure a high process yield. This is also supported by thicker growth substrates.

Selective Epitaxy

Selective epitaxy requires structuring of the substrate or a buffer layer on a substrate to grow structures which are small enough that tensile stress does not lead to cracking after cooling from growth temperature. From a simple viewpoint the structures must be smaller than the average crack distance in a continuous layer. In reality the dimensions turn out to be significantly larger which might be attributed to relaxation within the Si substrate (e.g., cracks are often observed in areas in between the overgrown regions, see Fig. 6 right). Selective growth can be performed either by a masking layer to form openings where device layer growth occurs (Dadgar et al. 2001a) or by etching the substrate with deep trenches leading to a discontinuous layer when the etch depth is sufficiently deep (Fig. 6; Zhang et al. 2007).

Here strain engineering buffer layers as described in the following section are beneficial to increase the possible growth area significantly. But such strain engineering layers can be reduced in thickness and amount and thicker layers without any further interruption are possible. The area size achievable is up to 1 mm^2 with layer thicknesses above $2 \text{ }\mu\text{m}$. The most difficult issue when using a mask is growth enhancement at the edges of the structures due to precursor diffusion on the masked layer. To reduce this either as narrow as possible masked regions between the LED layers or growth on etched substrates is indicated. For the latter the difficulty is the requirement to avoid melt-back etching, especially at the edges of the fields.

Several authors presented LEDs based on this method. Yang et al. presented one of the first LEDs on structured Si by metalorganic vapor phase epitaxy (MOVPE) (Yang et al. 2000), however with cracks. A first crack-free LED based on this method but only $100 \times 100 \text{ }\mu\text{m}^2$ in size was presented in the following year (Dadgar et al. 2001a, b, 2003). There an AlGaIn / GaN multilayer in the buffer also intended to act as Bragg-mirror for enhanced vertical light emission reduced tensile stress and helped in avoiding cracks in the up to $3.6 \text{ }\mu\text{m}$ thick structure. A $1.5 \text{ }\mu\text{m}$ thick LED structure with larger size ($200 \times 200 \text{ }\mu\text{m}^2$) was presented in the next year by Honda et al. (2002). The next big step was the demonstration of a relatively bright $2 \text{ }\mu\text{m}$ thick LED based on this growth method by Zhang et al. in 2007 (Zhang et al. 2007). This device showed an output power of 0.7 mW at 20 mA and 470 nm . Nevertheless, an identical structure on sapphire showed a 3–4 times higher output power, which can be expected due to absorption by the Si substrate. Latticepower has transferred this process to production and follows a thin film concept to minimize internal absorption (Jiang et al. 2009).

AlGaIn Buffer

AlN and AlGaIn have a smaller in-plane lattice parameter than GaN. By growing a layer with smaller in-plane lattice parameter the subsequent GaN (or AlGaIn layer with higher Ga content) can be under compression when it grows pseudomorphically or only partially relaxed. This was upon the first methods to increase the thickness of crack-free layers on silicon substrates and it is a logical consequence to use an AlGaIn intermediate layer when starting with AlN growth and aiming to grow GaN (Marchand et al. 1999; Ishigawa et al. 1999a, b; Kim et al. 2001). The composition is graded by different schemes as continuously or stepped with different impact on material quality and compression induced on the subsequently grown GaN layer. A high internal compression in the AlGaIn layers during growth as well as strain gradients at interfaces are also beneficial to incline or bend dislocations and increase their recombination probability. For large changes in lattice parameter often a higher degree in relaxation than for smaller changes is observed. Because a high compression of the layers is wanted well balancing the thickness and composition of each layer is required.

Today this method is usually always applied for GaN on Si growth but it is typically limited to GaN layer thicknesses of less than $3 \text{ }\mu\text{m}$, often too little for high quality LEDs.

AlN Interlayers

AlN interlayers, usually grown at low temperature (500–900 °C), were first applied on sapphire by Amano et al. (1998) to improve GaN material quality and enable crack free thick AlGaIn growth on GaN buffer layers. On Si they were also applied to avoid cracking of GaN (Dadgar et al. 2000). The about 10 nm thin layers enable crack elimination by inducing compression on the subsequently grown GaN layer. The mechanism is by full or partial relaxation of the AlN layer on top of a GaN layer (Bläsing et al. 2002; Reiher et al. 2003; Fritze et al. 2012a). The subsequent GaN layer ideally grows pseudomorphically and, because of the smaller in-plane lattice parameter of the interlayer, under compression. Usually some relaxation is observed at the upper AlN / GaN interface but less then at the lower interface which yields to a net compression in the top GaN layer.

Also alloying some GaN to AlN can improve the layer quality, a possible origin is the different surface structure and less relaxation at the upper interface. Fritze et al. have also demonstrated that the pre-stress of the GaN layer has an impact on the role of the interlayer (Fritze et al. 2012a): From a strain engineering layer, if no pre-stress is present, to a dislocation reducing interlayer, if a high compressive pre-stress is already present during growth.

Figure 7 shows the progress in LED layer structures from the first thick crack-free LED in 2002 to a crack-free LED structure in 2013 which only requires one interlayer and has a completely different growth scheme. Nowadays an AlN/AlGaIn buffer already provides some compression and better blocks meltback-etching to enable overall thicker layers which increases the freedom in LED design. The impact

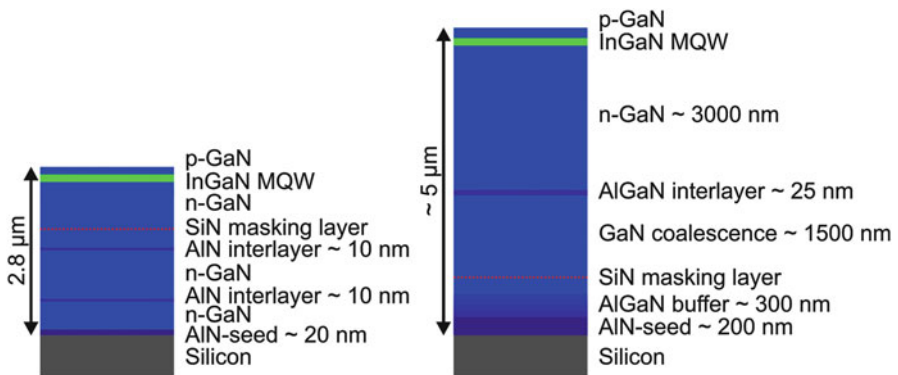


Fig. 7 Scheme of a crack-free LED structure in 2001/2002 (see also Fig. 8) with two AlN interlayers and an only 1–1.5 μm thick final n-GaN layer (*left*). Optimized LED structure as developed by OSRAM OS (2013) with only one strain engineering AlGaIn layer and 3 μm of uninterrupted GaN growth. Dislocation densities of the structures were in the low 10^9 cm^{-2} (*left*, 2002) and low 10^8 cm^{-2} (*right*, 2013). The structure on the right provides sufficient material for a thin film process where the buffer below the AlGaIn interlayer can be etched to provide a roughened surface for light extraction

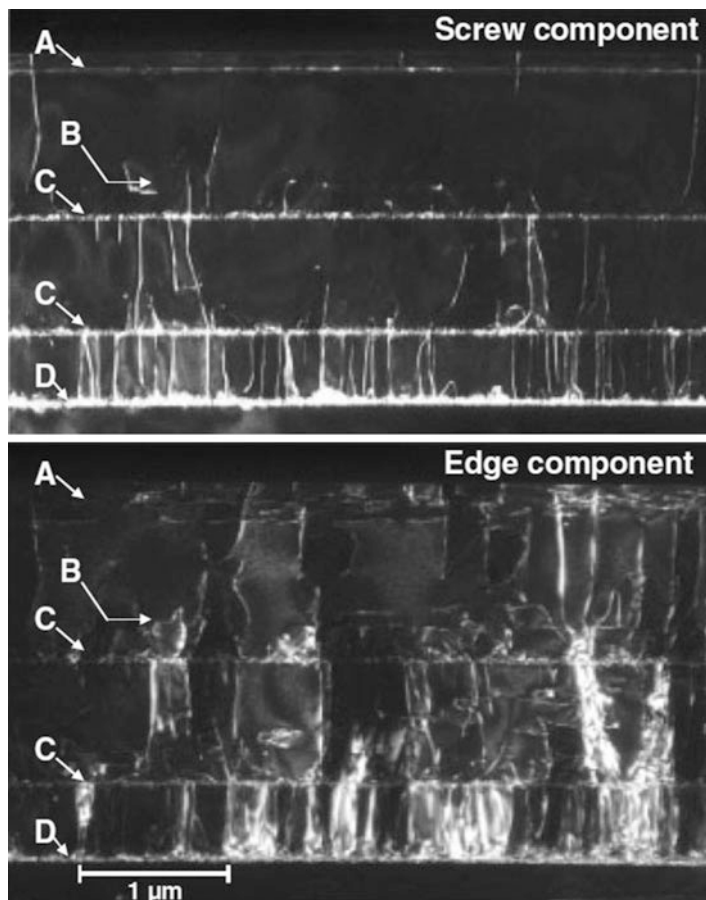


Fig. 8 TEM of a crack-free LED structure as grown in 2001/2002 (Fig. 7 left) with active layer A, SiN in-situ masking B for dislocation reduction and LT-AlN layers C for strain engineering and dislocation reduction on an AlN nucleation layer D on Si(111) substrate. Especially dislocations with a screw component are reduced at the AlN layers and SiN mask (TEM from Dadgar et al. (2002a) (Reprinted from *physica status solidi (a)* applications and materials science, A. Dadgar, M. Poschenrieder, O. Contreras, J. Christen, K. Fehse, J. Bläsing, A. Diez, F. Schulze, T. Riemann, F.A. Ponce, A. Krost, Bright, Crack-Free InGaN/GaN Light Emitters on Si(111), 6, Copyright (2002), with permission from Wiley)

of different layers as AlN and SiN in-situ masking on the edge and screw type dislocation density is visualized by transmission electron microscopy (TEM) in Fig. 8 (Dadgar et al. 2002a).

Because an interlayer can be applied again and again in a layer stack it enables very thick layer thicknesses as, e.g., a 14.5 μm thick GaN / LT-AlN layer stack with a final 4.5 μm thick GaN layer (Dadgar et al., *Phys. Stat. Sol. C* 8, 1503 (2011)). For such thicknesses a high pre-stress during growth is required that usually leads to plastic substrate deformation. Usually GaN layer thicknesses should be as low as

possible and necessary but a low dislocation density is a must for high quality LEDs thus a balance must be found and careful layer optimization is required. The progress demonstrated in the structures of Fig. 7 is mostly due to improvements in nucleation layer quality, introducing and optimizing thickness and composition of AlGaIn layers as well as the interlayer. Apart from SiN masking and other dislocation reducing layers which increase the overall thickness one of the most important layers to determine GaN quality is the nucleation layer which must be optimized first Volz et al. (2014).

Semi-polar LED Growth

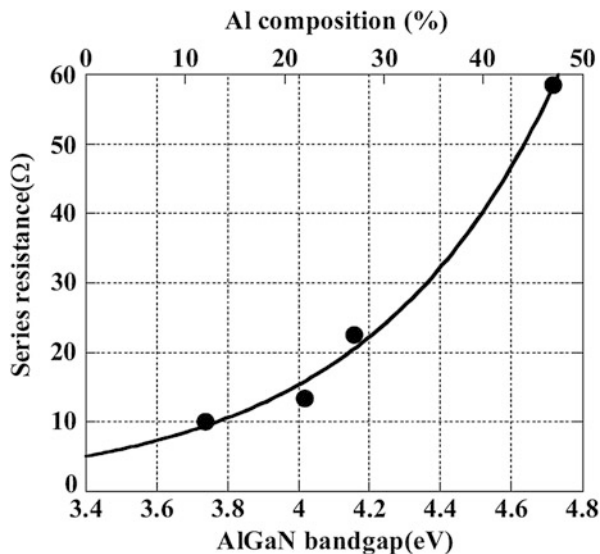
C-axis oriented GaN based LEDs usually suffer from the quantum confined Stark effect which reduces the radiative recombination rate, especially for longer wavelength emission. This can be avoided or at least reduced when using non- or semi-polar GaN orientations (Waltereit et al. 2000; Romanov et al. 2006). Such GaN orientations can be grown on Si either using thin low-temperature AlN nucleation layers on Si(111) ($h > 1$) (Ravash et al. 2009; Ravash et al. 2010, 2011) or by patterning non Si(111) oriented substrates to achieve Si(111) facets inclined to the surface normal. In this case MOVPE growth will result in c-axis oriented GaN on the Si(111) facets. Because the Si(111) facets are tilted to the surface normal this results in semi- or non-polar GaN layers (Chiu et al. 2011; Hikosaka et al. 2007; Murase et al. 2011; Tanikawa et al. 2008a, b; Sawaki and Honda 2011). In this method also lateral growth is performed resulting in a low dislocation density and because of the initial c-axis oriented growth it minimizes stacking fault formation.

A further benefit of this method is enhanced light scattering at the interface to the patterned Si substrate. LEDs were also demonstrated and showed a reduced droop which was attributed to the reduced pyro- and piezoelectric fields (Chiu et al. 2011).

Processing of GaN-on-Si LEDs

While by epitaxial growth on silicon the influence on light extraction is rather small, unless growth is performed on a three-dimensional patterned substrate, processing of LEDs is the most important part in regard of light extraction enhancement. By layer growth itself Bragg-mirror layers are suited for enhanced light extraction when a resonant cavity LED is realized (Ishikawa et al. 2004a, b) as are pre-patterned substrates which do either act as photonic bandgap structure (Orita et al. 2008) or simply enhance light scattering and with it light extraction probability. However, the growth of such layers is quite demanding and effort and gain usually is against the growth of them, especially since other methods exist which are also well suited. Also buffer layers as TiN have been tested but their reflectivity is about equal to Si in the blue wavelength region and reaches around 80 % in the yellow/red wavelength region (Chen et al. 2006). While there is no benefit in the blue wavelength region the

Fig. 9 Series resistance of a Si/AlGaN contact for different Al-content (From Honda et al. (2007))



benefit at longer wavelengths is low if only a simple LED (see, e.g., Fig. 9 left) is processed due to absorption losses already described in section “[Light Propagation and Extraction](#).”

Thermal and Electrical Considerations

For high power operation a good heat conduction and distribution is crucial. In general when very high wall plug efficiencies are achieved for GaN based LEDs (Narukawa et al. 2010) heat generation from losses gets less important and with further progress taking place in LED development this issue might get a minor one in the future. But today standard devices still generate enough losses and the generated heat can play an important role in regard of lifetime. This is not only the direct degradation of the semiconductor layer but also of encapsulation and by delamination of chips from their carrier due to overheating or excessive thermal stress by on-off-cycling of the device.

While this is a more general topic not specific for layers grown on Si substrates internal heat conduction can differ for layers grown on sapphire and on silicon. It has been already demonstrated that interfaces, especially between AlN and Si, lead to a reduction in heat conduction due to phonon mismatch (Kuzmík et al. 2005). But also ternary compounds are known to have a lower thermal conductivity than binary compounds. These two limiting factors might negatively influence high power device performance when a high number of (mismatched) interfaces or thick ternary compound layers are required within the structure and are between the heat generating layers and a heat sink. As already described ternary compound layers and additional interfaces are often required for strain engineering layers to achieve crack-free GaN on Si, thus can't be completely avoided.

With typical layer schemes on Si substrates these limiting factors are, however, no concern for thin film LEDs with p-side down mount. But they can be an issue for devices kept on the Si epi-substrate when operated at high current densities. Nevertheless, until now this is not a reported problem, which might be due to the typical fabrication of GaN on Si LED structures as thin film LED.

In regard of electrical performance limitations by the Si/III-N interface more research work has been conducted. Honda et al. demonstrated (2007) that the Si/AlN interface has a barrier height of 2.8 ± 0.4 eV (Ishikawa et al. 2003) leading to a substantial voltage drop which can be circumvented when using a thin (2 nm) AlN layer which forms a tunneling contact. In a systematic study Honda et al. also show that AlGaIn as nucleation layer reduces the contact resistance but high Ga-contents are required for low losses at this interface (Fig. 9). Unfortunately Gallium does increase the tendency of melt-back etching which makes this method less suited for thick layers in mass production. And also a very thin AlN nucleation layer, as demonstrated by Ishikawa et al., usually is not sufficient to protect the substrate from the melt-back etching reaction when thicker GaN layers are grown.

With another material combination a better contact resistance without the mentioned problems might be achieved. Here it is expected that for p-Si / n-InGaIn contacts with In contents around 40 % a tunneling contact is formed (Hsu and Walukiewicz 2008; Ager et al. 2009), which should result in low resistance across the interface. However, such layer scheme has not been demonstrated in sufficiently high quality for LED use yet and again melt-back etching is also a topic when the layers are grown by MOVPE. An alternative could be an AlInN layer with high In content around 50 % which helps in reducing meltback etching. In principle AlInN can be applied as nucleation layer on Si but again high In concentrations and a high material quality are difficult to achieve and further progress in material growth is required before such layers can be successfully applied.

AlN and AlGaIn interlayers as used for strain engineering are also a concern when current flows across them. Due to their large bandgap and the band offset a voltage drop of typically 0.5 V occurs at each layer increasing the internal heat up and decreasing the wall plug efficiency of a device. Thus in LED design such current paths must be avoided (see, e.g., the progress made in Fig. 7).

On Wafer LEDs

LED structures which are processed to devices with the Si growth substrate attached (Fig. 10) usually suffer from the before mentioned problems of light extraction. Using anti-reflecting top layers or roughening the p-GaN contacting layer can help to significantly increase brightness with an upper theoretical limit below 70 % extraction efficiency. Therefore this method is well suited for medium power LEDs but not for high brightness LEDs for general lighting purposes. Especially roughening the p-GaN layer is demanding with this layer usually only being 100–200 nm thick. Structuring ideally requires a thickness of more than 500 nm which then results in an increased series resistance due to the low conductivity of p-GaN. To minimize

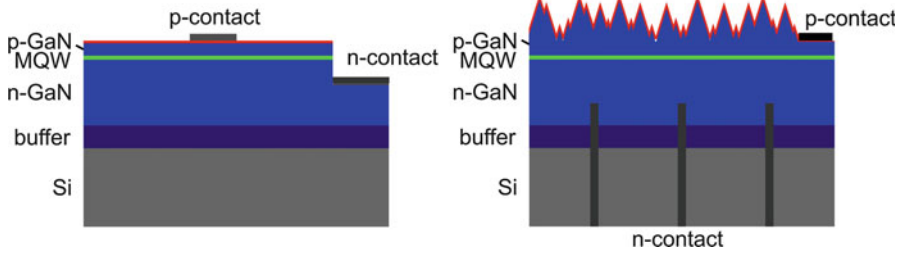


Fig. 10 Planar GaN on Si LED (*left*) with low extraction efficiency due to the losses described in section Light Propagation and Extraction and sketched in Figs. 1 and 4. An improvement can be achieved with surface roughening (*right*) which requires a thicker p-GaN layer. A top contact (red) with low absorption as, e.g., ITO is also required to minimize losses at the contact. In the example on the *right* an n-type contact realized by via etching through the Si substrate is applied. This minimizes losses of surface area, improves current spreading and also enables simpler mounting with only one bond wire required

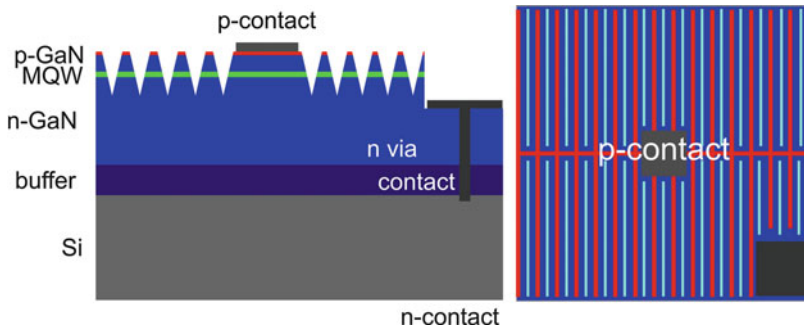


Fig. 11 Possible GaN on Si LED design for LEDs on Si substrates with improved light extraction efficiency (*left*: cross-section, *right*: plan-view). Multiple interior reflections are minimized by deep etched, ideally inclined, trenches. A high aspect ratio of etching depth to column or trench width is required to minimize losses as a roughened surface of the sidewalls is beneficial if long trenches are used. Here also a backside n-contact has been realized by etching through the GaN layer to the Si substrate

absorption from p-contact metallization an ITO layer instead of a conventional Ni/Au contact metallization is also advised (Narukawa et al. 2006).

Similar to surface roughening also small structures in the upper nanometer to low micrometer range, best etched through the MQW region and ideally with inclined sidewalls, can help to minimize absorption at the substrate and increase light extraction (Fig. 11). Here the main difficulty is p-contact formation and the reduced area of the active region.

For a low operation voltage of such structures different n-type contacting schemes can be applied:

- N-type top contacts after etching from the top to the n-type layer (Fig. 10 left). This requires a second bond wire.
- N-type via contacts etched from the back through the Si substrate and part of the III-nitride layers (Fig. 10 right).
- N-type back contacts etched from the front with a recess etched n-GaN layer to the Si substrate (Fig. 11) (Wei et al. 2010).

The main benefit of LEDs on Si substrates is their high electrostatic discharge strength. This is one reason why Sanken electric was selling low power blue GaN on Si LEDs for many years, which was the first commercial GaN on Si LED.

Thin Film LEDs

In LED manufacturing of GaN on sapphire grown LED structures two main routes have been established. One is the growth on structured sapphire which improves the material quality by lateral overgrowth and enhances light scattering and with this light extraction (Yamada et al. 2002), the other is thin film processing (Schnitzer et al. 1993; Haerle et al. 2004; Zhang et al. 2005; Zou et al. 2010; Egawa and Bin Abu Bakar Ahmad 2010; Alarcón-Lladó et al. 2010; Shaohua et al. 2013). Possible process steps for thin film manufacturing are sketched in Fig. 12. During this process the III-nitride layer stack is completely removed from the growth substrate after being bonded with its p-side with a highly reflecting metallic mirror onto a new carrier. Etching the now exposed nitrogen face of the III-N layer with an alkaline etch leads to a roughened top layer often with $\{10\bar{1}1\}$ facets with an angle to the surface plane close to 60° , best suited for high light extraction efficiencies (see Fig. 5).

For GaN on Si layers this is also the simplest and most efficient method to achieve a high brightness and wall plug efficiency. Here the p-contact layer covers the whole p-GaN layer ensuring ideal current distribution, usually one major bottleneck of conventional top contacted LEDs.

When Si carriers are used for thin film processing thermal matching between the Si substrate and carrier enables a layer transfer via a metal bond at elevated temperatures with little risk for cracking. However, thermal stress during device operation remains an issue. Thus also other concepts as copper carriers have been tested (Zhang et al. 2005; Dolmanan et al. 2011; Alarcón-Lladó et al. 2010). Copper, as most metals, has a significantly higher coefficient of thermal expansion which could lead to higher stresses upon operation. But its thermal conductivity is also much better than that of a Si carrier. Therefore the operation temperature of the LED can be lower which reduces thermal stresses between the thermally mismatched carrier and III-N layer.

For n-type contacts the use of a smaller contact area is less critical than for p-type contacts because of the significantly lower resistivity of the contacts and the layers. Depending on the structure processed the n-contact alloying temperature is limited, in the case of a thin film LED due to the metal bond to the new carrier. Therefore, as

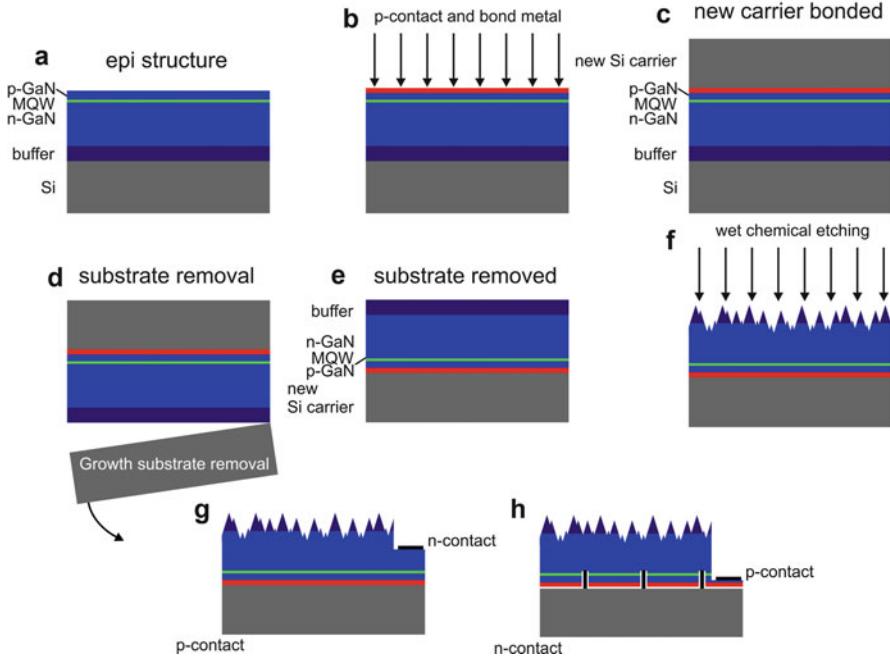


Fig. 12 Fabrication steps of a thin film LED: an epi structure (a) which is metalized with a highly reflecting metal in (b) to form a p-type contact which is covered with a thick bond metal. After bonding to a carrier in (c), preferably onto another Si substrate, the growth substrate is removed by grinding and wet chemical etching resulting in a N-face up LED structure. This structure can then be etched by alkaline etchants to form a pyramidal surface structure to enhance light extraction. Possible LED designs are shown in (g) and (h). (g) With a p-type back and n-type front contact as, e.g., the OSRAM ThinGaN[®] technology, and (h) a structure with minimized light absorption and n-type back and p-type front contact as realized, e.g., with the OSRAM UX:3[®] technology

high as possible electron carrier concentrations are required to enable a low contact formation temperature. For this purpose Ge doping has been demonstrated to be superior to Si doping because of a more than one order of magnitude higher electron concentration achievable (Dadgar et al. 2011; Fritze et al. 2012b) without layer degradation as well as the absence of tensile stress generation by doping induced edge type dislocation inclination during layer growth (Dadgar et al. 2004).

For thin film LEDs light extraction is mostly limited by losses due to the imperfect reflectivity of the Ag layer and the n-type contact metal. To enhance light extraction surface roughening by alkaline etchants usually leads to irregular sized pyramidal shaped surfaces. The etch depth as well as size and shape distribution is important for efficient light extraction and must be optimized. External quantum efficiencies above 60 % can be achieved today (Drechsel 2014). To achieve high wall plug efficiencies the operation voltage must be also as low as possible. By optimizing layers and doping contacts values below 3 V at nominal operation current (Drechsel 2014) can be achieved. For GaN on Si layers this means that low sheet

resistivities are required. Thus contacting must be performed in a depth of the structure where no strain engineering AlGa_N or AlN layer can increase the operation voltage (typically around 0.5 V for a ~10 nm AlN interlayer).

Closing Remarks

One argument for GaN on Si LEDs to reduce production cost has always been the possibility to use depreciated Si wafer fabs for fabrication. Until now it has, however not been demonstrated that the LED process is compatible with CMOS processing technology. Thus for a wafer fab that is not only intended to be used for GaN processing cross contamination from group-III elements and processing has to be considered as a potential risk which might exclude this approach. But if, also an integration of III-V optoelectronics with Si electronics will be opened.

What can be excluded today is a lowered lifetime of GaN on Si LEDs. In the beginning it was argued that strained layers can lead to a decreased lifetime. Nevertheless there has been no report that this is the case and in our lab out of 53 blue prototype GaN on Si LEDs in 5 mm plastic housings that were nearly continuously operated for 10 years at a current between 20 and 30 mA, 16 LEDs (partially) failed (most of them in the last 4 years) while the rest remained bright (Fig. 13).

In conclusion GaN on Si grown LED structures for high brightness LEDs are best manufactured with a thin film process. The presently achieved performance values demonstrate that after more than 10 years of development GaN on Si has evolved from a material full of cracks and with low output power to a real alternative to GaN on sapphire. However, to be successful growth of GaN on Si requires highly reproducible MOVPE growth systems and proper optimization of layers best using in-situ curvature control for stress monitoring. Although it is unlikely that GaN on Si will fully replace other substrates for LED growth it will most likely play a bigger role also promoted by developments in other fields as high power GaN on Si electronics.

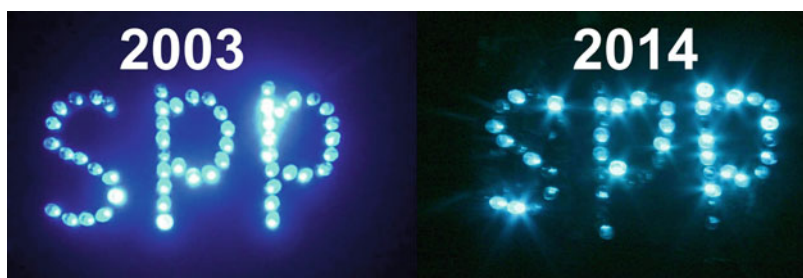


Fig. 13 53 blue-green GaN on Si prototype LEDs arranged in 2003 for the cover of *physica status solidi c* 0, No. 6, 1565–1949 (2003) and their aging after nearly 10 years of uninterrupted operation at 20–30 mA/LED @3.7 V. In the last 4 years LED failure started and until today 16 LEDs failed (operation voltage increased). Because of a narrow emission angle some LEDs do not appear as bright as they are. Different color stems from different cameras with different white balance settings

References

- Ager JW III, Reichertz LA, Cui Y, Romanyuk YE, Kreier D, Leone SR, Yu KM, Schaff WJ, Walukiewicz W (2009) Electrical properties of InGaN-Si heterojunctions. *Phys Status Solidi C* 6:S413
- Alarcón-Lladó E, Bin-Dolmanan S, Lin VKX, Teo SL, Dadgar A, Krost A, Tripathy S (2010) Temperature rise in InGaN/GaN vertical light emitting diode on copper transferred from silicon probed by Raman scattering. *J Appl Phys* 108:114501
- Amano H, Iwaya M, Kashima T, Katsuragawa M, Akasaki I, Han J, Hearne S, Floro JA, Chason E, Figiel J (1998) Stress and Defect Control in GaN Using Low Temperature Interlayers. *Jpn J Appl Phys* 37:L1540
- Bläsing J, Reiher A, Dadgar A, Diez A, Krost A (2002) The origin of stress reduction by low-temperature AlN interlayers. *Appl Phys Lett* 81:2722
- Bykov P (1972) Spontaneous Emission in a Periodic Structure. *Sov Phys JETP* 35:269
- Chen NC, Lien WC, Shih CF, Chang PH, Wang TW, Wu MC (2006) Nitride light-emitting diodes grown on Si (111) using a TiN template. *Appl Phys Lett* 88:191110
- Chiu C-H, Lin D-W, Lin C-C, Li Z-Y, Chang W-T, Hsu H-W, Kuo H-C, Lu T-C, Wang S-C, Liao W-T, Tanikawa T, Honda Y, Yamaguchi M, Sawaki N (2011) Reduction of Efficiency Droop in Semipolar (1101) InGaN/GaN Light Emitting Diodes Grown on Patterned Silicon Substrates. *Appl Phys Exp* 4:012105
- Compound Semiconductor, Lattice Power Corporation (2010) Silicon-based LEDs leap from lab to fab. 16(4):30
- Dadgar A, Krost A (2013) Epitaxial growth and benefits of GaN on silicon. In: Gil B (ed) III-nitride semiconductors and their modern devices. Oxford University Press, Oxford. ISBN 978-0-19-968172-3
- Dadgar A, Bläsing J, Diez A, Alam A, Heuken M, Krost A (2000) Metalorganic Chemical Vapor Phase Epitaxy of Crack-Free GaN on Si (111) Exceeding 1 μm in Thickness. *Jpn J Appl Phys* 39:L1183
- Dadgar A, Alam A, Riemann T, Bläsing J, Diez A, Poschenrieder M, Straßburg M, Christen J, Krost A (2001a) Crack-Free InGaN/GaN Light Emitters on Si(111). *Phys Status Solidi A* 188:155–158
- Dadgar A, Poschenrieder M, Bläsing J, Fehse K, Riemann T, Diez A, Christen J, Krost A (2001b) Bright Crack-Free $300\times 300\ \mu\text{m}^2$ InGaN Light Emitters on Si(111). MRS fall meeting, Boston, 14.7
- Dadgar A, Alam A, Christen J, Riemann T, Richter S, Bläsing J, Diez A, Heuken M, Krost A (2001c) Bright blue electroluminescence from an InGaN/GaN multiquantum-well diode on Si (111): Impact of an AlGaIn/GaN multilayer. *Appl Phys Lett* 78:2211
- Dadgar A, Poschenrieder M, Contreras O, Christen J, Fehse K, Bläsing J, Diez A, Schulze F, Riemann T, Ponce FA, Krost A (2002a) Bright, Crack-Free InGaN/GaN Light Emitters on Si (111). *Phys Status Solidi A* 192:308
- Dadgar A, Poschenrieder M, Bläsing J, Fehse K, Diez A, Krost A (2002b) Thick, crack-free blue light-emitting diodes on Si(111) using low-temperature AlN interlayers and in situ SixNy masking. *Appl Phys Lett* 80:3670
- Dadgar A, Strittmatter A, Bläsing J, Poschenrieder M, Contreras O, Veit P, Riemann T, Bertram F, Reiher A, Krtschil A, Diez A, Hempel T, Finger T, Kasic A, Schubert M, Bimberg D, Ponce FA, Christen J, Krost A (2003) Metalorganic chemical vapor phase epitaxy of gallium-nitride on silicon. *Phys Status Solidi C* 0:1583
- Dadgar A, Clos R, Strassburger G, Schulze F, Veit P, Hempel T, Bläsing J, Krtschil A, Daumiller I, Kunze M, Kaluza A, Modlich A, Kamp M, Diez A, Christen J, Krost A (2004) Strains and stresses in GaN heteroepitaxy – sources and control. In: Kramer B (ed) *Advances in solid state physics*, vol 44. Springer, Heidelberg, p 313
- Dadgar A, Hums C, Diez A, Bläsing J, Krost A (2006) Growth of blue GaN LED structures on 150 mm Si(111). *J Cryst Growth* 297:279
- Dadgar A, Bläsing J, Diez A, Krost A (2011) Crack-Free, Highly Conducting GaN Layers on Si Substrates by Ge Doping. *Appl Phys Exp* 4:011001

- Dadgar A, Fritze S, Schulz O, Hennig J, Bläsing J, Witte H, Diez A, Heinle U, Kunze M, Daumiller I, Haberland K, Krost A (2013) Anisotropic bow and plastic deformation of GaN on silicon. *J Cryst Growth* 370:278
- de Almeida RMC, Baumvol IJR (2000) Reaction-diffusion model for thermal growth of silicon nitride films on Si. *Phys Rev B* 62, R16255
- Dolmanan SB, Teo SL, Lin VK, Hui HK, Dadgar A, Krost A, Tripathy S (2011) Thin-film InGaN/GaN vertical light emitting diodes using GaN on silicon-on-insulator substrates. *Electrochem Solid-State Lett* 14:H460
- Drechsel P (2014) Metallorganische Gasphasen-Epitaxie von Gruppe III-Nitrid-basierten LED Strukturen auf Silizium. PhD thesis, HU-Berlin, Dr. Hut, Munich, ISBN 978-3843917995
- Egawa T, Bin Abu Bakar Ahmad Shuhaimi (2010) High performance InGaN LEDs on Si (111) substrates grown by MOCVD. *J Phys D Appl Phys* 43:354008
- Egawa T, Zhang B, Nishikawa N, Ishikawa H, Jimbo T, Umeno M (2002) InGaN multiple-quantum-well green light-emitting diodes on Si grown by metalorganic chemical vapor deposition. *J Appl Phys* 91:528
- Feltin E, Dalmasso S, de Mierry P, Beaumont B, Lahrière H, Bouillé A, Haas H, Leroux M, Gibart P (2001) Green InGaN Light-Emitting Diodes Grown on Silicon (111) by Metalorganic Vapor Phase Epitaxy. *Jpn J Appl Phys* 40:L738
- Fischer AM, Wu Z, Sun K, Wie Q, Huang Y, Senda R, Iida D, Iwaya M, Amano H, Ponce F (2009) Misfit Strain Relaxation by Stacking Fault Generation in InGaN Quantum Wells Grown on m-Plane GaN. *Appl Phys Exp* 2:041002
- Fritze S, Drechsel P, Stauss P, Rode P, Markurt T, Schulz T, Albrecht M, Bläsing J, Dadgar A, Krost A (2012a) *J Appl Phys* 111:124505
- Fritze S, Dadgar A, Witte H, Bügler M, Rohrbeck A, Bläsing J, Hoffmann A, Krost A (2012b) Role of low-temperature AlGaIn interlayers in thick GaN on silicon by metalorganic vapor phase epitaxy. *Appl Phys Lett* 100:122104
- Guha S, Bojarczuk NA (1998a) Ultraviolet and violet GaN light emitting diodes on silicon. *Appl Phys Lett* 72:415
- Guha S, Bojarczuk NA (1998b) Multicolored light emitters on silicon substrates. *Appl Phys Lett* 73:1487
- Haerle V, Hahn B, Kaiser S, Weimar A, Bader S, Eberhard F, Plössl A, Eisert D (2004) High brightness LEDs for general lighting applications Using the new ThinGaIn™ Technology. *Phys Status Solidi A* 201(12):2736
- Hikosaka T, Honda Y, Yamaguchi M, Sawaki N (2007) Al doping in (1–101)GaIn films grown on patterned (001)Si substrate. *J Appl Phys* 101:103513
- Honda Y, Kuroiwa Y, Yamaguchi M, Sawaki N (2002) Growth of GaN free from cracks on a (111) Si substrate by selective metalorganic vapor-phase epitaxy. *Appl Phys Lett* 80:222
- Honda Y, Kato S, Yamaguchi M, Sawaki N (2007) Series resistance in a GaN/AlGaIn/n-Si structure grown by MOVPE. *Phys Status Solidi C* 4:2740
- Hsu L, Walukiewicz W (2008) Modeling of InGaIn/Si tandem solar cells. *J Appl Phys* 104:024507
- Ishigawa H, Zhao GY, Nakada N, Egawa T, Soga T, Jimbo T, Umeno M (1999a) High-Quality GaN on Si Substrate Using AlGaIn/AlN Intermediate Layer. *Phys Status Solidi A* 176:599
- Ishigawa H, Zhao G-Y, Nakada N, Egawa T, Jimbo T, Umeno M (1999b) GaN on Si Substrate with AlGaIn/AlN Intermediate Layer. *Jpn J Appl Phys* 38:L492
- Ishikawa H, Yamamoto K, Egawa T, Soga T, Jimbo T, Umeno M (1998) Thermal stability of GaN on (111) Si substrate. *J Cryst Growth* 189–190:178
- Ishikawa H, Zhang B, Egawa T, Jimbo T (2003) Valence-Band Discontinuity at the AlN/Si Interface. *Jpn J Appl Phys* 42:6413
- Ishikawa H, Asano K, Zhang B, Egawa T, Jimbo T (2004a) Improved characteristics of GaN-based light-emitting diodes by distributed Bragg reflector grown on Si. *Phys Status Solidi A* 201:2653
- Ishikawa H, Zhang B, Asano K, Egawa T, Jimbo T (2004b) Characterization of GaInN light-emitting diodes with distributed Bragg reflector grown on Si. *J Cryst Growth* 272:322

- Ishikawa H, Jimbo T, Egawa T (2008) GaInN light emitting diodes with AlInN/GaN distributed Bragg reflector on Si. *Phys Status Solidi C* 5:2086
- Jiang F, Wang L, Wang X, Mo C, You X, Zheng C, Liu W, Zhou Y, Xiong C, Tang Y, Fang W, Lu B (2009) H7, 8th international conference on nitride semiconductors (ICNS-8), Jeju
- Kikuchi A, Kawai M, Tada M, Kishino K (2004) InGaN/GaN Multiple Quantum Disk Nanocolumn Light-Emitting Diodes Grown on (111) Si Substrate. *Jpn J Appl Phys* 43:L 1524
- Kim M-H, Do Y-G, Kang HC, Noh DY, Park S-J (2001) Effects of step-graded Al_xGa_{1-x}N interlayer on properties of GaN grown on Si(111) using ultrahigh vacuum chemical vapor deposition. *Appl Phys Lett* 79:2713
- Krost A, Dadgar A (2002) GaN-based optoelectronics on Silicon substrates. *Mater Sci Eng B* 93:77
- Kuzmík J, Bychikhin S, Neuburger M, Dadgar A, Krost A, Kohn E, Pogany D (2005) Transient thermal characterization of AlGaIn/GaN HEMTs grown on silicon. *IEEE Trans Electron Dev* 52:1698
- Li J, Lin JY, Jiang HX (2006) Growth of III-nitride photonic structures on large area silicon substrates. *Appl Phys Lett* 88:171909
- Li T, Mastro M, Dadgar A (eds) (2010) III–V compound semiconductors: integration with silicon-based microelectronics. CRC-Press, Boca Raton, Florida, USA, ISBN 978-1439815229
- Lin VKX, Tripathy S, Teo SL, Dolmanan SB, Dadgar A, Noltemeyer M, Franke A, Bertram F, Christen J, Krost A (2010) Luminescence Properties of Photonic Crystal InGaIn/GaN Light Emitting Layers on Silicon-on-Insulator. *Electrochem Solid-State Lett* 13:H343
- Liu R, Ponce FA, Dadgar A, Krost A (2003) Atomic arrangement at the AlN/Si (111) interface. *Appl Phys Lett* 83:860
- Marchand H, Zhang N, Zhao L, Golan Y, Rosner SJ, Girolami G, Fini PT, Ibbetson JP, DenBaars SP, Speck JS, Mishra UK (1999) Structural and optical properties of GaN laterally overgrown on Si(111) by metalorganic chemical vapor deposition using an AlN buffer layer. *MRS Internet J Nitride Semicond Res* 4:2
- Murase T, Tanikawa T, Honda Y, Yamaguchi M, Amano H, Sawaki N (2011) Drastic Reduction of Dislocation Density in Semipolar (1122) GaN Stripe Crystal on Si Substrate by Dual Selective Metal–Organic Vapor Phase Epitaxy. *Jpn J Appl Phys* 50:01AD04
- Narukawa Y, Narita J, Sakamoto T, Deguchi K, Yamada T, Mukai T (2006) Ultra-High Efficiency White Light Emitting Diodes. *Jpn J Appl Phys* 45:L1084
- Narukawa Y, Ichikawa M, Sanga D, Sano M, Mukai T (2010) White light emitting diodes with super-high luminous efficacy. *J Phys D Appl Phys* 43:354002
- Orita K, Takase Y, Fukushima Y, Usuda M, Ueda T, Takigawa S, Tanaka T, Ueda D, Egawa T (2008) Integration of Photonic Crystals on GaN-Based Blue LEDs Using Silicon Mold Substrates. *IEEE J Quan Electron* 44:984
- OSRAM OS (2012) Press release “Success in research: first gallium-nitride LED chips on silicon in pilot stage”
- Ravash R, Bläsing J, Hempel T, Noltemeyer M, Dadgar A, Christen J, Krost A (2009) Metal organic vapor phase epitaxy growth of single crystalline GaN on planar Si(211) substrates. *Appl Phys Lett* 95:242101
- Ravash R, Bläsing J, Dadgar A, Krost A (2010) Semipolar single component GaN on planar high index Si(11h) substrates. *Appl Phys Lett* 97:142102
- Ravash R, Bläsing J, Hempel T, Noltemeyer M, Dadgar A, Christen J, Krost A (2011) Impact of AlN seeding layer growth rate in MOVPE growth of semi-polar gallium nitride structures on high index silicon. *Phys Status Solidi B* 248:594
- Reiher A, Bläsing J, Dadgar A, Diez A, Krost A (2003) Efficient stress relief in GaN heteroepitaxy on Si(111) using low-temperature AlN interlayers. *J Cryst Growth* 248:563
- Romanov AE, Baker TJ, Nakamura S, Speck JS (2006) Strain-induced polarization in wurtzite III-nitride semipolar layers. *J Appl Phys* 100:023522
- Sawaki N, Honda Y (2011) Semi-polar GaN LEDs on Si substrate. *Sci China Technol Sci* 54:38
- Schnitzer I, Yablonoivitch E, Caneau C, Gmitter TJ, Scherer A (1993) 30% external quantum efficiency from surface textured, thin-film light-emitting diodes. *Appl Phys Lett* 63:2174

- Schulze F, Dadgar A, Krtschil A, Hums C, Reissmann L, Diez A, Christen J, Krost A (2008) MOVPE growth of blue In_xGa_{1-x}N/GaN LEDs on 150 mm Si(001). *Phys Status Solidi C* 5:2238
- Shaohua Z, Bo F, Qian S, Hanmin Z (2013) Preparation of GaN-on-Si based thin-film flip-chip LEDs. *J Semicond* 34:053006
- Strittmatter A, Krost A, Straßburg M, Türck V, Bimberg D (1999) Low-pressure metal organic chemical vapor deposition of GaN on silicon(111) substrates using an AlAs nucleation layer. *Appl Phys Lett* 74:1242
- Tanikawa T, Hikosaka T, Honda Y, Yamaguchi M, Sawaki N (2008a) Growth of semi-polar (11-22) GaN on a (113)Si substrate by selective MOVPE. *Phys Status Solidi C* 5:2966
- Tanikawa T, Rudolph D, Hikosaka T, Honda Y, Yamaguchi M, Sawaki N (2008b) Growth of non-polar (11 $\bar{2}$ 0)GaN on a patterned (110)Si substrate by selective MOVPE. *J Cryst Growth* 310:4999
- Tran CA, Osinski A, Karlicev RF Jr, Berishev I (1999) Growth of InGaN/GaN multiple-quantum-well blue light-emitting diodes on silicon by metalorganic vapor phase epitaxy. *Appl Phys Lett* 75:1494
- Tripathy S, Dadgar A, Zang KY, Lin VKX, Liu YC, Teo SL, Yong AM, Soh CB, Chua SJ, Bläsing J, Christen J, Krost A (2009) GaN-based deep green light emitting diodes on silicon-on-insulator substrates. *Phys Status Solidi C* 6:S822
- Umeno M, Egawa T, Ishikawa H (2001) GaN-based optoelectronic devices on sapphire and Si substrates. *Mater Sci Semicond Process* 4:459
- Volz K, Stolz W, Dadgar A, and Krost A (2014) Growth of III/Vs on Silicon:Nitrides, Phosphides, Arsenides, and Antimonides in *Handbook of Crystal Growth: Thin Films and Epitaxy*, Tom Kuech editor, Elsevier, Amsterdam, NL, ISBN: 978-0444633040
- Waltereit P, Brandt O, Trampert A, Grahn HT, Menniger J, Ramsteiner M, Reiche M, Ploog KH (2000) Nitride semiconductors free of electrostatic fields for efficient white light-emitting diodes. *Nature* 406:865
- Wei J, Zhang B, Wang G, Fan B, Yang L, Rao W, Huang Z, Yang W, Chen T, Egawa T (2010) Vertical GaN-Based Light-Emitting Diodes Structure on Si(111) Substrate with Through-Holes. *Jpn J Appl Phys* 49:072104
- Yablonoitch E (1987) Inhibited Spontaneous Emission in Solid-State Physics and Electronics. *Phys Rev Lett* 58:2059
- Yamada M, Mitani T, Narukawa Y, Shioji S, Niki I, Sonobe S, Deguchi K, Sano M, Mukai T (2002) InGaN-Based Near-Ultraviolet and Blue-Light-Emitting Diodes with High External Quantum Efficiency Using a Patterned Sapphire Substrate and a Mesh Electrode. *Jpn J Appl Phys* 41: L1431
- Yang JW, Lunev A, Simin G, Chitnis A, Shatalov M, Kahn MA, Van Nostrand JE, Gaska R (2000) Selective area deposited blue GaN-InGaN multiple-quantum well light emitting diodes over silicon substrates. *Appl Phys Lett* 76:273
- Zhang B, Egawa T, Ishikawa H, Yang L, Jimbo T (2005) Thin-film InGaN multiple-quantum-well light-emitting diodes transferred from Si (111) substrate onto copper carrier by selective lift-off. *Appl Phys Lett* 86:071113
- Zhang B, Liang H, Wang Y, Feng Z, Ng KW, Lau KM (2007) High-performance III-nitride blue LEDs grown and fabricated on patterned Si substrates. *J Cryst Growth* 298:725
- Zou XB, Liang H, Lau KM (2010) Light extraction enhancement from GaN-based thin-film LEDs grown on silicon after substrate removal using HNA solution. *Phys Status Solidi C* 7:2171

Thin-GaN LED Materials

Ray-Hua Horng

Contents

Introduction	150
Background of Light-Emitting Diodes	150
Wafer Transfer Technology for III–V Compound Semiconductors	150
Experiments	152
MOCVD-Grown LED Structures (GaN-Based LEDs)	152
Device Fabrication	153
Surface Texturing Techniques	157
Characterization of Vertical Mirror-Structure n-Side-Up Nitride-Based LEDs	158
Design of Mirror Structure for Vertical n-Side-Up Nitride-Based LEDs	158
Surface Texturing for Vertical n-Side-Up Nitride-Based LEDs	165
Characteristics of Vertical n-Side-Up Nitride-Based LEDs	169
Summary	175
Conclusions	176
References	176

Abstract

In this chapter, thin film GaN-based light emitting diodes (LEDs) fabricated into n-side-up and p-side-up LEDs on mirror-substrate structures using a combination of wafer bonding, laser lift-off and surface texturing techniques were described. The effects of Pd, ITO/Al, NiO/Ag, NiO/Ag/Ni, and NiO/Au/Ag mirrors on the n-side-up GaN/mirror/Si LED properties were studied. It was found that the characteristics of the vertical-conducting n-side-up GaN/mirror/Si LEDs with a NiO/Ag/Ni mirror structure showed the best performance than the other mirror ones. After the thermal anneal process, the specific contact resistance of NiO/Ag/Ni to p-GaN can be reduced. The output power of the n-side-up GaN/mirror/Si LED shows nearly three times in magnitude as compared with that of the original

R.-H. Horng (✉)
National Chung Hsing University, Taichung, Taiwan
e-mail: huahorng@dragon.nchu.edu.tw

GaN/sapphire sample. On the other hand, the p-side-up GaN LEDs were fabricated using a combination of omni-directional reflector (ODR) and double-sided textured surface (both p-GaN and undoped-GaN) techniques. An Essential Macleod program was used to simulate the optimum thickness of the ODR structure. The reflectivity value of ODR structure used in work can reach 99%. On the top-side textured surface, the p-type GaN with hexagonal cavities was grown under low temperature (LT) conditions using metalorganic chemical vapor deposition. The GaN LED with a suitable LT p-GaN cap layer thickness was also studied. Experimental results indicate that the GaN LED sample with the 200-nm hexagonal cavity GaN layer on the surface exhibits a 50% enhancement in luminance intensity. The luminance efficiency can be improved. This indicates that the thin-film structure can enhance the light extraction efficiency of GaN-based LEDs, especially for large chip sizes.

Introduction

Background of Light-Emitting Diodes

Light-emitting diode (LED) is a semiconductor device that emits near-ultraviolet, visible, or infrared light when an electric current passes through it. The first blue LED was fabricated by Pankove et al. in 1972 using III-nitrides materials with a metal in structure (Pankove et al. 1972). Since then, related researches went on continually. However, the device performance was limited by the poorly conducting p-type GaN. Until the late 1980s, Aksamski and Amano et al. developed the low-temperature buffer layer and low-energy electron beam interaction techniques to obtain conductive p-type GaN, the first GaN blue LED constructed of a real p-n junction, which greatly improved the device performance (Amano et al. 1986, 1989). In 1992, Nakamura et al. achieved conductive p-type GaN with high-temperature thermal annealing in nitrogen ambient (Nakamura et al. 1992). Details of the evolution of major LED development (Craford 1997). Full-color solid-state lighting has become a realization with these great progresses of techniques.

Wafer Transfer Technology for III-V Compound Semiconductors

Wafer Transfer Application in GaN LEDs

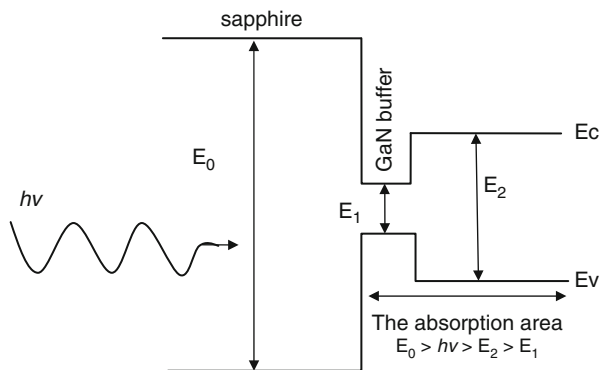
GaN-based semiconductors are attractive materials for LEDs and laser diodes in the blue-ultraviolet spectral region (Nakamura et al. 2000; Morita et al. 2002). These devices make new display technology or high-density data storage on compact disks be realized. Due to the lack of native substrates, films of GaN and related nitride compounds are commonly grown on sapphire wafers. The use of a sapphire substrate also complicates processing steps, such as formation of cleaved-edge facets and electrical backside contacts. The extraction of heat from an operating device through the sapphire substrate is also hampered. Therefore, free-standing GaN

optoelectronics without sapphire are most desirable. Thin-film laser lift-off (LLO) technique has recently been established as an effective tool for GaN-based heteroepitaxial structures, eliminated the sapphire constraint (Kelly et al. 1996; Wong et al. 1998, 1999, 2001; Chu et al. 2003). In 1998, Wong et al. used single 38 ns KrF excimer laser pulses directed through the transparent substrate in the range of 400–600 mJ/cm² to separate the GaN and sapphire substrate (Wong et al. 1998). The absorption of the 248 nm radiation by the GaN at the interface induces rapid thermal decomposition of the interfacial layer, yielding metallic Ga and N₂ gas. In 1999, Kelly et al. accomplished the free-standing GaN substrate grown on sapphire by hydride vapor phase epitaxy (Kelly et al. 1999). The thick GaN substrate was separated from the growth sapphire substrate by LLO, using a pulsed laser with 355 nm radiation to thermally decompose a thin layer of GaN at the film-substrate interface. Sequentially scanned pulses were employed and the lift-off was performed at elevated temperature (>600 °C) to relieve postgrowth bowing. In 2003, Chu et al. presented the performance of free-standing InGaN/GaN multiple quantum wells (MQWs) LEDs mounted on a Cu substrate (Chu et al. 2003). The InGaN/GaN MQWs LEDs structures, which grew on a sapphire substrate originally, are transferred to a Cu substrate by the LLO process into two different configurations, namely p-side-up and p-side down with the same Ni/Pd/Au p-contact layers. However, in these reports, either the n-side-up or p-side-up GaN light-emitting device is mounted on higher thermal-conductivity substrates, such as Si or Cu. The bottom reflector between the LED and absorbing substrate was not fully investigated. As mentioned above, the LLO technique is an essential method for the fabrication of wafer-transferred GaN/sapphire LEDs.

Laser Lift-Off (LLO) Process

A LLO process, first demonstrated by Kelly et al. (1997), with the incident beam directed through the transparent sapphire can be used to rapidly and effectively separate GaN thin films from its growth substrates. Since this first demonstration, many groups have also used LLO to separate GaN thin films from sapphire substrates. Other groups have reported the transfer of GaN-based LEDs, prefabricated on sapphire, onto copper and Si substrate (Tavernier et al. 1999a; Song et al. 1999a, b). However, the lift-off of a functioning on p-n diode using LLO from sapphire onto a different substrate is first demonstrated by Kelly et al. (1998). On the other hand, the LLO process has also been demonstrated on other thin-film materials system such as (Pb,La)(Zr,Ti)O₃, Pb(Zr,Ti)O₃, and ZnO in which the laser source is XeCl or KrF pulsed-excimer laser, or the third harmonic of a excimer laser to separate films deposited onto MgO substrate (Tsakalakos and Sands 1999, 2000; Tavernier et al. 1999b). In this study, the whole wafer (2 in. in diameter) is bonded to the Si by thermal-pressure bonding and then subjected to the LLO process. Figure 1 describes the fundamental mechanism of the LLO process, demonstrating the efficacy of the bandgap-selective absorbency using a KrF pulsed-excimer laser. Characterization of the GaN thin film before and after laser processing has been investigated and shows no detectable degradation in the GaN crystal quality after the lift-off and transfer processes.

Fig. 1 Schematic illustration of demonstrating the efficacy of the bandgap-selective absorptency



Ohmic Contact and High Reflection Mirror

In the vertical LEDs with mirror structure, two issues are very important. One is the ohmic contact with p-type or n-type layer and the other is the high reflection mirror. For the GaN-based LEDs, the high reflective mirror and ohmic contact with p-type GaN layer cannot be easily obtained simultaneously. For examples, Al and Ag are good ohmic-contact metals with the n-type GaN and with high reflectivity ($>90\%$), because the n-type GaN layer has high electron concentration. It is useful to fabricate the p-side-up wafer-transferred GaN with a high reflective mirror structure. However, for p-type GaN, it is difficult to obtain a low-resistance ohmic contact and high reflectivity at the same time due to the following two reasons. First, it is difficult to achieve a high hole carrier concentration in p-type GaN. The p-type GaN is typically achieved by doping magnesium (Mg) in MOCVD or molecular beam epitaxy. Mg has the relative high optical activation energy of around 250 meV compared to other acceptors (Strite and Morkoc 1992; Nakamura et al. 1991). The formation of Mg-H complexes during MOCVD growth is generally suggested to be responsible for the low activation efficiency of Mg that results in the relative low p-type GaN carrier concentration. Second, the absence of suitable metals that have higher work function than that of p-type GaN (~ 7.5 eV) resulting the specific contact resistance only around $10^{-2} \Omega\text{-cm}^2$. Therefore, the major issues in the n-side-up GaN epilayer-transferred structure concentrate on finding a suitable mirror for the p-type GaN layer.

Experiments

MOCVD-Grown LED Structures (GaN-Based LEDs)

The GaN LEDs wafers used in this work were grown on a *c*-face (0001) sapphire substrate by MOCVD. Trimethylindium, trimethylgallium, trimethylaluminum, and ammonia are used as the sources of In, Ga, Al, and N, respectively. Bis-cyclopentadienyl magnesium and silane are used as the p-type and n-type doping sources, respectively. The LEDs structure for 470 nm emission consists of a 30-nm-thick GaN

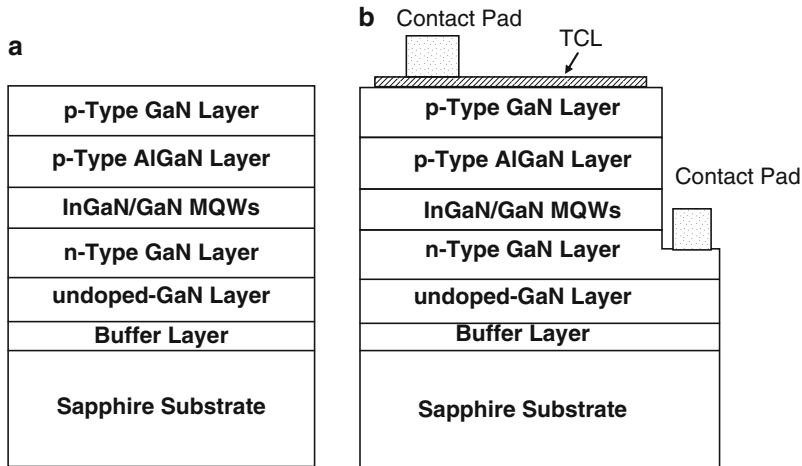


Fig. 2 Schematic drawing of the layer structure and device of (a) a conventional GaN LEDs structure, (b) a conventional GaN LEDs device

nucleation layer grown at 550 °C, a 1- μm -thick undoped GaN layer grown at 1,050 °C, a 3- μm -thick Si-doped n-type GaN layer grown at 1,050 °C, a 5-pair $\text{In}_{0.3}\text{Ga}_{0.7}\text{N}/\text{GaN}$ MQWs structure grown at 770 °C, a 50-nm-thick Mg-doped p-type $\text{Al}_{0.15}\text{Ga}_{0.85}\text{N}$ electron blocking layer grown at 1,050 °C, and a 0.2- μm -thick Mg-doped p-type GaN cladding layer grown at 1,050 °C. In addition, the carrier concentrations of the n-GaN, p-type $\text{Al}_{0.15}\text{Ga}_{0.85}\text{N}$, and p-type GaN were 2×10^{18} , 3×10^{17} , and $5 \times 10^{18} \text{ cm}^{-3}$, respectively (Morkoc 1999; Huang et al. 2006). Figure 2 shows the standard structure of the InGaN/GaN-based LEDs and conventional LEDs device.

Device Fabrication

Conventional GaN/Sapphire LEDs

The GaN LEDs used in this dissertation are grown on *c*-face (0001) sapphire substrate by MOCVD. Detail of GaN structure has been described in previous section. Figure 3 shows the fabrication process of the conventional GaN/sapphire LEDs with planar electrodes (Chang et al. 2004). Prior to the chip process, the samples were cleaned in a standard solvent, dipped into HCl solution (HCl: H_2O = 1: 1) for 1 min, immersed into boiling aqua regia (HCl: HNO_3 = 3: 1) for 10 min, and then rinsed in running de-ionized water (Kim et al. 1998). Following, the Ni/Au (5 nm/5 nm) and indium tin oxide (ITO) as the p-type GaN ohmic contact layers were deposited and annealed in an oxygen environment at 500 °C for 10 min. Here, the Ni/Au and ITO layers are transparent conductive layer (TCL) for p-type GaN (Ho et al. 1999; Horng et al. 2001; Sheu et al. 2001). Then, a square mesa structure of $300 \times 300 \mu\text{m}^2$ to $1,000 \times 1,000 \mu\text{m}^2$ was fabricated by using a photolithograph and

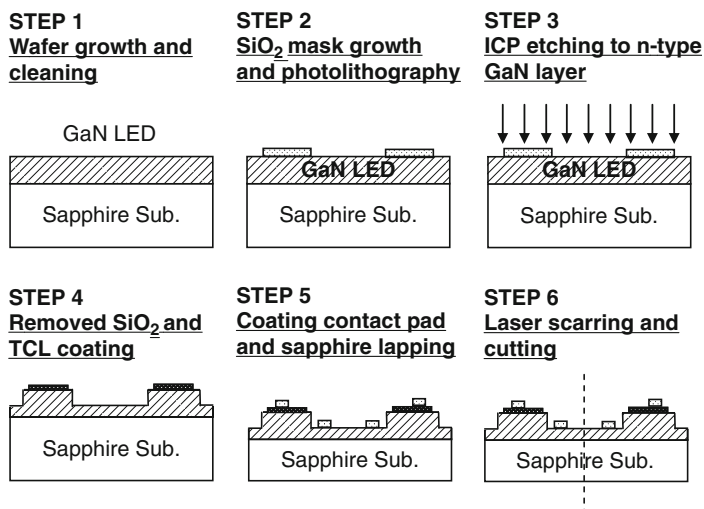


Fig. 3 Schematic drawing of the conventional GaN LED devices fabricated

inductively coupled-plasma (ICP) etcher etching to n-type GaN contact layer. Finally, the contact Ti/Al/Ti/Au metals were evaporated on the TCL and n-type GaN surface by thermal evaporating system.

n-Side-Up GaN/Mirror/Si LEDs by Metal Bonding

Figure 4 shows the flow chart of the fabrication processes for the n-side-up GaN/mirror/Si LEDs with vertical conducting electrodes (Huang et al. 2005, 2006; Lin et al. 2005). First, the sample was cleaned using standard solvent, dipped into HCl solution (HCl: H₂O = 1: 1) for 1 min, and immersed into boiling aqua regia (HCl: HNO₃ = 3: 1) for 10 min, then rinsed in running de-ionized water. After the cleaning process, a 1- μ m-thick SiO₂ mask was grown by plasma-enhanced chemical vapor deposition (PECVD) onto the p-GaN top layer and fabricated using photolithography to design the 1 mm width cross-pattern at the center of the LED wafer. Then, the cross-pattern area was etched down to sapphire by using an ICP system. Here, this cross-isolation pattern purposed to release the bubble during bonding process. Then, the various kinds of mirror were deposited on the p-type GaN surface. Here, the mirror materials not only can supply the high reflective mirror, but also must be ohmic contact with the p-type GaN layer. After deposition, the mirror materials were annealed at a suitable temperature to achieve ohmic contact characteristic under air or N₂ ambience. Next, E-beam evaporated Ti/Au (20/1,000 nm) metals were deposited onto LED wafer. Here, the Ti/Au metals play the role of the bonding process was bonding metal, but in electroplating process was seed-layer metal. In the bonding process, the Si wafer also loading into deposited the Ti/Au with LED wafer. After completion of the LED structure, the whole wafer (2 in. in diameter) was bonded to the Si substrate by thermal-pressure bonding and then subjected to the LLO process. The bonding temperature and pressure were 380 °C

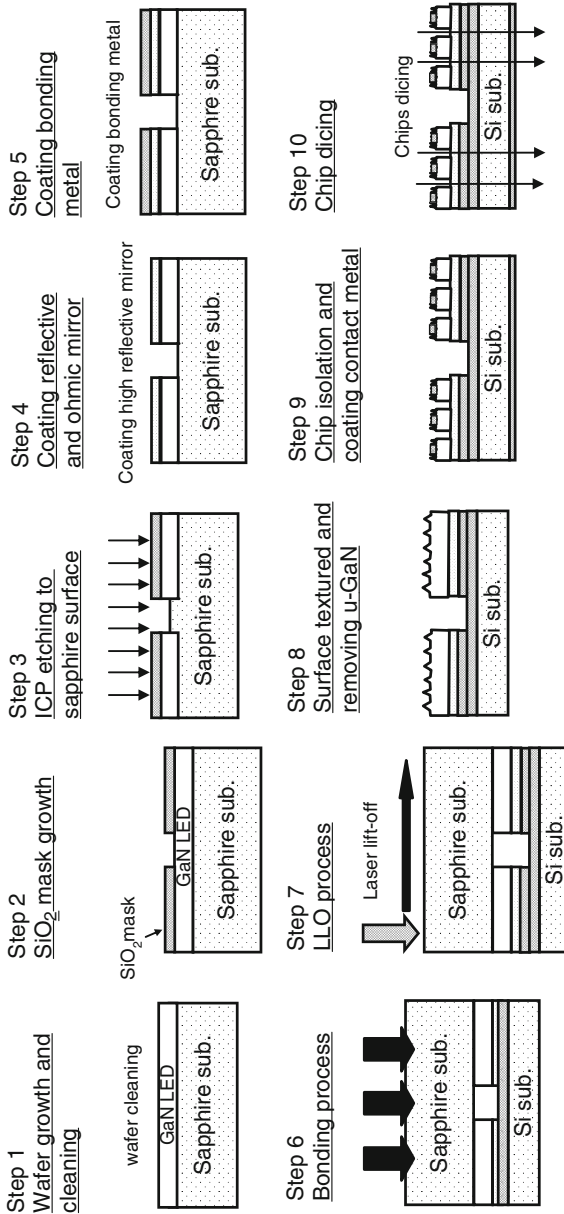


Fig. 4 Schematic drawing of the vertical structure n-side-up GaN/mirror/Si LED devices fabricated

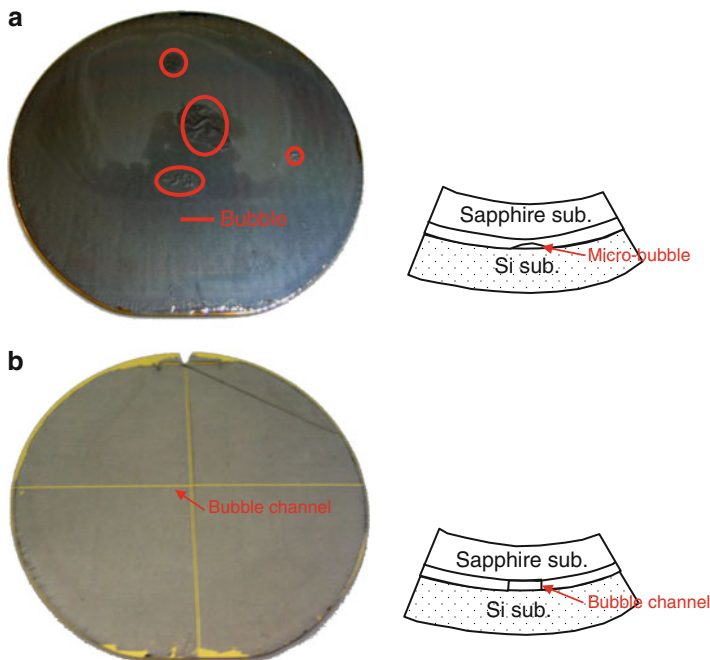
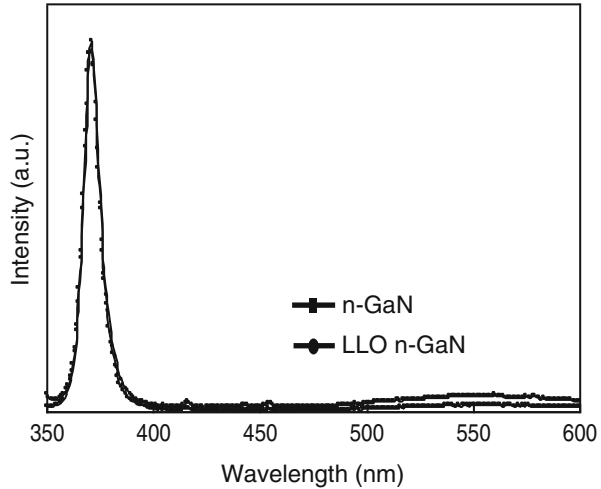


Fig. 5 Photographs of the 2-in. GaN epilayer with reflective mirror on Si substrates

for 1 h and 2.4 kg/cm^2 , respectively. The LLO process was performed using a UV laser. The laser light was irradiated from the back surface of the sapphire substrate, and GaN was locally heated close to the sapphire/GaN interface. After the entire LED wafer was scanned with the laser beam, the sapphire substrate was separated from the LED structure. Then, the GaN LED epilayers were transferred to the Si substrate with a reflective mirror. Here, the top layer of the LED/mirror/ n^+ -Si substrate structure is an undoped GaN epilayer. To enable n-contact formation, the undoped GaN was etched away to expose the n-GaN layer by wet chemical etching. The KOH solution was used to etch away the undoped GaN layer and roughen the n-GaN surface. After the texturing process, the n-GaN surface was cleaned with H_2SO_4 : H_2O_2 : H_2O (5: 1: 1) solution. The separation channel was defined by the thick photoresist via conventional photolithography, in which a square mesa structure with dimension of $1 \times 1 \text{ mm}^2$ was fabricated using ICP etching for electrical current isolation. Hot H_3PO_4 solution was employed in final mesa isolation process. Finally, the electron-beam evaporated Ti/Al layers on the n-GaN epilayer were used as the n-contacts. Finally, the high-power GaN/mirror/Si LEDs structure for vertical current injection was then completed. During the dicing process, the $1 \times 1 \text{ mm}^2$ square LED chip was packaged to measure the characteristic of the vertical-structure type LED device. To make a comparison, the conventional GaN/sapphire LED with the same chip size ($1 \times 1 \text{ mm}^2$) was fabricated in the use of ITO as the transparent p-electrode. Figure 5a,b show the pictures of 2 in. GaN epi-layer with reflective mirror on Si substrate. Figure 5a is whole

Fig. 6 PL spectrum of the n-GaN before and after the laser lift-off process



2 in. GaN wafer bonding with the Si substrate. Different from the Fig. 5a, b is 2 in. GaN wafer with cross-pattern at the center of the wafer bonding with Si substrate. When two pieces whole 2 in. wafer bonded together, some bubble will be locked inside which cannot release outside. Cross-channel in the center of the wafer can release the bubble effectively. Therefore, we can see the Fig. 5b with almost 100 % bonding yield. After LLO process, the normalized room-temperature band-to-band photoluminescence (PL) intensity of the GaN thin film before and after the LLO process was measured as shown in Fig. 6. The He-Cd laser ($\lambda = 325$ nm) was used in PL measurements. When applying the n-type GaN thin film to the PL, the emission wavelength observed is around 370–372 nm. As the laser signal is getting stronger, the full width at half maximum (FWHM) is narrower, indicating that both luminescence and quality of the thin film are better. From Fig. 6, we can see that the maximum PL intensity near band-edge luminescence of n-type GaN is around 372 nm, which is the same as that of GaN thin film after LLO process. Both of the FWHM were around 8 nm. This indicates that the luminescence of GaN thin film after LLO from sapphire substrate will not severely deteriorate the thin film quality. Therefore, in general, the LLO technique would not severely affect the GaN luminescence.

Surface Texturing Techniques

In conventional LEDs, the external efficiency is limited by the total internal reflection in the semiconductor to air interface because of the different refractive indices between the semiconductor and air. Thus, the internal light has difficulty in escaping into the air from the semiconductor. The total light reflective effect on the light extraction efficiency for the LEDs with large areas (1×1 mm²) is more obvious as compared to that of the LEDs with small areas (300×300 μ m²). Figure 7 shows the normalized extraction efficiency as a function of LED chip area at a current density

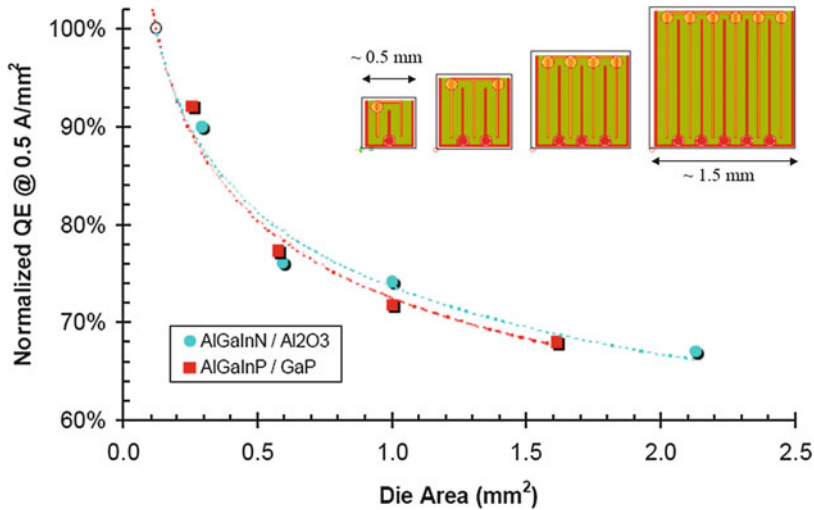


Fig. 7 Normalized extraction efficiency as a function of LEDs chip area at 0.5 A/mm² (Huh et al. 2003) (Quoted from Lumileds Lighting LLC Company)

of 0.5 A/mm² (Horng et al. 2006). When the chip area increases, the normalized extraction efficiency decreases. Because the LED structure can be regarded as a lateral waveguide, it results in increasing the probability of reabsorption. Since the emitted light can be scattered from the textured semiconductor surface, the textured surface can reduce the optical loss from the waveguide effect and enhances the light extraction efficiency (Windisch et al. 2000, 2001; Schnitzer et al. 1993; Huh et al. 2003; Pan et al. 2003; Fujii et al. 2004; Gao et al. 2004). In order to achieve a textured surface, several approaches have been attempted, such as using natural lithography method, epitaxial textured growth in an MOCVD system, or chemical wet etching (Horng et al. 2005; Wu et al. 2003).

Characterization of Vertical Mirror-Structure n-Side-Up Nitride-Based LEDs

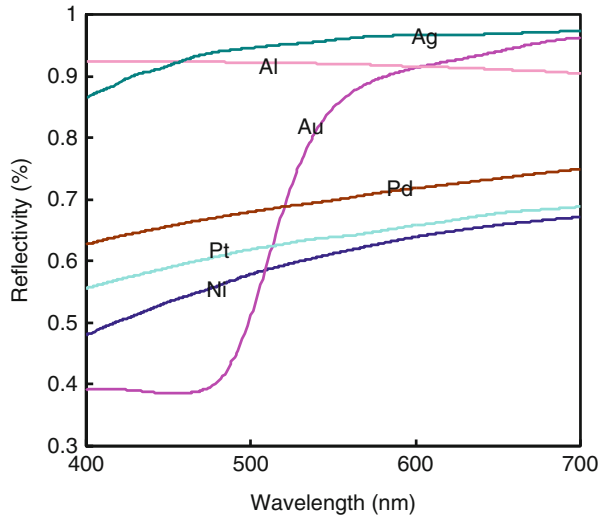
In this chapter, we attempt to combine the techniques of wafer bonding or electroplating, LLO and texturing an n-GaN surface by wet etching (without additional photoillumination) with various mirror materials to realize high-performance and high-brightness blue LEDs.

Design of Mirror Structure for Vertical n-Side-Up Nitride-Based LEDs

In the section [Introduction](#), we discussed the vertical n-side-up GaN LED has two important parts in which ohmic contact with p-type GaN layer and high reflective

Table 1 Characteristics of vertical GaN/mirror/LED with various mirror structure

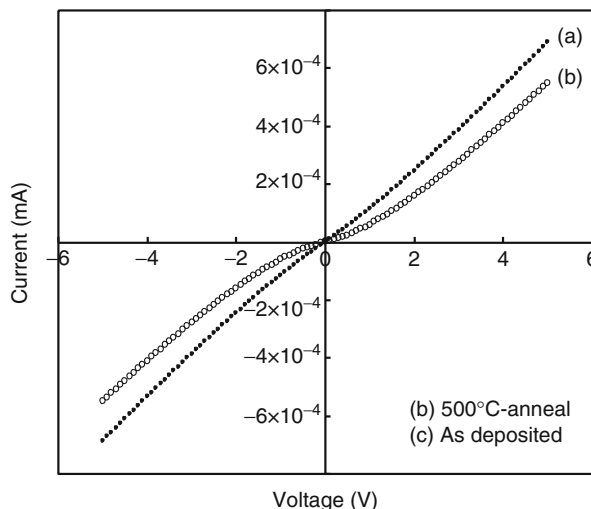
Vertical GaN LED on Si with various mirror	Reflectivity (%) at 470 nm	Output power (mW)	Forward voltage (V)	External quantum efficiency (%)
With Pd mirror	66	4.2	2.9	7.2
With NiO/Au/Ag mirror	61	3.9	2.9	6.7
With ITO/Al mirror	92	8.9	2.78	16.0
With NiO/Ag/Ni mirror	93	13	2.76	23.6

Fig. 8 Simulation results of the Au, Ni, Pd, Pt, Al, and Ag mirrors

mirror. In order to obtain the ohmic contact with p-type GaN layer, the metal must be with high work function characteristic. Table 1 shows the work function of commonly used metals. In Table 1, the Au, Ni, Pd, and Pt metals are usually used to contact with p-type GaN layer and also have higher work function than the other metals (Mori et al. 1996; Kim et al. 1998; Jang and Seong 2000). To know the reflective characteristics of each metal, we used the Essential Macleod program to simulate that. Figure 8 shows the simulation of reflectivity results of the Au, Ni, Pd, Pt high work function and Al, Ag high reflection metals. Although the high work function metals are easier with ohmic contact with p-type GaN layer than Al and Ag metals, these metals show poor optical reflectivity. Here, the reflectivity of Au, Ni, Pd, and Pt are 39, 54, 66, and 60 %, respectively. We chose the Pd metal in our study to ohmic contact with p-type GaN layer. Because this metal has higher reflectivity than the other high work function metals.

In this study, the Pd metal is in direct contact with the p-type GaN layer. For the vertical structure n-side-up GaN LEDs, the Pd layer acts not only as the mirror but also as the ohmic contact layer. Thus, it is important to evaluate the electrical properties of the Pd metal and p-type GaN layers. In this study, the Pd/p-type GaN

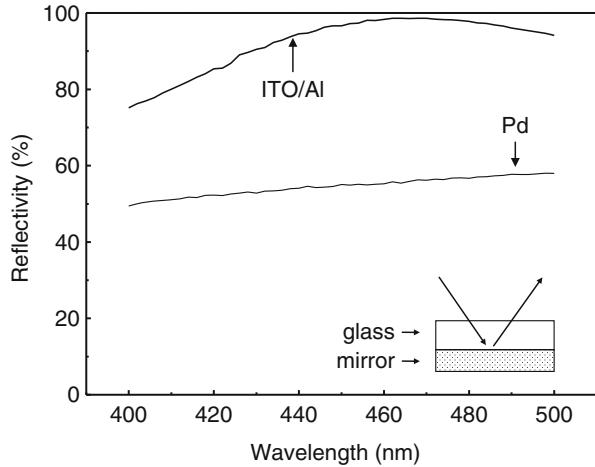
Fig. 9 I-V characteristics for Pd contact on p-type GaN samples before and after 500 °C thermal annealing



samples were annealed at 200 °C, 300 °C, 400 °C, and 500 °C. It was found that only the samples treated at 500 °C can exhibit ohmic contact characteristics. Figure 9 shows the typical contact characteristics of Pd metal and p-GaN layers for samples before and after annealed at 500 °C. Clearly, the as-deposited Pd/p-GaN sample can not exhibit the ohmic contact property. After the samples are annealed at 500 °C, the Pd metal layer exhibits linear contact characteristics with the p-GaN layer. The specific contact resistance between Pd and p-GaN was $8.1 \times 10^{-3} \Omega\text{-cm}^2$, as evaluated by a rectangular transmission-line method.

Although Pd metal can be ohmic contact with p-type GaN layer, the reflective characteristic is poor. Recently, the high transparency material was widely used as the p-GaN TCL such as oxidation of Ni/Au and Ni/ITO (Ho et al. 1999; Horng et al. 2001). The oxidation of Ni/Au TCL can obtain a specific contact resistance as low as $4.0 \times 10^{-6} \Omega\text{-cm}^2$ for p-GaN ohmic contact layer. The contact property of Ni/ITO showed low specific contact resistivity of $8.6 \times 10^{-4} \Omega\text{-cm}^2$ and high transparency (above 80 % for 450–550 nm) for p-GaN ohmic contact. On the other hand, the Si-doped InGaN/GaN short-period superlattice (SPS) tunneling contact layers are used instead of high-resistively p-GaN as a top contact layer (Sheu et al. 2001). It was found that ITO can be directly Ohmic contact on n⁺-SPS structure to obtain a reasonably small specific contact resistance of $1.6 \times 10^{-3} \Omega\text{-cm}^2$ and provide an extremely high transparency (above 93 % at 465 nm). In order to achieve the high reflective mirror in the n-side-up vertical structure GaN/mirror/Si LEDs, the high reflective metal (Al) and high transparency ohmic contact material (ITO) were combined to fabricate the high performance wafer-bonded n-side-up GaN/mirror/Si LEDs. For the LED with the highest reflectivity mirror among the ohmic contact metals, Pd is chosen to compare with the n-side-up GaN/mirror/Si with the ITO/Al reflectivity mirror. Figure 10 shows ITO/Al and Pd reflectivity as a function of wavelength. These mirror samples were deposited on the glass substrates

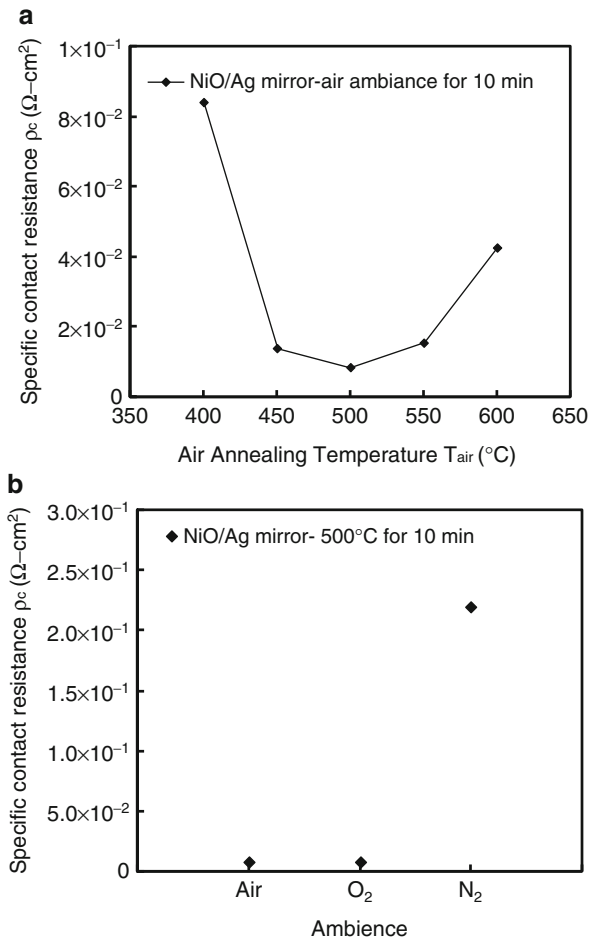
Fig. 10 Reflectivity as a function of wavelength for various metallic mirrors



for the reflectivity measurements where the incident light penetrated the glass and was reflected from the mirror materials. It was found that the reflectivity of the ITO/Al mirror showed 92 % at 470 nm and it was higher than that of Pd mirror (56 %).

Besides the TCL collocated with high reflectivity materials, it is well known that Ag is a very attractive material for mirror application due to its high reflectance (>95 %). However, Ag is not easily to make ohmic contact with p-GaN due to its lower work function. The instability of Ag after high temperature processing has also been observed even though it can achieve ohmic with the p-GaN layer (Jang and Lee 2004). Several approaches in reducing the contact resistance have been obtained using the Ag-based materials or combination such as Ni/Ag, Ni/Au/Ag, AgAl, Ag/Cu and Pd/Ag, etc. (Jang and Lee 2004; Hibbard et al. 2003; Kim et al. 2006, 2007; Adivarahan et al. 2001). Although the AgAl and AgCu complex materials show the better thermal stability, the reflectivity of the complex Ag-based materials is lower than that of the pure Ag mirror. In this study, a thermally stable mirror structure (NiO/Ag/Ni) is proposed for the vertical-conducting GaN/mirror/Si LEDs where no complex Ag-based materials are included. This multilayer mirror structure can achieve ohmic contact with p-GaN and serve high reflectivity characteristic. Details on the fabrication process and device performance of the vertical-conducting GaN/mirror/Si LEDs were described in section [Experiments](#). The specific contact resistance of Ni/Ag (2.5 nm/200 nm) on p-GaN as a function of air-annealing temperature (T_{air}) is shown in Fig. 11a. The specific contact resistance was measured to be $8.40 \times 10^{-2} \Omega\text{-cm}^2$ at $T_{\text{air}} = 400 \text{ }^\circ\text{C}$. When the T_{air} increased from 400 $^\circ\text{C}$ to 500 $^\circ\text{C}$, the specific contact resistance decreased rapidly. Since the Ni-to-NiO formation process is an essential criterion for p-GaN ohmic contact, lower T_{air} will make the Ni oxidation process incompletely. During the oxidation process, Ag will combine with the outdiffusing Ga atoms via the formation of Ag-Ga solid solution, producing a number of Ga vacancies in the p-GaN layer (Jang and Lee 2004). Therefore, the incomplete oxidation process would make the higher contact resistance with the p-GaN layer. The optimum contact resistance was measured to be

Fig. 11 Specific contact resistance (ρ_c) characteristics of the Ni/Ag metal contacts with p-GaN layer as a function of the (a) annealing temperature in air ambient and (b) annealing ambient



$8.37 \times 10^{-3} \Omega\text{-cm}^2$ at $T_{\text{air}} = 500 \text{ }^{\circ}\text{C}$. However, the contact resistance increased when the T_{air} increased above $500 \text{ }^{\circ}\text{C}$. This could be attributed to the Ag film being destroyed under high T_{air} conditions. Figure 11b shows the specific contact resistance of Ni/Ag (2.5 nm/200 nm) on p-GaN with various atmospheres at $500 \text{ }^{\circ}\text{C}$. It is worthy to mention that the annealing atmosphere (e.g. air, pure O₂ or N₂) has evident effect in the performance of Ni/Ag to p-GaN contact. The results show that the metal annealed in the air ambient presents the best ohmic contact properties.

The surface morphologies of the NiO/Ag and NiO/Ag/Ni contact layers on the glass substrates after oxidation annealing ($T_{\text{air}} = 500 \text{ }^{\circ}\text{C}$ for 10 min) were shown in Fig. 12a, b, respectively. The annealed NiO/Ag sample cannot maintain a film structure due to the Ag with lower melting point and inferior thermal stability. However, the annealed NiO/Ag/Ni sample still showed a smooth surface (Jang et al. 2007). To know the surface roughness, the NiO/Ag and NiO/Ag/Ni surface after oxidation annealing ($T_{\text{air}} = 500 \text{ }^{\circ}\text{C}$ for 10 min) were measured by atomic force

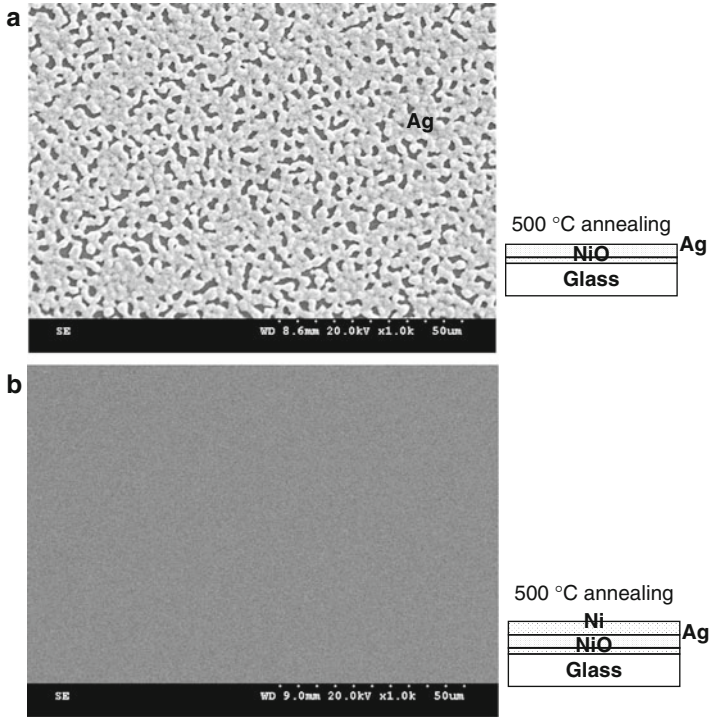


Fig. 12 Surface morphologies of the NiO/Ag and NiO/Ag/Ni contact metals coating on the glass after oxidation annealing at 500 °C for 10 min

microscope as shown in Fig. 13a, b. The corresponding root-mean-square roughness of the NiO/Ag mirror with and without a Ni cap layer was 1.5 and 112.2 nm, respectively. Obviously, the NiO/Ag mirror with a Ni top layer can protect the Ag mirror from oxidation and keeps the smooth surface morphology. The present result could be due to that fact that not only the bottom Ni layer was oxidized during the annealing process but also the top Ni layer. A similar result was also reported by Wang et al., where the Ag film sandwiched by the top and bottom oxidation layers could achieve a higher thermal stability (Wang and Alford 1999).

Figure 14 shows the cross-section TEM image and EDX analyses of the GaN epilayer and NiO/Ag/Ni mirror interface. In the EDX analysis of the EDX01 area, the first Ni is very thin (~2 nm) and meantime the O signal is found. The O signal confirmed that the thin Ni layer has become NiO during 500 °C annealing in the air ambience. On the other hand, the EDX02 area is Ag mirror layer. The EDX analysis of the EDX02 area, we did not find the O signal in this area. Therefore, we can be sure the top Ni layer with good protection for the Ag mirror does not oxidation during annealing process. Be noted to the signal of the EDX02 area has Ga signal in the Ag mirror layer. For this reason, we can confirm that the Ga has interdiffused to the Ag area during the annealing process to help for ohmic contact.

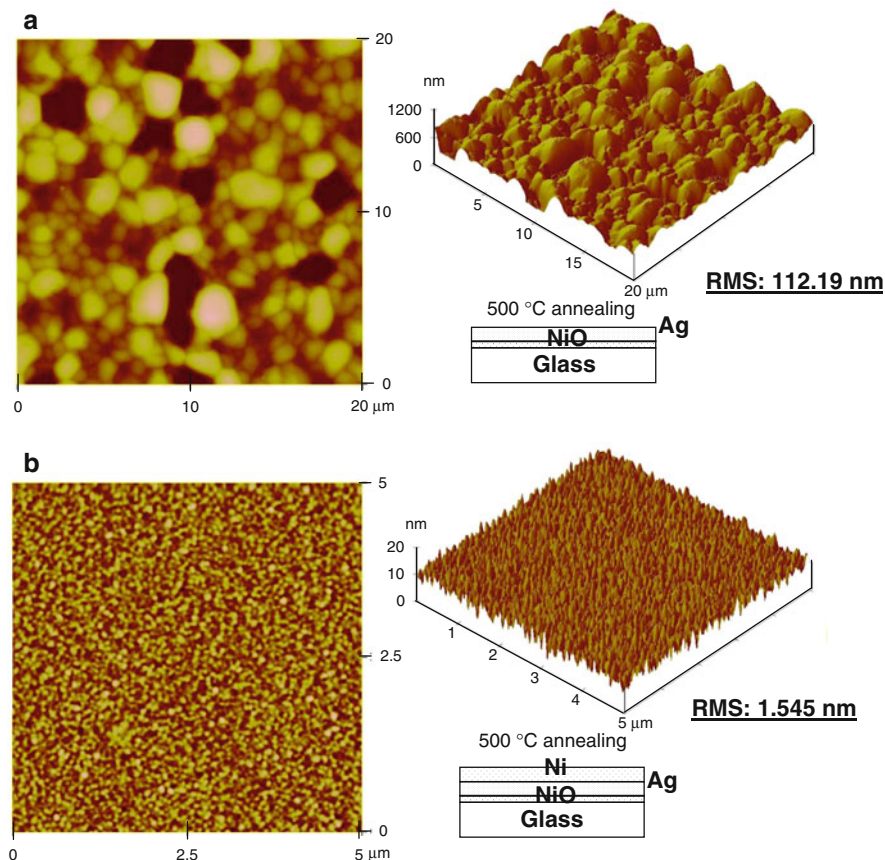


Fig. 13 AFM roughness of the NiO/Ag and NiO/Ag/Ni contact metals coating on the glass after oxidation annealing at 500 °C for 10 min

Figure 15 shows the reflectivity of NiO/Ag/Ni (2.5/200/100 nm), NiO/Ag (2.5/200 nm), and NiO/Au/Ag (5/5/200 nm) mirror structures as a functions of wavelength from 400 to 600 nm. These mirror samples were deposited on the glass substrates for the reflectivity measurements where the incident light penetrated the glass and was reflected from the mirror materials. It was found that the reflectivity of the NiO/Ag mirror showed only 8.6 % at 470 nm. This could be resulted from the Ag mirror being destroyed after the high temperature annealing and the Ag mirror surface became very rough as observed in Fig. 12a. However, the reflectivity of the NiO/Ag/Ni was measured to be 92 % at 470 nm, which agreed well with the smooth surface morphology as discussed in Fig. 12b.

To compare characteristics of the various reflective mirrors such as Pd, ITO/Al and NiO/Ag/Ni, the vertical structure n-side-up GaN/mirror/Si LEDs with different mirror was fabricated by wafer-bonding and LLO technologies. Fig. 16 shows the reflectivity of each mirror as a function of the wavelength from 400 to 600 nm.

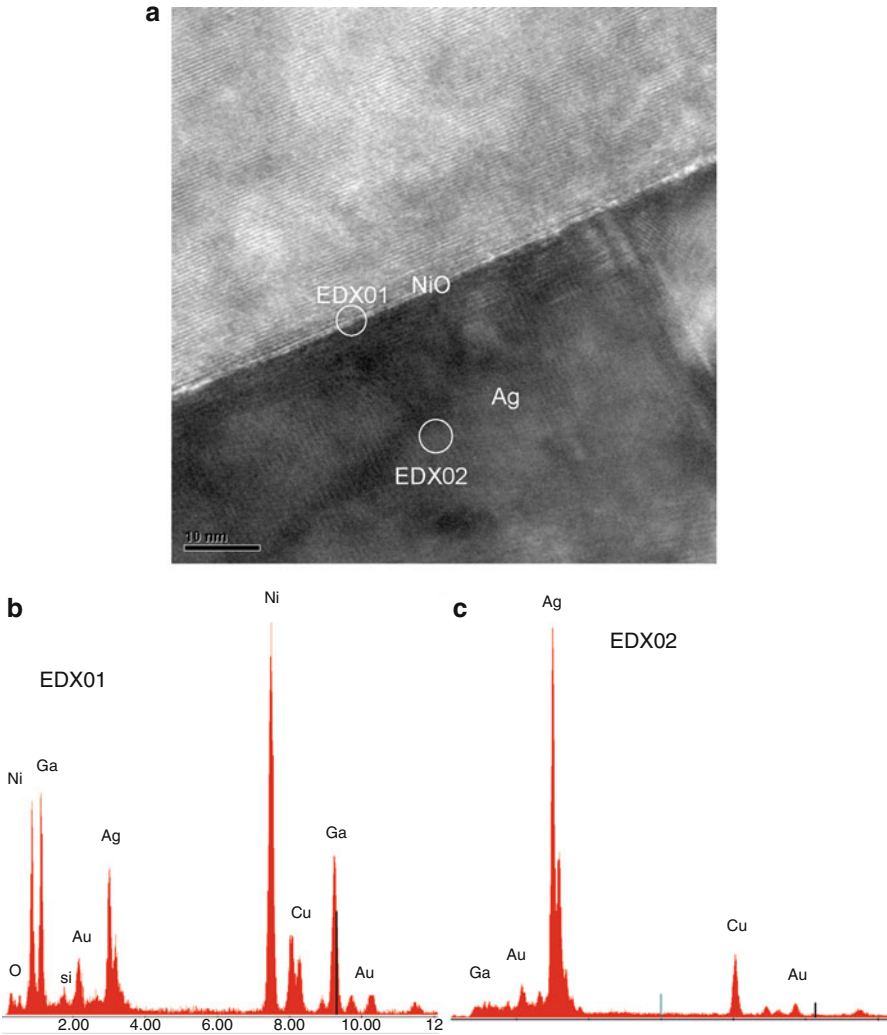


Fig. 14 (a) Cross-section TEM image and EDX analysis of (b) GaN and NiO/Ag/Ni mirror interface and (c) Ag mirror

The reflectivity was measured from glass to mirror materials, too. Note that the reflectivity of the NiO/Au/Ag sample was only 61 % at 470 nm. Obviously, the NiO/Ag/Ni mirror structure (oxidized from Ni/Ag/Ni) and ITO/Al show the best reflector performance for the vertical-conducting n-side-up GaN/mirror/Si LEDs.

Surface Texturing for Vertical n-Side-Up Nitride-Based LEDs

For surface-textured LEDs with a rear reflector, an external quantum efficiency of 40 % has already been reported (Windisch et al. 2000). For highly efficient

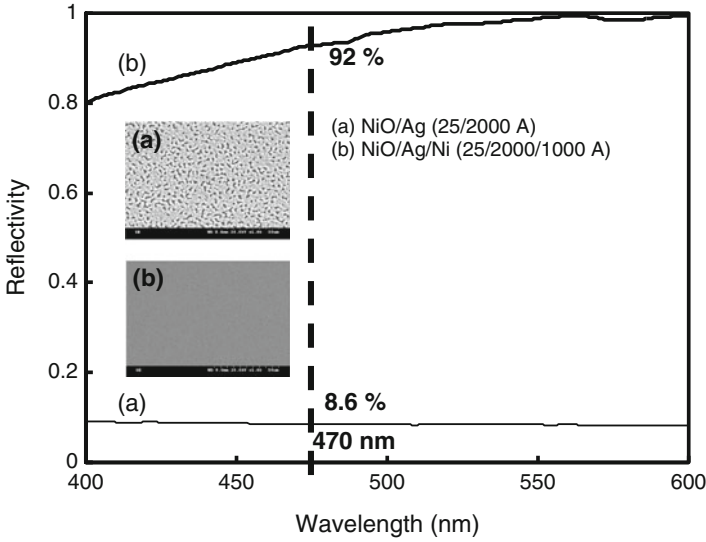
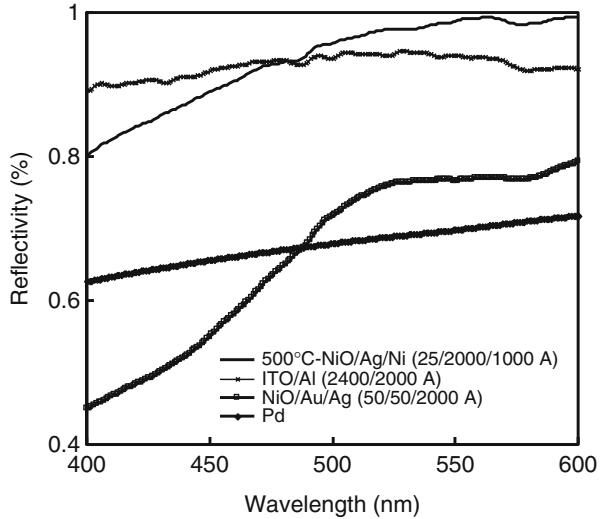


Fig. 15 Reflectivity of the NiO/Ag and NiO/Ag/Ni after annealing at 500 °C for 10 min as a function of the wavelength from 400 to 600 nm

Fig. 16 Reflective characteristics of ohmic contact to p-GaN mirror as a function of the wavelength from 400 to 600 nm



red/amber LEDs, the textured surfaces can be directly obtained by plasma etching the top of the LED structure epilayer. However, the thickness of the p-GaN cladding top layer is thin, about 0.3 μm . It is not desirable to directly etch the p-GaN cladding layer, because it will increase the resistance of the p-GaN layer after the surface texturing process, and such treatment might cause electrical deterioration. Use of the wet etching process can avoid the occurrence of plasma damage on the LED surface.

Photochemical etching for roughening an LED surface has been reported previously (Sharma et al. 2005; Haberer et al. 2004). The basic mechanism for the photoenhanced etching is the oxidative dissociation of the semiconductor into its component elements (thereby consuming the photogenerated holes) and the subsequent reduction of the oxidizing agent in the solution by reaction with the photogenerated electrons. This results in the etching being nonuniform. Moreover, there are some etching rate problems, such as that a multi-quantum well is easily etched and it degrades the device performance. If the etching is performed using a hot chemical solution, the nonuniform etching problem can be avoided and the etching rate can be easily controlled by adjusting the solution temperature.

After the vertical mirror structure, GaN structure was obtained and the top surface of the n-GaN layer was etched with a wet chemical solution in order to obtain the textured surface. Fig. 17a, b show the scanning electron microscopy images of n-GaN layers which were surface textured by etching with 37 % and 50 % KOH solutions, respectively, at 150 °C for 1 min, 2 min, and 3 min. For both samples, a surface covered with hexagonal conelike structures could be obtained. However, for the sample etched with 37 % KOH solution, the size of the hexagonal conelike structures is larger than that for the sample etched with 50 % KOH solution. It is well known that the roughness of the textured surface of LEDs must be near to or larger than the emission wavelength in order to enhance the external quantum efficiency. If the top surface of the n-GaN layer was etched with 37 % KOH solution, the encompassed size of based plane size of the hexagonal conelike structures of over 400 nm is easily and quickly obtained. This is attributed to the fact that at higher KOH concentrations, negative charge can build up on the n-GaN surface and hinder the diffusion of OH⁻ ions; this results in a decrease in the etching rate. This phenomenon was also observed in the previous study of Stocker et al. (1998). In our study, the 400–500 nm structure size could be easily obtained for the samples etched with 37 % KOH solution at 150 °C for 1 min. It is noteworthy that large hexagonal conelike structures can be obtained as the etched time increases, and this result in a thinning of the n-GaN layer. The optimum etching parameters are 1 min etching time and 37 % KOH solution.

It is important to evaluate the effect of surface texture on LED performance. Figure 18 shows the L-I characteristic of the standard LEDs and vertical GaN LEDs without a textured surface and with a textured surface (1 min etching in 37 % KOH solution). Here, the vertical GaN LED chips were not encapsulated for the electrical and optical measurements. The vertical GaN LEDs with surface texturing exhibited a maximum luminance intensity of 130 mcd (at 20 mA) with a forward voltage of 3.2 V. The luminance intensity is over two times larger than that of the original planar GaN/sapphire LED (at 20 mA). Clearly, the standard LEDs exhibit the smallest brightness and easily saturate when injected at high current. Correspondingly, the vertical GaN LEDs can be operated with high current injection without saturation. The Si substrate plays an important role in the thermal heat sink. The vertical GaN LEDs with textured surface exhibit the highest brightness. This means that light can escape from the n-GaN/air interface because of the change in light path caused by surface scattering due to the surface texture. Because of the use of the surface

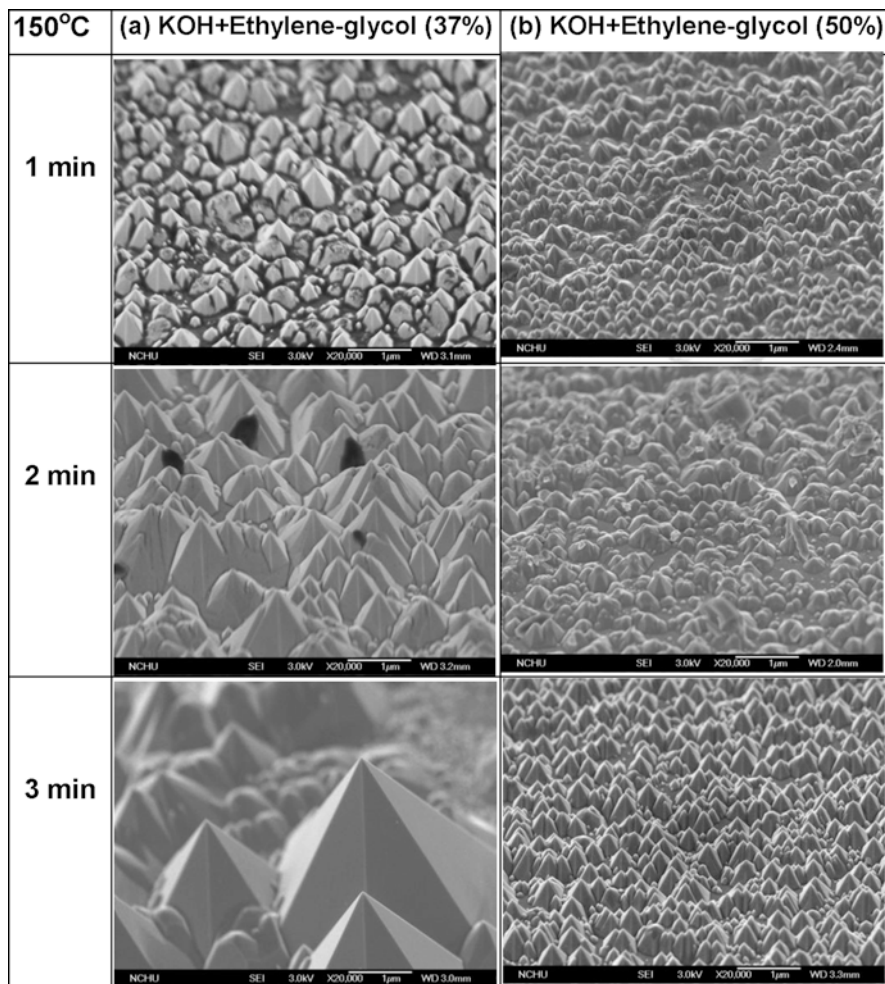


Fig. 17 SEM images of n-GaN etched with (a) 37 % and (b) 50 % KOH solutions at 150 °C for 1 min, 2 min, and 3 min

texturing technique, the light can escape easily from the air/semiconductor interface, which increases the light extraction efficiency.

Figure 19 shows the output power characteristics of the standard, vertical structure GaN LEDs without a textured surface and with a textured surface. Clearly, the vertical structure GaN LEDs with a textured surface exhibited a maximum output power of 4.3 mW (at 20 mA) that was higher than the output powers of 2.6 and 3.2 mW of standard and vertical structure GaN LEDs without a textured surface, respectively. For high current injection, the vertical structure GaN LEDs exhibit better performance than the standard LEDs. These results can be attributed to the vertical structure GaN and surface texturing processes, which result in the reflective

Fig. 18 L-I curves of VB-type GaN LEDs with and without surface texturing as compared with that of conventional GaN LED with sapphire substrate

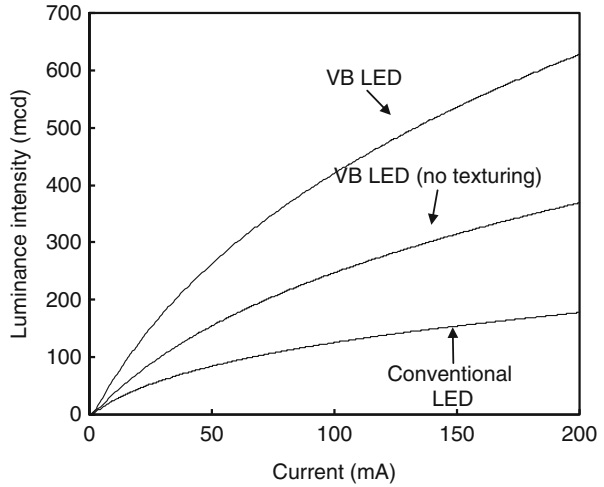
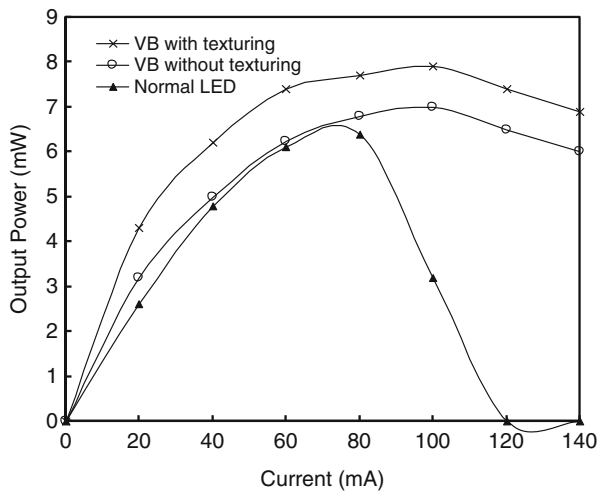


Fig. 19 Output power as function of injection current for VB-type GaN LEDs with and without surface texturing as compared with that for conventional GaN LED with sapphire substrate

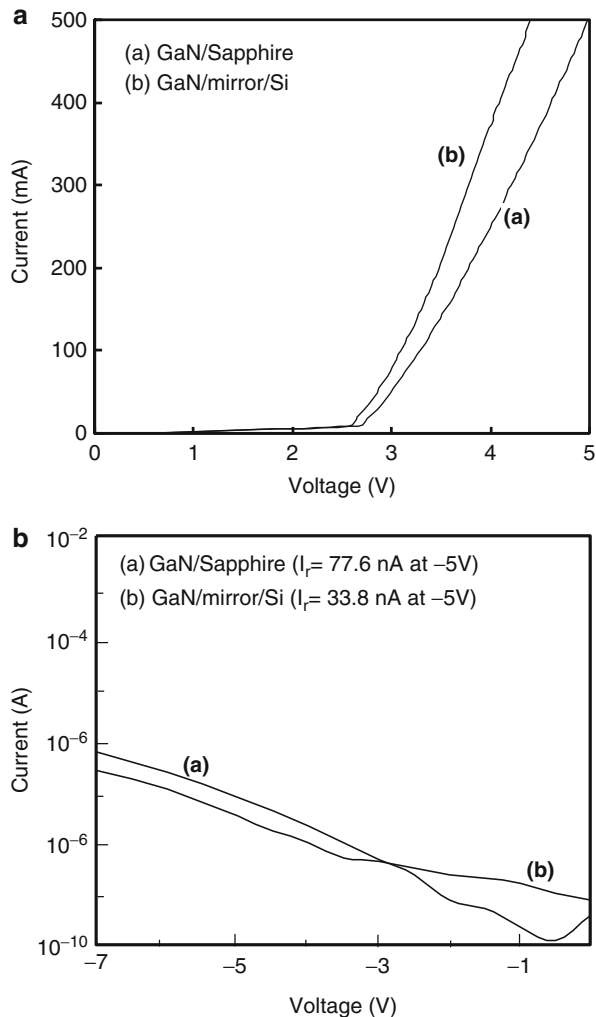


thermal dissipating substrate and increasing extraction efficiency. On the other hand, it was found that the encapsulated LEDs cannot be operated with high current injection; their L-I characteristics exhibit markedly early saturation (as shown in Fig. 19) as compared with LEDs without encapsulation (as shown in Fig. 18).

Characteristics of Vertical n-Side-Up Nitride-Based LEDs

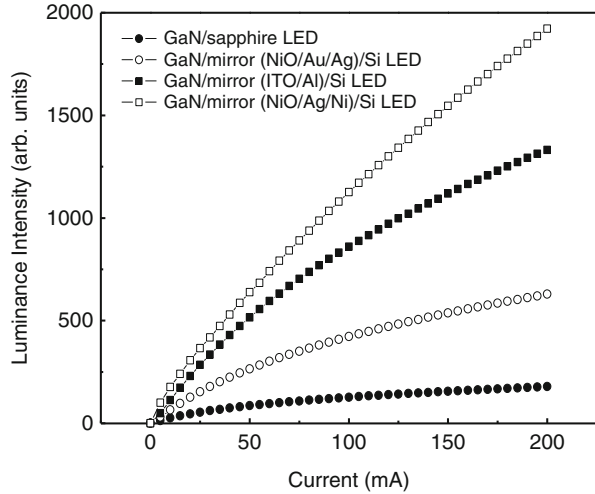
Figure 20a, b show the forward and reverse electrical characteristics of the vertical structure GaN/mirror/Si and conventional GaN/sapphire LED. Figure 20a shows the current-voltage characteristics of the vertical structure GaN/mirror (NiO/Ag/Ni)/Si

Fig. 20 Forward and reverse electrical characteristics of the vertical structure GaN/mirror/Si and conventional GaN/sapphire LED



and original planar GaN/sapphire LED samples. Here, the LED chips were not encapsulated for the electrical and optical measurements. The forward voltage (at 20 mA) of the original GaN/sapphire and vertical structure n-side-up GaN/mirror/Si LED samples were 2.76 and 2.64 V, respectively. Especially, the forward voltage of the GaN/mirror/Si LED sample was lower than the original GaN/sapphire sample. This suggests that the series connection resistance of the GaN/mirror/Si LEDs does not arise with the present vertical conducting structure. Furthermore, there was no abnormal phenomenon in the curve slope of the GaN/mirror/Si LEDs with increased injection current to 500 mA. Figure 20b shows

Fig. 21 L-I characteristics of the original sapphire LEDs and vertical structure n-side-up GaN/mirror/Si LEDs with NiO/Ag/Ni, ITO/Al, and NiO/Au/Ag reflectors



the reverse characteristic, the leakage current of the n-side-up GaN/mirror/Si LED at a reverse voltage of 5 V was 33.8 nA, and the conventional GaN/sapphire LED was 77.6 nA. The vertical LED has the normal reverse characteristic which compared with the conventional LED. That means the n-side-up GaN/mirror/Si did not destroy by vertical structure process.

Figure 21 presents the L-I characteristics of the original sapphire LEDs and vertical structure n-side-up GaN/mirror/Si LEDs with NiO/Ag/Ni, ITO/Al, and Pd reflectors. The LED chips were not encapsulated for the electrical and optical measurements. The luminance intensity of the original LEDs saturated early at about 50 mA of injection current. In contrast, the luminance intensity of vertical structure n-side-up GaN LEDs increased as the injection current increased. This occurs because the thermal conduction of the Si substrate is better than that of the sapphire substrate. Moreover, the GaN/NiO/Ag/Ni/Si LEDs with surface texturing presented maximum luminance intensity (at 20 mA) with a forward voltage of 2.7 V. The luminance intensity is over five times the magnitude of the original planar GaN/sapphire LEDs. Furthermore, the GaN/mirror (NiO/Ag/Ni)/Si LEDs present over two times the luminance intensity of the GaN/mirror (Pd)/Si LEDs.

Figure 22 shows output power vs. injection current for the GaN/mirror/Si with a NiO/Ag/Ni mirror and the original GaN/sapphire LEDs, where the chips were encapsulated into lamp form. The output power of the GaN/mirror (NiO/Ag/Ni)/Si LEDs were 13 mW at 20 mA; that were higher than the original LEDs. It is well know that the device performance in the lamp form condition degrades much easier due to excessive heating. With increasing current, the light output power for the GaN/mirror/Si LED was more stable than that of the original LEDs. This result suggests again that the GaN/mirror/Si LED has a higher current capability than the original GaN/sapphire LED due to the higher thermal conductivity of the Si substrate.

Fig. 22 Output power versus injection current for the GaN/mirror/Si with a NiO/Ag/Ni mirror and the original GaN/sapphire LEDs

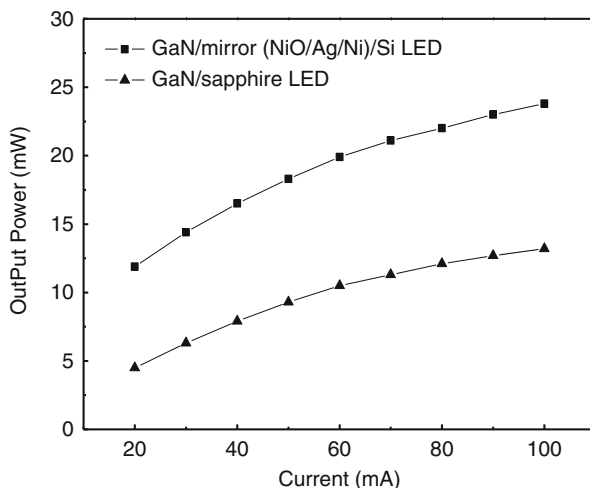


Table 1 summarizes the characteristics of various mirror material in vertical mirror-structure n-side-up GaN LEDs. Compared with the Pd, NiO/Au/Ag, ITO/Al, and NiO/Ag/Ni mirror structure in vertical GaN/mirror/Si, with NiO/Ag/Ni mirror GaN LED has the best performance in the optical characteristic. With NiO/Ag/Ni GaN mirror LED has 23.6 % external quantum efficiency that is higher than the other mirror structure.

As shown in Fig. 23, the junction heating effect on both LEDs can be further interpreted using the variation in emission peak wavelength measured under different injection currents. For the driving current less than 100 mA, a small blue shift (1 nm) for the GaN/sapphire LED was observed due to a band-filling effect of the localization energy states caused by indium composition fluctuation in the InGaN MQWs (Mukai et al. 1998). When the driving current increased from 300 to 800 mA, the peak wavelength of the GaN/sapphire LED showed a drastic red shift from 475 to 484 nm. However, the GaN/mirror/Si LEDs showed a continued blue shift, which is very similar to that observed in the commercial GaN/SiC LEDs. We have also measured the top surface temperature of both LED lamps (via epoxy) using a thermal couple. At 350 mA, the surface temperature of the GaN/sapphire LED showed 20 °C higher than that of the GaN/mirror/Si LED. These suggest that the excessive heating from the active layer during high current operation can be dissipated via the Si substrate. This indicates that large-area Si substrate LEDs have better thermal management and good heat sink. Furthermore, to obtain the real junction temperature at high current operation is necessary. A theoretical model of the dependence of the diode forward voltage has been developed by Xi and Schubert (2004). Excellent agreement between the theoretical and experiment temperature coefficients of the forward voltage ($dV_f = dT$) is found. In our study, the diode forward voltage is used to assess the junction temperature of GaN/mirror/Si and GaN/sapphire LEDs. The forward voltage method consists of two series of

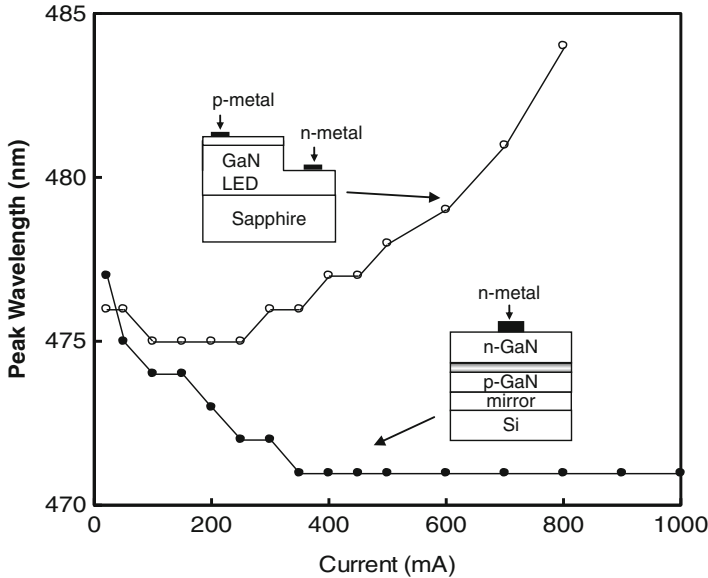


Fig. 23 Electroluminescence peak wavelength as a function of injection current for large-area vertical GaN/mirror/Si and GaN/sapphire LED samples

measurements: one is a calibration and the other is an actual junction temperature measurement. In the calibration, a pulsed forward current drives the LED sample located in a temperature controlled oven. Figure 24a, b show the relationship of measured forward voltage vs. junction temperature of GaN/mirror/Si and GaN/sapphire LEDs. The temperature coefficient of the forward voltage at low current is -2.40 and -2.35 mV/K for the GaN/mirror/Si and GaN/sapphire, respectively. Figure 25 shows the junction temperature vs. the forward current of GaN/mirror/Si and GaN/sapphire LED. As the forward current increases from 20 to 200 mA, the junction temperature of the GaN/mirror/Si increases from 23.6 °C to 86.8 °C which is smaller than GaN/sapphire increase from 37.8 °C to 157.8 °C. It is obvious that the GaN/ mirror/Si LED structure has better thermal dissipation than the GaN/sapphire LED.

Finally, it is important to study the effect on the lifetime of vertical structure GaN LEDs of being processed by wafer bonding, LLO, and surface texturing techniques. The long-term reliability characteristics of 12 mil vertical structure GaN LEDs stressed at 20 mA, at ambient temperature was measured. Figure 26 shows the luminance intensity and variation in forward voltage of the vertical structure GaN LEDs with surface texturing as functions of aging time. After 500 h, the luminance intensity and voltage variation were not over ± 20 %. The results of the lifetime test indicate that these processes do not deteriorate the roughened vertical structure GaN LED performance.

Fig. 24 Forward voltage versus oven temperature under different pulsed injection currents for (a) GaN/mirror/Si and (b) GaN/sapphire LEDs

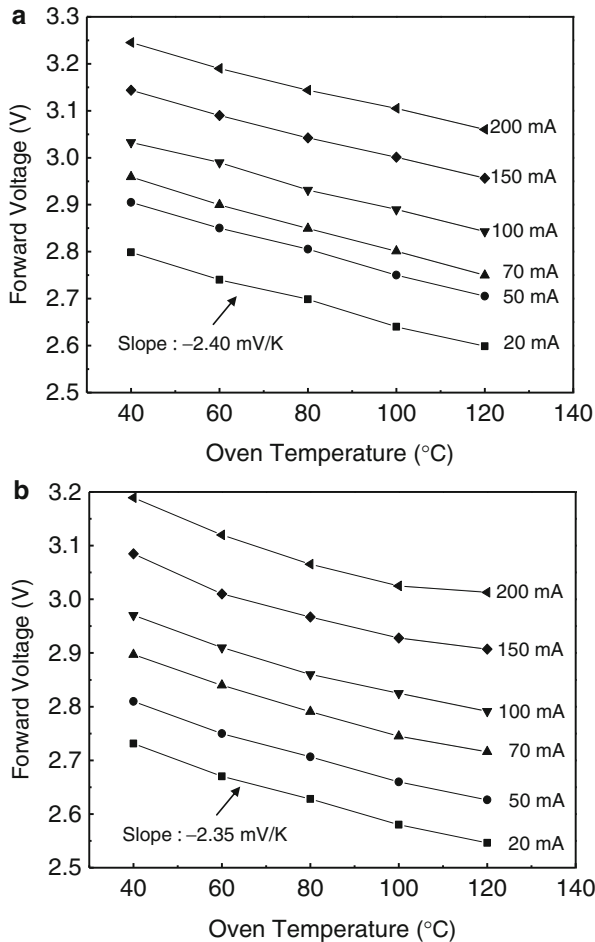


Fig. 25 Junction temperature as a function of forward current for the GaN/mirror/Si and GaN/sapphire LEDs

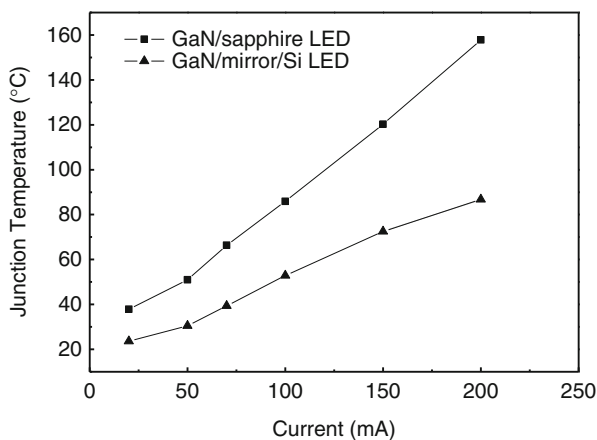
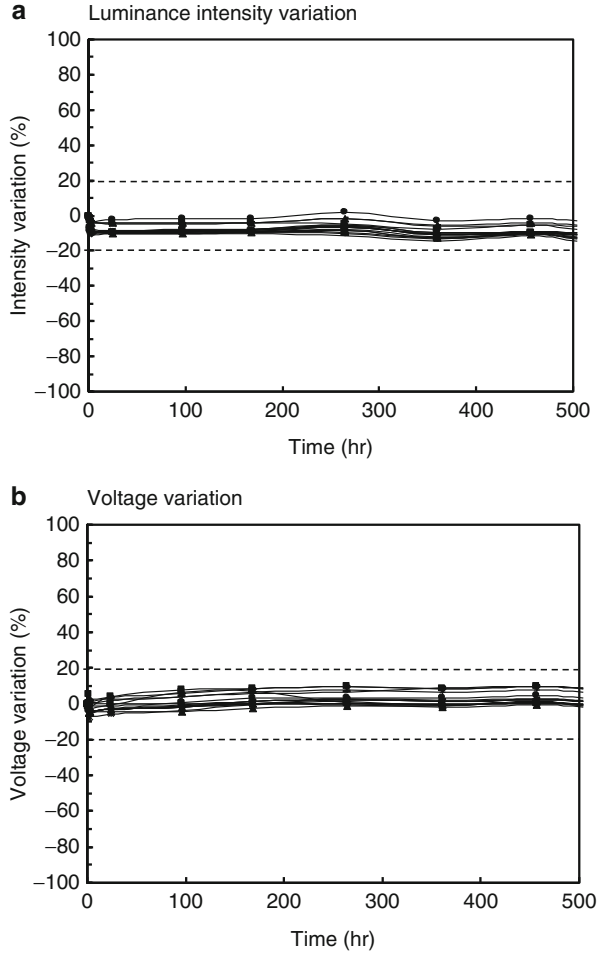


Fig. 26 (a) Luminance intensity and (b) variation in forward voltage as functions of stress time for vertical structure GaN/mirror/Si LEDs



Summary

In summary, n-side-up GaN/mirror/Si LED with vertical electrode was fabricated by wafer bonding and LLO processes. Using hot KOH solution, a textured n-GaN surface could be obtained. It was found that a sample with a textured surface exhibited better performance than a sample without a textured surface. In the mirror design, the vertical GaN LEDs with high reflective mirror (NiO/Ag/Ni) achieve over five times and two times light intensity than the original GaN/sapphire and Pd-mirror/Si LEDs, respectively. We used the theoretical model of the dependence of the diode forward voltage to calculate the junction temperature. The GaN/mirror/Si structure has lower junction temperature than original GaN/sapphire LED that mean with Si substrate has better heat transfer than sapphire substrate. Furthermore, the long-term reliability characteristics of the vertical GaN LEDs indicate that the

wafer bonding, LLO, and surface texturing techniques used to fabricate the vertical GaN LEDs with a roughened surface are viable for applications requiring high reliability.

Conclusions

The n-side-up vertical structure GaN/mirror/Si LEDs can achieve better performance during high power operation. However, it is not easy to obtain a high reflectivity mirror and ohmic contact with p-type GaN at the same time. Much of the literature indicates that Ag-based materials such as Ag, AgAl, AgCu, and Ni/Ag can achieve high reflectivity and ohmic contact. The AgAl and AgCu complex materials also have better thermal stability than the Ag metal alone. For this dissertation, we studied the ITO/Al structure for ohmic contact with p-type GaN and obtained >90 % reflectivity in the n-side-up structure with vertical-conducting electrodes. We also used wafer bonding and laser lift-off techniques to successfully transfer the 2-in. GaN epilayer on the Si substrate. To improve the bonding yield, a bubble-channel at the center of the 2-in. GaN wafer was utilized to help with bubble release. The achieved output power of the GaN/ITO/Al/Si LED is 8.9 mW, which is higher than that of conventional GaN/sapphire (4.5 mW) LED at 20 mA. Under a high current injection, the surface-textured GaN/ITO/Al/Si LED also shows more linear luminance intensity and output power. In the Ag-based mirror structure, we provided a new mirror structure (NiO/Ag/Ni) which demonstrated both ohmic contact with the p-type GaN layer and high thermal stability. To make a more quantitative understanding, the junction temperature was measured by the diode forward method. As a result, when the dc forward current increases to 200 mA, the junction temperatures of the GaN/mirror/Si and GaN/sapphire LED samples show 87 °C and 158 °C, respectively.

References

- Adivarahan V, Lunev A, Khan MA, Yang J, Simin G, Shur MS, Gaska R (2001) Very-low-specific-resistance Pd/Ag/Au/Ti/Au alloyed ohmic contact to p GaN for high-current devices. *Appl Phys Lett* 78:2781–2783
- Amano H, Sawaki N, Akasaki I, Toyoda Y (1986) Metalorganic vapor phase epitaxial growth of a high quality GaN film using an AlN buffer layer. *Appl Phys Lett* 48:353–355
- Amano H, Sawaki N, Akasaki I, Toyoda Y (1989) P-type conduction in Mg-doped GaN treated with low-energy electron beam irradiation (LEEBI). *Jpn J Appl Phys* 28:L2112–L2214
- Chang SJ, Chang CS, Su YK, Chuang RW, Lai WC, Kuo CH, Hsu YP, Lin YC, Shei SC, Lo HM, Ke JC, Sheu JK (2004) Nitride-based LEDs with an SPS tunneling contact layer and an ITO transparent contact. *IEEE Photon Technol Lett* 16:1002–1004
- Chu CF, Yu CC, Cheng HC, Lin CF, Wang SC (2003) Comparison of p-side down and p-side-up GaN light-emitting diodes fabricated by laser lift-off. *Jpn J Appl Phys* 42:L147–L150
- Craford MG (1997) High brightness light emitting diodes, vol 48. Academic, San Diego

- Fujii T, Gao Y, Sharma R, Hu EL, DenBaars SP, Nakamura S (2004) Increase in the extraction efficiency of GaN-based light-emitting diodes via surface roughening. *Appl Phys Lett* 84:855–857
- Gao Y, Fujii T, Sharma R, Fujito K, DenBaars SP, Nakamura S, Hu EL (2004) Roughening hexagonal surface morphology on laser lift-off (LLO) N-face GaN with simple photo-enhanced chemical wet etching. *Jpn J Appl Phys* 43:L637–L639
- Haberer ED, Sharma R, Stonas AR, Nakamura S, DenBaars SP, Hu EL (2004) Removal of thick (>100 nm) InGaN layers for optical devices using band gap selective photoelectrochemical etching. *Appl Phys Lett* 85:762–764
- Hibbard DL, Jung SP, Wang C, Ullery D, Zhao YS, Lee HP, So W, Liu H (2003) Low resistance high reflectance contacts to p-GaN using oxidized Ni/Au and Al or Ag. *Appl Phys Lett* 83:311–313
- Ho JK, Jong CS, Chiu CC, Huang CN, Shih KK, Chen LC, Chen FR, Kai JJ (1999) Low-resistance ohmic contacts to p-type GaN. *Appl Phys Lett* 74:1275–1277
- Hornig RH, Wu DS, Lien YC, Lan WH (2001) Low-resistance and high-transparency Ni/indium tin oxide ohmic contacts to p-type GaN. *Appl Phys Lett* 79:2925–2927
- Hornig RH, Yang CC, Wu JY, Huang SH, Lee CE, Wu DS (2005) GaN-based light-emitting diodes with indium tin oxide texturing window layers using natural lithography. *Appl Phys Lett* 86:221101
- Hornig RH, Huang SH, Yang CC, Wu DS (2006) Efficiency improvement of GaN-based LEDs with ITO texturing window layers using natural lithography. *IEEE J Sel Top Quantum Electron* 12:1196–1201
- Huang SH, Hornig RH, Hsu SC, Chen TY, Wu DS (2005) Surface texturing for wafer-bonded vertical-type GaN/Mirror/Si light-emitting diodes. *Jpn J Appl Phys* 44:3028–3031
- Huang SH, Hornig RH, Wu DS (2006) Improvements of n-side-up GaN light-emitting diodes performance by indium–tin–oxide/Al mirror. *Jpn J Appl Phys* 45:3449–3452
- Huh C, Lee KS, Kang EJ, Park SJ (2003) Improved light-output and electrical performance of InGaN-based light-emitting diode by microroughening of the p-GaN surface. *J Appl Phys* 93:9383–9385
- Jang HW, Lee JL (2004) Mechanism for ohmic contact formation of Ni/Ag contacts on p-type GaN. *Appl Phys Lett* 85:5920–5922
- Jang JS, Seong TY (2000) Electronic transport mechanisms of nonalloyed Pt Ohmic contacts to p-GaN. *Appl Phys Lett* 76:2743–2745
- Jang HW, Son JH, Lee JL (2007) Highly reflective low resistance Ag-based Ohmic contacts on p-type GaN using Mg overlayer. *Appl Phys Lett* 90:012106
- Kelly MK, Ambacher O, Dahlheimer B, Groos G, Dimitrov R, Angerer H, Stutzmann M (1996) Optical patterning of GaN films. *Appl Phys Lett* 69:1749–1751
- Kelly MK, Ambacher O, Dimitrov R, Handschuh R, Stutzmann M (1997) Optical process for liftoff of group III-nitride film. *Phys Status Solidi A* 159:R3
- Kelly MK, Ambacher O, Dimitrov R, Angerer H, Handschuh R, Stutzmann M (1998) Laser-processing for patterned and free-standing nitride films. *Mater Res Soc Symp Proc* 482:973
- Kelly MK, Vaudo RP, Phanse VM, Gorgens L, Ambacher O, Stutzmann M (1999) Large free-standing GaN substrates by hydride vapor phase epitaxy. *Jpn J Appl Phys* 38:L217–L219
- Kim JK, Lee JL, Lee JW, Shin HE, Park YJ, Kim T (1998) Low resistance Pd/Au ohmic contacts to p-type GaN using surface treatment. *Appl Phys Lett* 73:2953–2955
- Kim JY, Na SI, Ha GY, Kwon MK, Park IK, Lim JH, Park SJ (2006) Thermally stable and highly reflective AgAl alloy for enhancing light extraction efficiency in GaN light-emitting diodes. *Appl Phys Lett* 88:043507
- Kim H, Baik KH, Cho J, Lee JW, Yoon S, Kim H, Lee SN, Sone C, Park Y, Seong TY (2007) High-reflectance and thermally stable AgCu alloy p-Type reflectors for GaN-based light-emitting diodes. *IEEE Photon Technol Lett* 19:336–338
- Lin WY, Wu DS, Pan KF, Huang SH, Lee CE, Wang WK, Hsu SC, Su YY, Huang SY, Hornig RH (2005) High-power GaN-mirror-Cu light-emitting diodes for vertical current injection using laser lift-off and electroplating techniques. *IEEE Photon Technol Lett* 17:1809–1811

- Mori T, Kozawa T, Ohwaki T, Yaga Y, Nagai S, Yamasaki S, Asami S, Shibata N, Koike M (1996) Schottky barriers and contact resistances on p-type GaN. *Appl Phys Lett* 69:3537–3539
- Morita D, Sano M, Yamamoto M, Murayama T, Nagahama S, Mukai T (2002) High output power 365 nm ultraviolet light emitting diode of GaN-free structure. *Jpn J Appl Phys* 41:L1434–L1436
- Morkoc H (1999) Nitride semiconductors and devices. Springer, Berlin
- Mukai T, Yamada M, Nakamura S (1998) Current and temperature dependence of electroluminescence of InGaN-based UV/blue/green light emitting diodes. *Jpn J Appl Phys* 37:L1358–L1361
- Nakamura S, Senoh M, Mukai T (1991) Highly p-typed Mg-doped GaN films grown with GaN buffer layers. *Jpn J Appl Phys* 30:L1708–L1711
- Nakamura S, Mukai T, Senoh M, Jwasa N (1992) Hole compensation mechanism of p-type GaN films. *Jpn J Appl Phys* 31:1258–1266
- Nakamura S, Senoh M, Nagahama S-I, Iwasa N, Matsushita T, Mukai T (2000) Blue InGaN-based laser diodes with an emission wavelength of 450 nm. *Appl Phys Lett* 76:22–24
- Pan SM, Tu RC, Fan YM, Yeh RC, Hsu JT (2003) Improvement of InGaN-GaN light-emitting diodes with surface-textured indium-tin-oxide transparent ohmic contacts. *IEEE Photon Technol Lett* 15:649–651
- Pankove JI, Miller EA, Berkeyheiser JE (1972) GaN blue light-emitting diodes. *J Lumin* 5:84
- Schnitzer I, Yablonovitch E, Caneau C, Gimttter TJ (1993) 30 % external quantum efficiency from surface textured, thin-film light-emitting diodes. *Appl Phys Lett* 63:2174–2176
- Sharma R, Haberer ED, Meier C, Hu EL, Nakamura S (2005) Vertically oriented GaN-based air-gap distributed Bragg reflector structure fabricated using band-gap-selective photoelectrochemical etching. *Appl Phys Lett* 87:051107
- Sheu JK, Tsai JM, Shei SC, Lai WC, Wen TC, Kou CH, Su YK, Chang SJ, Chi GC (2001) Low-operation voltage of InGaN-GaN light-emitting diodes with Si-doped $\text{In}_{0.3}\text{Ga}_{0.7}\text{N}/\text{GaN}$ short-period superlattice tunneling contact layer. *IEEE Electron Device Lett* 22:460–462
- Song YK, Diagne M, Zhou H, Nurmikko AV, Carter-Coman C, Kern RS, Kish FA, Krames MR (1999a) A vertical injection blue light emitting diode in substrate separated InGaN heterostructures. *Appl Phys Lett* 74:3720–3722
- Song YK, Zhou H, Diagne M, Ozden I, Vertikov A, Nurmikko AV, Carter-Coman C, Kern RS, Kish FA, Krames MR (1999b) A vertical cavity light emitting InGaN quantum well heterostr. *Appl Phys Lett* 74:3441–3443
- Stach EA, Kelsch M, Nelson EC, Wong WS, Sands T, and Cheung NW, Structural and Chemical Characterization of Free-standing GaN Films Separated from Sapphire Substrates by Laser Lift-off, *Appl. Phys. Lett.* 77 (2000) pp. 1819–1821
- Stocker DA, Schubert EF, Redwing JM (1998) Crystallographic wet chemical etching of GaN. *Appl Phys Lett* 73:2654–2656
- Strite S, Morkoc H (1992) GaN, AlN, and InN: a review. *J Vac Sci Technol B* 10:1237
- Tavernier PR, Hansen MC, DenBaars SP, Clarke DR (1999a) GaN LEDs transferred to copper substrates using laser assisted debonding. *J Electron Mater* 28:1003
- Tavernier PR, Verghese PM, Clarke DR (1999b) Photoluminescence from laser assisted debonded epitaxial GaN and ZnO films. *Appl Phys Lett* 74:2678–2680
- Tsakalagos L, Sands T (2000) Epitaxial ferroelectric (PbLa)(Zr, Ti)O₃ thin films on stainless steel by excimer laser liftoff. *Appl Phys Lett* 76:227–229
- Wang Y, Alford TL (1999) Formation of aluminum oxynitride diffusion barriers for Ag metallization. *Appl Phys Lett* 74:52–54
- Windisch R, Dutta B, Kuijk M, Knobloch A, Meinlschmidt S, Schoberth S, Kiesel P, Borghs G, Dohler GH, Heremans P (2000) 40 % efficient thin-film surface-textured light-emitting diodes by optimization of natural lithography. *IEEE Trans Electron Devices* 47:1492–1498
- Windisch R, Rooman C, Meinlschmidt S, Kiesel P, Zipperer D, Dohler GH, Dutta B, Kuijk M, Borghs G, Heremans P (2001) Impact of texture-enhanced transmission on high-efficiency surface-textured light-emitting diodes. *Appl Phys Lett* 79:2315–2317
- Wong WS, Sands T, Cheung NW (1998) Damage-free separation of GaN thin films from sapphire substrates. *Appl Phys Lett* 72:599–601

- Wong WS, Cho Y, Weber ER, Sands T, Yu KM, Krüger J, Wengrow AB, Cheung NW (1999) Structural and optical quality of GaN/metal/Si heterostructures fabricated by excimer laser lift-off. *Appl Phys Lett* 75:1887–1889
- Wong WS, Kneissl M, Mei P, Treat DW, Teepe M, Johnson NM (2001) Continuous-wave InGaN multiple-quantum-well laser diodes on copper substrates. *Appl Phys Lett* 78:1198–1200
- Wu LW, Chang SJ, Su YK, Chuang RW, Hsu YP, Kuo CH, Lai WC, Wen TC, Tsai JM, Sheu JK (2003) InGaN/GaN MQW LEDs with a low temperature GaN cap layer. *Solid State Electron* 47:2027–2030
- Xi Y, Schubert EF (2004) Junction-temperature measurement in GaN ultraviolet light-emitting diodes using diode forward voltage method. *Appl Phys Lett* 85:2163–2165

Phosphors for White LEDs

Chun Che Lin, Wei-Ting Chen, and Ru Shi Liu

Contents

Definition of Phosphor	182
Terminology	182
Luminescence Phenomena	183
Host Lattice	185
Configurational Coordinate Diagram	186
Selection Rule	188
Activator	189
Fundamentals of Phosphor	197
Requirements for Phosphor-Converted LEDs (pc-LEDs)	197
Classification of Phosphors for pc-WLEDs	205
Applications of Phosphors	207
Prospects of Phosphors	209
Red-Emitting Phosphor Materials	209
Packaging Technology for WLEDs	213
Conclusion	217
References	218

C.C. Lin

Department of Chemistry, National Taiwan University, Taipei, Taiwan

Institute of Organic and Polymeric Materials, National Taipei University of Technology, Taipei, Taiwan

e-mail: cclin0530@gmail.com

W.-T. Chen

Department of Chemistry, National Taiwan University, Taipei, Taiwan

e-mail: d98223129@ntu.edu.tw

R.S. Liu (✉)

Department of Chemistry, National Taiwan University, Taipei, Taiwan

Department of Mechanical Engineering and Graduate Institute of Manufacturing Technology, National Taipei University of Technology, Taipei, Taiwan

e-mail: rsliu@ntu.edu.tw

Abstract

White Light-Emitting Diodes (WLEDs) is a promising conserve energy device for altering the traditional illuminating apparatus because of their high efficiency, high flexibility, long lifetime, low energy consumption, and friendly environment. Of course you can frequently find WLEDs in your daily life. Phosphor is an important component of WLEDs and has been investigated broadly. This chapter introduces readers who begin meeting these fields to understand phosphor including history, principle, application, and perspective. The first part is a fundamental definition to luminescent materials. The second part provides requirements, classifications, and applications of phosphors for phosphor-converted LEDs (pc-LEDs). Finally, we propose some prospects and challenges of optical materials in the future.

Definition of Phosphor

The word “phosphor” was coined in the early seventeenth century by an Italian alchemist, Vincentinus Casciarolo, who found a heavy crystalline stone near a volcano and attempted to refine it as a noble metal. Casciarolo observed red light emitting from the sintered stone after it was exposed to sunlight (Shionoya and Yen 1998). The same discovery was reported in Europe, where these light-emitting stones were called phosphors. The word “phosphor” means “light bearer” in Greek. The definitions of relative terms depend on the user during different periods. In this section, we introduce the terminology, different luminescence phenomena, fundamental concepts (e.g., host lattice, configurational coordinate diagram, and selection rule), and the features of various activators.

Terminology

- (A) The word “fluorescence” denotes the direct transition from an excited state to a ground state, which is typically a short phenomenon. At present, fluorescence is defined as the imperceptible short afterglow of calcium fluoride (CaF_2) after excitation (Shionoya and Yen 1998).
- (B) The word “phosphorescence” was derived from the word “phosphor” which is the transition from an excited state to a ground state that passes through an intermediate state, in which electron is trapped. Consequently, a phenomenon longer than fluorescence is induced. At present, phosphorescence is defined as the afterglow that can be detected by the human eye after the excitation of light has stopped (Shionoya and Yen 1998). In several extreme cases, fluorescence can occur longer than phosphorescence.
- (C) The word “luminescence” was derived from the Latin word “lumen” which means “light” (Shionoya and Yen 1998). Luminescence extensively means phosphorescence and fluorescence. At present, luminescence is defined as a

phenomenon where electrons are excited by external energy and are released back to their original state in light form.

- (D) The word “light” refers to the electromagnetic waves from near-ultraviolet (UV) to near-infrared (IR) regions, including the visible region at wavelengths ranging from 400 to 700 nm (Shionoya and Yen 1998).

The word “phosphor” is not clearly defined and is dependent on the user. In this study, however, phosphor refers to the inorganic phosphors in polycrystalline powder form. The aforementioned definitions of phosphorescence and fluorescence are applied in inorganic-related fields. For organic molecules, the emission from a triplet excited state is defined as phosphorescence, whereas that from a singlet excited state is called fluorescence.

Luminescence Phenomena

- (A) *Bioluminescence*: Fireflies are a family of insects belonging to the beetle order Coleoptera, which are also called winged beetles and lightning bugs. Fireflies attract mates or prey using their conspicuous bioluminescence, which is a chemical reaction in their lower abdomen. The lighting process occurs when the enzyme luciferase acts on the compound luciferin in the presence of magnesium ions, adenosine triphosphate, and oxygen, which produces light. Bioluminescence is classified as “cold light” without IR or UV frequencies, and its region may be yellow, green, or pale red, with wavelengths ranging from 510 to 670 nm (HowStuffWorks 2001; BBC News; Stanger-Hall et al. 2007).
- (B) *Chemical luminescence*: Twenty-five years ago, Michael Rauhut and Laszlo Bollyky discovered bis(2,4,5-trichloro-6-carbopentoxypheyl) oxalate (Rauhut 1969), which is a portable, cheap, and safe substance used in the rave scene (e.g., for light necklaces, light glasses, and light ropes). The oxidation products of bis(2,4,5-trichloro-6-carbopentoxypheyl) oxalate are responsible for the chemiluminescence in light sticks, as shown in Fig. 1 (Wilson 1999; Chandross 1963).
- (C) *Triboluminescence*: The transient light of sugar is generated through the breaking of chemical bonds when a solid matrix is rubbed, crushed, ripped, scratched, or pulled apart. The electric discharge ionizes the surrounding air, which causes a flash of light during charge recombination. To date, this effect has not yet been clearly explained, but it seems to be caused by the separation and reunification of electric charges based on experimental, crystallographic, and spectroscopic demonstrations. Several researchers reported that asymmetric materials are triboluminescent because of charge separation and poor conduction. However, the symmetric hexakis(antipyrine)terbium iodide also emits light, which breaks the aforementioned rule. This particular condition can be explained by the impurities contained in this compound (Clegg et al. 2002).
- (D) *Thermally-induced luminescence*: The mineral form of CaF_2 is called fluorite, which is classified as a halide mineral. Many types of crystal structure consist of isometric cubic habit, octahedral, and complex isometric forms. Fluorite is a

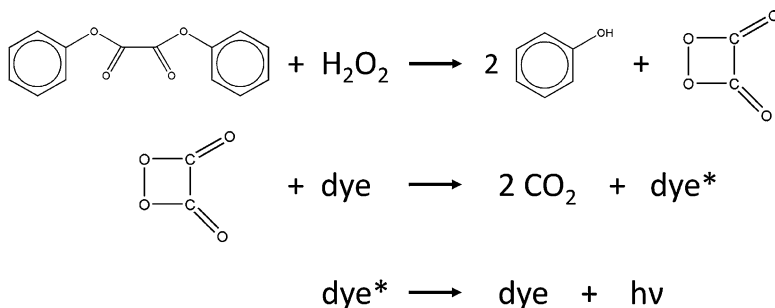


Fig. 1 Oxidation of diphenyl oxalate (*top*), decomposition of 1,2-dioxetanedione (*middle*), and relaxation of dye (*bottom*) (Clark 2000)

colorful material under UV to visible light. The fluorescence of fluorite arises from its impurities, such as organic matter, yttrium, and ytterbium. Boudeile et al. (Boudeile et al. 2008) reported the complete spectral properties of $\text{MeF}_2:\text{Yb}^{3+}$ ($\text{Me} = \text{Ca}$ and Sr) under high-power pumping with or without the laser effect. $\text{CaF}_2:\text{Yb}^{3+}$ exhibits better performance than $\text{SrF}_2:\text{Yb}^{3+}$ at an emission wavelength of 980 nm under high-power continuous wave lasers. $\text{CaF}_2:\text{Yb}^{3+}$ also possesses higher thermal conductivity and lower thermo-optic response than $\text{SrF}_2:\text{Yb}^{3+}$.

- (E) *Radioactivity*: Uranium was mainly used in small amounts for yellow glass and pottery glazes before the discovery of radioactivity (The Homer Laughlin China Company 2011). Then, radium was extracted by Marie Curie after discovering and isolating uranium, which was used in paints for clock and aircraft dials (Newscientist.com). After the radioactivity of uranium was discovered, uranium metal was used for X-ray targets to generate high-energy X-rays (Hammond 2000). Moreover, uranium is suitable for estimating the age of the earliest igneous rocks because of the long half-life of the isotope uranium-238 (4.51×10^9 years). Uranium was also used as a toner (particularly uranium nitrate) in photographic chemicals, in lamp filaments for stage lighting bulbs, in dentures, and as stains and dyes in the wood industry (US Environmental Protection Agency; Uranium containing dentures).
- (F) *Cathodoluminescence*: Cathodoluminescence is an optical and electromagnetic phenomenon in which electrons can affect the emission wavelength of luminescent materials such as $\text{Y}_2\text{O}_3\text{S}:\text{Eu}^{3+}$, $\text{Gd}_2\text{O}_3\text{S}:\text{Tb}^{3+}$, $\text{SrGa}_2\text{S}_4:\text{Eu}^{3+}/\text{Ce}^{3+}$, $(\text{Zn}/\text{Cd})\text{S}:\text{Cu},\text{Al}$, and $\text{ZnS}:\text{Ag},\text{Cl}$. Sulfide-based phosphors have been utilized as efficient low-voltage field emission displays (FEDs) (Zhang et al. 1998; Vecht et al. 1999). Light originates from the excitation of electrons on the phosphor-coated inner surface of the screen. The operation principle of FEDs is basically similar to that of conventional cathode ray tubes (CRTs) but has many effects on the cathodoluminescent properties of phosphors, such as their morphology, size, composition, surface, and crystallization (Li et al. 2010). The energy and probability of a photon relatively depend on the aforementioned factors.

Table 1 Maximum value of Eu^{3+} CT transition in several host lattices (Blasse 1972)

Host lattice	Maximum Eu^{3+} CT (10^3 cm^{-1})
YPO_4	45
YOF	43
Y_2O_3	41.7
LaPO_4	37
La_2O_3	33.7
LaOCl	33.3
$\text{Y}_2\text{O}_2\text{S}$	30

Host Lattice

The same activator is located at different host lattices. The optical properties of the luminescent center are typically different because the surroundings of this center change with the variations in crystal structure. Several factors are considered in this subsection. The first factor is covalency (Jørgensen 1962a, 1971; Lever 1984; Duffy 1990). An increase in covalency induces a reduction in the interaction among electrons because they spread out over wide orbitals. Accordingly, electronic transition shifts to a lower energy with the increase in covalency. This situation is known as the nephelauxetic effect, which will be explained later. For example, the charge transfer (CT) absorption band of Eu^{3+} in YF_3 has higher energy than that in Y_2O_3 because oxide has a more covalent structure. Table 1 shows the maximum value of Eu^{3+} CT transition in several host lattices (Blasse 1972). The second factor is crystal field, which denotes the influence of the host lattice on the optical properties of a dopant. For example, Cr_2O_3 and $\text{Al}_2\text{O}_3:\text{Cr}^{3+}$ are green and red, respectively. Cr^{3+} ions occupy smaller Al^{3+} sites, which results in a stronger crystal field than that in Cr_2O_3 . The mechanism of the interactions is expressed as follows: different host lattices \rightarrow different crystal fields \rightarrow different splittings (Blasse and Grabmaier 1994).

In the design viewpoint, the host is composed of several different cations combined with several different anions. The host cation is restricted to ions with a rare gas electron configuration or closed electron shells and, thus, is optically inactive. Moreover, we need anions that have a balanced charge with cations to stabilize the host crystal. The anions can be divided into two groups, namely, anions that are optically inert and anions that are optically active. The optically inert anions are used to design phosphors because nearly all anions are encompassed by oxygen. Many types of compounds involve silicate, borate, phosphate, aluminate, germinate, and oxide.

The complete composition of phosphors is shown in Fig. 2 (Jüstel 2005). Three types of cation, namely, rare earth (RE), transition metal (TM), and S^2 ions (e.g., Sn, Sb, Tl, Pb, and Bi), can work as activators in phosphor. An important point that should be emphasized is the collocation of the host and the activator in phosphor. For example, Ce-doped lutetium yttrium aluminum garnet ($\text{Lu}_3\text{Al}_5\text{O}_{12}:\text{Ce}$; LuAG) is a green-yellow light-emitting phosphor, which has an isostructure with $\text{Y}_3\text{Al}_5\text{O}_{12}:\text{Ce}$. The Lu^{3+} of lanthanides can be used to form the host crystal because of the outer closed electron shells (e.g., $\text{Lu}^{3+} = 4f^{14}5s^25p^6$).

Composition \Rightarrow Host Lattice + Activator

Activator = RE ions, TM ions, s^2 ions

													13	14	15	16	17	18								
1																	2									
1	H																	He								
3	Li	Be												B	C	N	O	F	Ne							
11	Na	Mg												Al	Si	P	S	Cl	Ar							
19	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr								
37	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe								
55	Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn								
87	Fr	Ra	Ac	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg															
													58	59	60	61	62	63	64	65	66	67	68	69	70	71
													Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
													90	91	92	93	94	95	96	97	98	99	100	101	102	103
													Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr

Fig. 2 Phosphor composition in the periodic table (Jüstel 2005)

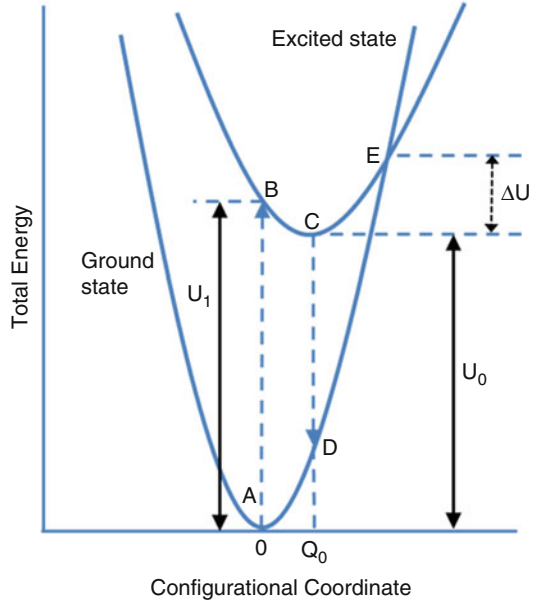
Configurational Coordinate Diagram

The optical properties of a localized center are determined using a configurational coordinate model, particularly the effect of lattice vibrations, as shown in Fig. 3. For simplicity, we can practically regard ions as isolated molecules by neglecting the effects of other distant ions in this model. Based on this concept, the large number of actual vibrational modes of the lattice can be approximated by a small number or a combination of specific normal coordinates. These normal coordinates are called configurational coordinates. In the schematic of the configurational coordinate model, the variable on the abscissa is the interatomic distance and that on the ordinate is the total energy, which is the sum of the electron energy and ion energy. We can determine that the bonding force between a luminescent ion and its nearest neighbor ion depends on Hooke's law. The deviation from the equilibrium position of the ions is considered the configurational coordinate, which is denoted as Q . The total energy of the ground state, U_g , and the total energy of the excited state, U_e , are given by the following relations (Klick and Schulman 1997; Curie 1963; Maeda 1963):

$$U_g = K_g \frac{Q^2}{2} \quad (1)$$

$$U_e = K_e \frac{(Q - Q_0)^2}{2} + U_0 \quad (2)$$

Fig. 3 Configurational coordinate diagram. $A \rightarrow B$ indicates the absorption transition (vertical broken lines) and $C \rightarrow D$ indicates the emission of light



where K_g and K_e are the force constants of the chemical bond, Q_0 is the interatomic distance at the equilibrium of the ground state, and U_0 is the total energy at $Q = Q_0$.

This model has several qualitative facts, which are as follows:

1. Absorption: At absolute zero, the optical absorption proceeds from the equilibrium position of the ground state to the excited state, as exhibited by the arrow in $A \rightarrow B$.
2. Emission: State B relaxes to the equilibrium position C of the excited state before it emits luminescence, which is followed by the emission process $C \rightarrow D$ and the relaxation process $D \rightarrow A$, thereby completing the cycle.
3. Stokes shift: The energy of absorption is higher than the energy of emission because nonradiative relaxation occurs in most cases. This difference further induces a change in the equilibrium position and the force constant of the ground and excited states.
4. Activation energy (ΔU): The energy between two vibration curves intersects E with the lower excited state of each vibration curve assisted by thermal energy and can relax to the ground state nonradiatively (Kamimura et al. 1969).

Based on the aforementioned description, we can derive the transition probability (N) of a nonradiative relaxation process, as follows:

$$N = s \exp \frac{-\Delta U}{kT} \quad (3)$$

where s can be regarded as a constant because it is weakly dependent on temperature, is called the frequency factor, and is typically in the order of 10^{13} s^{-1} . By using Equation (3) and letting W be the luminescence probability, luminescence efficiency η can be expressed as follows:

$$\eta = \frac{W}{W + N} = \left[1 + \frac{s}{W} \exp \frac{-\Delta U}{kT} \right]^{-1} \quad (4)$$

The nonradiative process on the equilibrium position of the excited state C is located outside the configurational coordinate curve of the ground state. The excited state intersects the ground state in the relaxation process from B to C.

Selection Rule

The electronic configurations of RE ions, which normally include 15 lanthanide ions from La to Lu, Sc, and Y ions, are shown in Table 2 (Shionoya and Yen 1998). The azimuthal quantum number (l) of the $5d$ orbitals is 2, which results in 5 ($2l + 1$) orbitals, whereas that of $4f$ is $2 \times 3 + 1 = 7$ orbitals. The spin orbit coupling of the electronic state can be expressed by $^{2S+1}L_J$, where L is the orbital angular momentum that represents the term symbols of S, P, D, F, G, H, ..., which correspond to $L = 0$,

Table 2 Electronic configurations of trivalent RE ions in the ground state (Shionoya and Yen 1998)

Ions	Corresponding elements	$4f$ electrons	$(\sum S)$	$(\sum l)$	$(\sum J)$ ($L \pm S$)
Sc ³⁺	Ar		0	0	0
Y ³⁺	Kr		0	0	0
La ³⁺			0	0	0
Ce ³⁺	Xe	↑	1/2	3	5/2
Pr ³⁺	Xe	↑ ↑	1	5	4
Nd ³⁺	Xe	↑ ↑ ↑	3/2	6	9/2
Pm ³⁺	Xe	↑ ↑ ↑ ↑	2	6	4
Sm ³⁺	Xe	↑ ↑ ↑ ↑ ↑	5/2	5	5/2
Eu ³⁺	Xe	↑ ↑ ↑ ↑ ↑ ↑	3	3	0
Gd ³⁺	Xe	↑ ↑ ↑ ↑ ↑ ↑ ↑	7/2	0	7/2
Tb ³⁺	Xe	↑↓ ↑ ↑ ↑ ↑ ↑ ↑	3	3	6
Dy ³⁺	Xe	↑↓ ↑↓ ↑ ↑ ↑ ↑ ↑	5/2	5	15/2
Ho ³⁺	Xe	↑↓ ↑↓ ↑↓ ↑ ↑ ↑ ↑	2	6	8
Er ³⁺	Xe	↑↓ ↑↓ ↑↓ ↑↓ ↑ ↑ ↑	3/2	6	15/2
Tm ³⁺	Xe	↑↓ ↑↓ ↑↓ ↑↓ ↑↓ ↑ ↑	1	5	6
Yb ³⁺	Xe	↑↓ ↑↓ ↑↓ ↑↓ ↑↓ ↑↓ ↑	1/2	3	7/2
Lu ³⁺	Xe	↑↓ ↑↓ ↑↓ ↑↓ ↑↓ ↑↓ ↑↓	0	0	0

1, 2, 3, 4, 5, . . . , respectively; S is the spin angular momentum, $2S + 1$ is the multiplicity; and J is the total angular momentum. The selection rule can be represented as follows:

$$\Delta S = 0$$

$$\Delta L = 0 \text{ or } \pm 1$$

$$\Delta J = 0 \text{ or } \pm 1 (J = 0 \rightarrow J = 0 \text{ is forbidden})$$

the J value of the lowest ground state is determined as follows:

$J = L + S$ when the number of $4f$ electrons is greater than 7

$J = L - S$ when the number of $4f$ electrons is less than 7

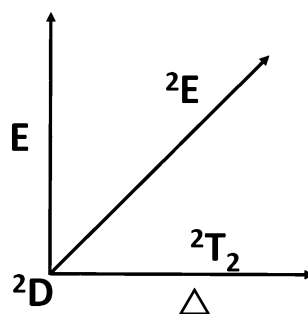
Activator

The activator in phosphors can emit visible light after down-converting the absorbed radiation from an incident light source. In this section, several specific ions are discussed in detail to understand the luminescent properties of phosphors. The transition process is categorized into three types, namely, $3d-3d$, $4f-4f$, and $4f-5d$.

$3d-3d$ Transitions of TM Ions

The electron configuration of TM ions is d^n ($0 < n < 10$), which has an incompletely filled d shell. The energy levels were calculated by Tanabe and Sugano (Kamimura et al. 1969) based on mutual interaction among d electrons, such as the crystal field. In this subsection, we briefly describe the configurations d^1 , d^3 , and d^5 . The free ion has a fivefold orbital degeneracy (2D), which is split into two levels (2E and 2T_2), as shown in Fig. 4 (d^1 configuration). In the octahedral coordination situation, the possible absorption transition is from 2T_2 to 2E , which is equal to the crystal field strength (Δ) (Shriver et al. 1990; Cotton 1990). The crystal field strength value is approximately $20,000 \text{ cm}^{-1}$ for a trivalent TM. For example, the absorption

Fig. 4 Energy levels of the d^1 configuration as a function of the octahedral crystal field Δ . 2D is the free ion level (Blasse and Grabmaier 1994)



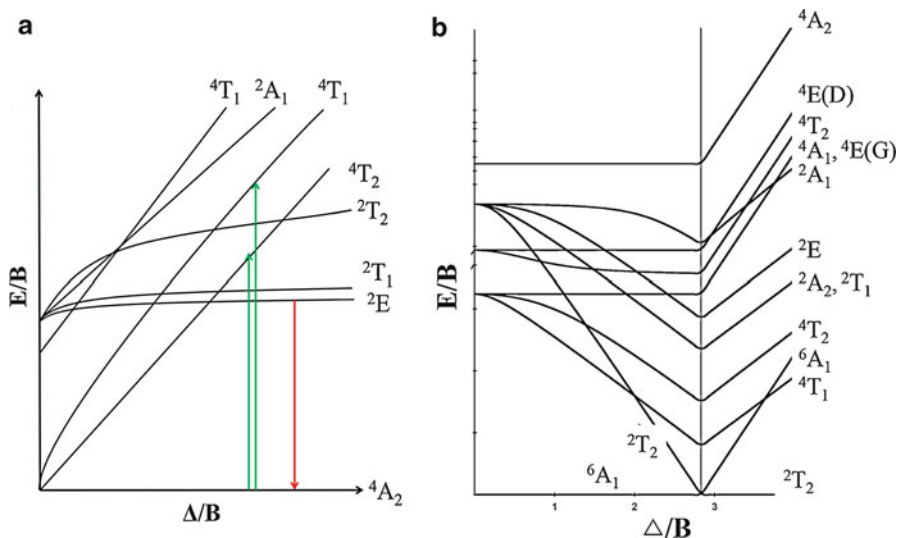


Fig. 5 Tanabe-Sugano energy level diagrams for (a) d^3 and (b) d^5 ions in the octahedral coordinate environment (Kamimura et al. 1969; Figgis and Hitchman 2000)

band of Ti^{3+} ($3d^1$) is approximately $20,000 \text{ cm}^{-1}$ (Blasse and Grabmaier 1994). The Tanabe-Sugano energy level diagrams of different electron numbers (d^3 and d^5 ions) are shown in Fig. 5. Based on the aforementioned selection rules, the absorption spectrum of Mn^{4+} is composed of two broad peaks in K_2SiF_6 compound. These two peaks are spin-allowed transition from ${}^4A_{2g}$ to ${}^4T_{1g}$ (4F and 4P) and ${}^4A_{2g}$ to ${}^4T_{2g}$ (Fig. 5a), which have absorption wavelengths of approximately 350 and 450 nm, respectively. The spin quartet ($2S + 1 = 4$) level occurs and coincides with the spin selection rule. The ground state of Mn^{2+} (d^5 configuration) is 6A_1 in Fig. 5b. All absorption transitions are spin forbidden and parity forbidden. Notably, the ${}^6A_1 \rightarrow {}^4A_1$ and 4E bands are narrow. By contrast, the ${}^6A_1 \rightarrow {}^4T_1$ and 4T_2 bands are broad. The condition involves coupling with vibrations. The broad excitation peak may reflect fluctuations in a large metal–ligand (or cation) distance.

4f–4f Transitions of RE Ions (Line Emission)

Based on an incompletely filled 4f shell, the 4f–4f transitions of trivalent RE ions are small because the $5s^2$ and $5p^6$ orbitals shield the 4f electrons. The influence of the host lattice on optical transitions is small for the $4f^n$ configuration. Therefore, the 4f–4f transitions of trivalent RE ions are normally the same in various host lattices. Thus, different energy levels and emission peaks can be precisely assigned in Dieke's diagram, as shown in Fig. 6 (Dieke 1968).

The Franck-Condon principle describes the intensity of vibronic transitions in spectroscopy and quantum chemistry. Vibronic transitions in a molecule are the simultaneous changes in electronic and vibrational energy levels caused by the absorption or emission of a photon with suitable energy. The rule mainly states the

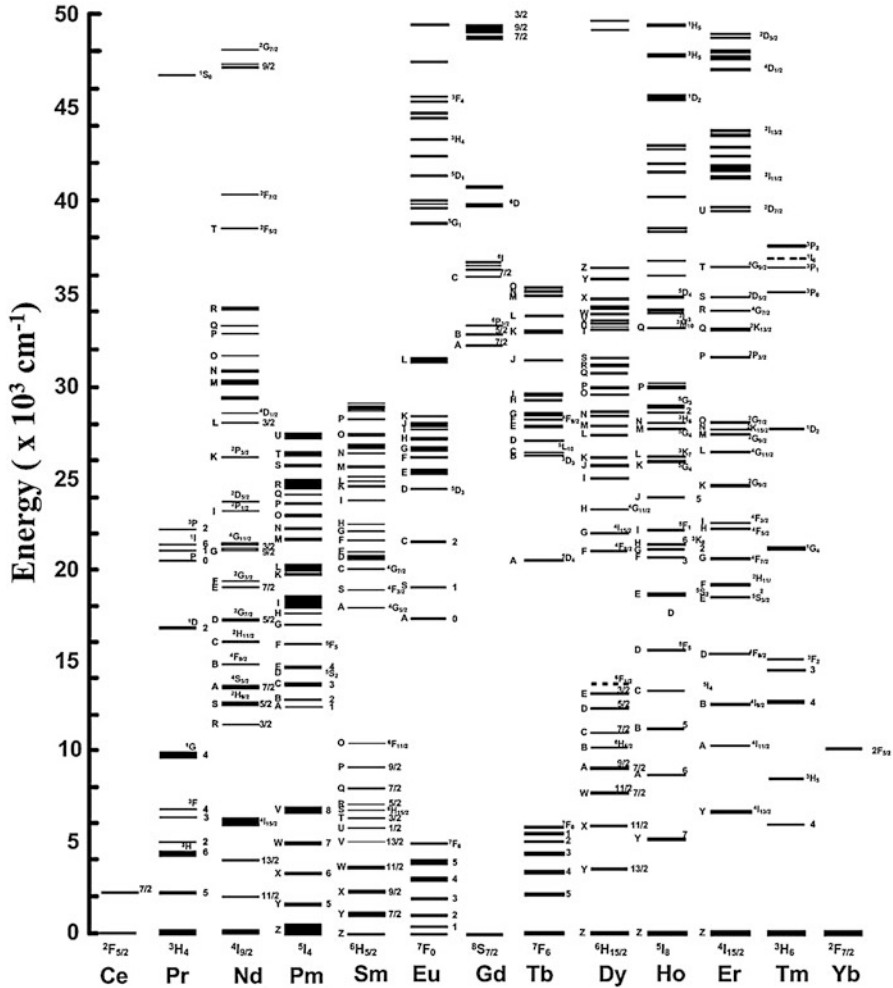
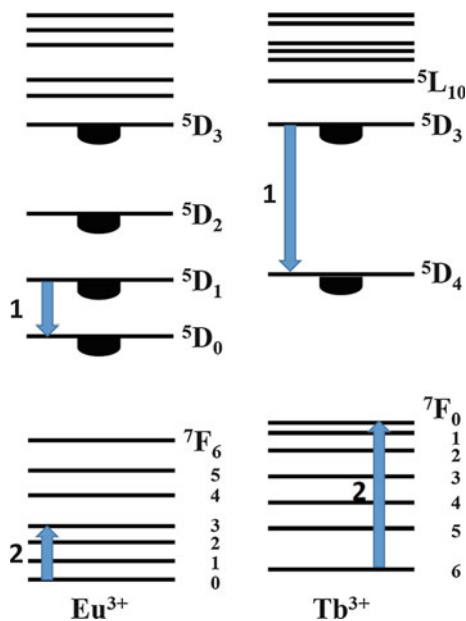


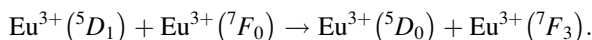
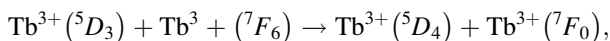
Fig. 6 Dieke's diagram of free ion energy levels for trivalent RE ions (Dieke 1968)

change between two overlapping vibrational wave functions during electronic transition. Based on the Franck-Condon principle (Nakazawa 2006; Condon 1926), the ΔR between the ground and excited states of $4f-4f$ transitions is zero. This transition is called zero transition or non-phonon transition. Thus, sharp line emissions are observed in the $4f-4f$ transitions of trivalent RE ions. The emitting intensities of these transitions are weak because of the forbidden transitions in the selection rule ($\Delta f = 0$, parity forbidden). Furthermore, the cross-relaxation process (Shionoya and Yen 1998; Nakazawa 2006) of $4f-4f$ transition metal-based phosphors significantly influences emission properties and color changes. If some high energy-level emissions of Eu^{3+} and Tb^{3+} are quenched by transferring energy to the other ions, then this process is

Fig. 7 Cross-relaxation of the energy levels of Eu^{3+} (left) and Tb^{3+} (right) (Blasse and Grabmaier 1994)



called cross-relaxation. The following reaction formulas are frequently derived when the concentration of activators is high (Blasse and Grabmaier 1994):



For low Eu^{3+} concentrations (approximately 0.1 mol%), the 5D_3 , 5D_2 , 5D_1 , and 5D_0 emissions of Eu^{3+} ions are observed in YBO_3 and Y_2O_3 compounds. By contrast, only 5D_0 dominates the emission of Eu^{3+} ions when Eu^{3+} content is 3 %. The high energy of Eu^{3+} ions is quenched by transferring energy to other ions, which are promoted to the 7F_3 level, as shown in Fig. 7 (left). The increase in the concentration of the Tb activator will result in a color change from blue to green because the excited energy of $^5D_3 \rightarrow ^5D_4$ is transferred to other ions, which are promoted to the 7F_0 level, as shown in Fig. 7 (right). This phenomenon occurs because of the resonance between 5D_J (5D_3 to 5D_4) and 7F_J (7F_6 to 7F_0) states.

Considering other transition processes, the CT band, which denotes the position of the trivalent RE ion in the host lattice, can be predicted by Jørgensen's expression (Jørgensen 1962b), as follows:

$$E_{\text{CT}} = [\chi_{\text{opt}}(X) - \chi_{\text{uncorr}}(M)] 30 \times 10^3 \text{ cm}^{-1} \quad (5)$$

where $\chi_{\text{opt}}(X)$ is the optical electronegativity of the ligand ion (similar to Pauling's electronegativity) and $\chi_{\text{uncorr}}(M)$ can be calculated using Su's expression (Su 1991), as follows:

$$E_{\text{Ln}}^0 \left[\text{Ln}^{n+} \rightarrow \text{Ln}^{(n-1)+} \right] = 4.273\chi_{\text{uncorr}}(M) - 7.776 \quad (6)$$

An example is the binding of a Tb^{3+} RE ion with oxygen and nitrogen ligands, where $E_{\text{Tb}}^0 (\text{Tb}^{3+} \rightarrow \text{Tb}^{2+})$ is determined to be -3.7 eV (Su 1991). $\chi_{\text{uncorr}}(\text{Tb})$ is predicted to be 0.75. Thus, E_{CT} values are predicted to be 64,500 and 52,500 cm^{-1} , which correspond to 155 and 190 nm, respectively, whereas $\chi_{\text{opt}}(\text{O})$ and $\chi_{\text{opt}}(\text{N})$ are approximately 3.1 and 2.7, respectively.

4f–5d Transitions of RE Ions (Broad Emission)

In several broad emission bands of RE ions, $\text{Ce}^{3+} (4f^1)$ and $\text{Eu}^{2+} (4f^7)$ RE ions are the most common 4f–5d transition activators in the host lattice of phosphors, in which the excited states are $4f^65d^1$ and $4f^65d^1$, respectively. We analyze the influences of optical properties and discuss the emission transitions of trivalent ion (Ce^{3+}) and divalent ion (Eu^{2+}).

The transitions between the ground and excited states are parity allowed (i.e., $\Delta f = \pm 1$, the emissions of Eu^{2+} - and Ce^{3+} -doped phosphors have broad spectra). If these types of activator are under an environment with crystal field strength, then the five 5d orbitals (d_{xy} , d_{xz} , d_{yz} , $d_{x^2-y^2}$, and d_z^2) shown in Fig. 8 (Shionoya and Yen 1998) will split within a gravity center. If the local environment is an octahedral site, then the five 5d orbitals will degenerate into two levels (high potential E_g level ($d_{x^2-y^2}$ and d_z^2) and low potential T_{2g} level (d_{xy} , d_{xz} , and d_{yz})). Crystal field splitting between E_g and T_{2g} is represented by Δ_o , which is equal to $10 Dq$. The Dq value can be expressed as follows:

$$Dq = \frac{35Ze}{4R^5} \quad (7)$$

where Dq is the crystal field strength, Z is the valence state of the anions, e is the electron charge, and R is the average distance between the activator and the coordinated anions. Crystal field strength is determined by the type of coordinated ligand. The spectrochemical series determines the splitting of the 5d orbitals and directly influences the emission wavelength in the transition process from the lowest 5d excited state to the 4f ground state. The series listing of ligands from weak to strong fields is depicted as follows:

Spectrochemical series:

(Weak field) $\text{I}^- < \text{Br}^- < \text{S}^{2-} < \text{SCN}^- < \text{Cl}^- < \text{NO}_3^- < \text{N}_3^- < \text{F}^-$
 $< \text{OH}^- < \text{C}_2\text{O}_4^{2-} \approx \text{H}_2\text{O} < \text{NCS}^- < \text{CH}_3\text{CN} < \text{py}$ (pyridine) $< \text{NH}_3$
 $< \text{En}$ (ethylenediamine) $< \text{bipy}$ (2,2'-bipyridine) $< \text{phen}$ (1,10-phenanthroline)
 $< \text{NO}_2^- < \text{PPh}_3 < \text{CN}^- \approx \text{CO}$ (strong field)

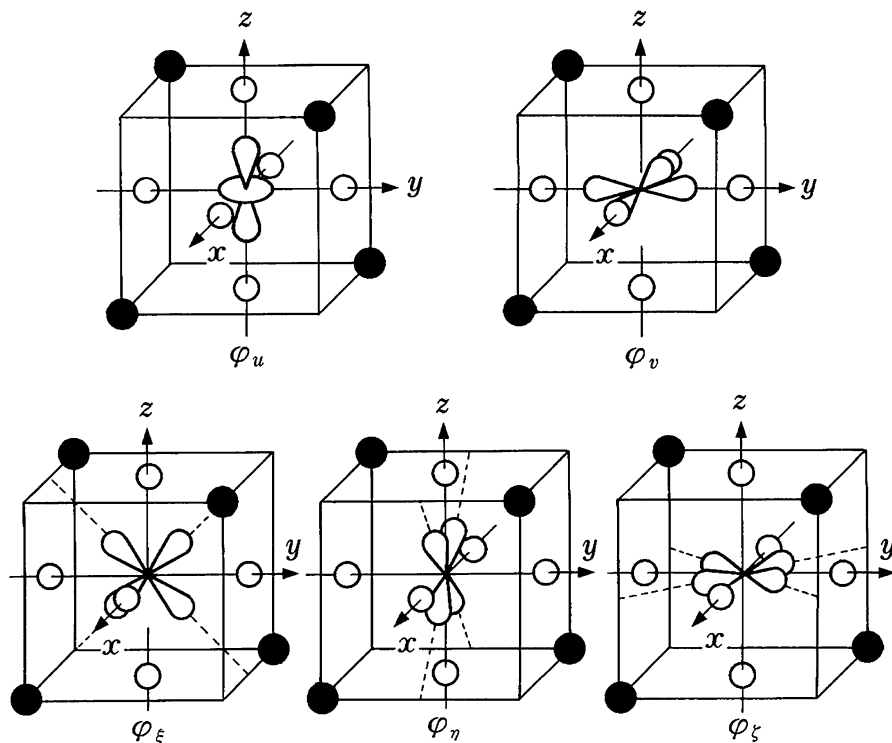


Fig. 8 Schemes of the $5d$ orbitals (d_{yz} , d_{xz} , d_{xy} , $d_{x^2-y^2}$, and d_{z^2}) and ligand positions. *White spheres*: ligands for octahedral (Oh) symmetry. *Black spheres*: ligands for tetrahedral (Td) symmetry (Shionoya and Yen 1998)

Another well-known effect that influences the gravity of $5d$ orbitals is the nephelauxetic effect. The word “nephelauxetic” means “cloud expanding” in Greek. The nephelauxetic effect refers to the covalency of the bond between the activator and the ligands where many unpaired electrons of the metal ion will strongly interact with the electrons of ligands. For the same metal ion, the spin-pairing energy for different ligands follows the sequence of $F^- > H_2O > NH_3 > Cl^- > CN^- > Br^- > I^-$, which is called a nephelauxetic series. The empirical prediction equation for the spin-pairing energy of the nephelauxetic effect is expressed as follows:

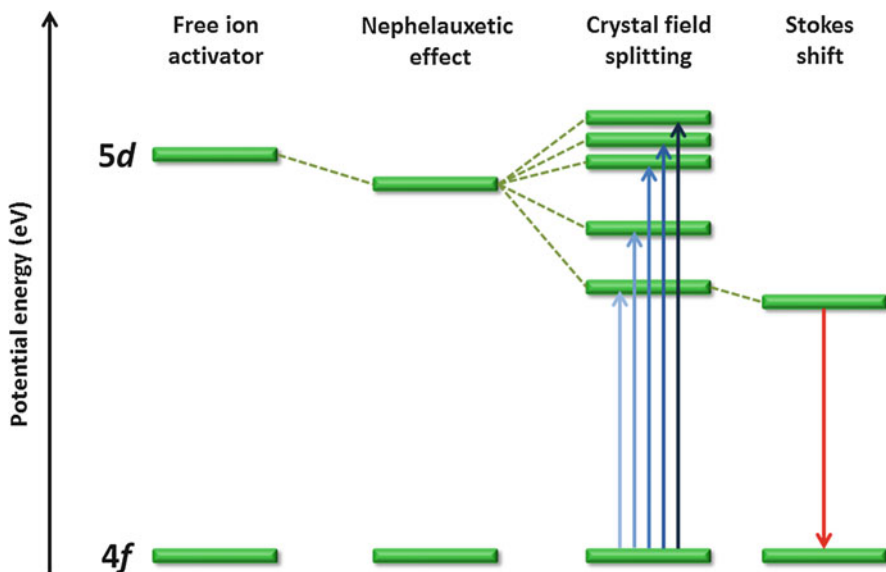
$$B = B_0 \times (1 - h \times k) \quad (8)$$

where B is the spin-pairing energy; B_0 is the spin-pairing energy of B for the free ion; and h and k are the empirical interelectronic repulsion parameters for the ligands and metal ions, respectively. The h values are shown in Table 3.

Thus, the gravity of $5d$ orbitals downshifts toward a low potential position (i.e., the centroid shift) because of the strong interaction of the electron cloud between the

Table 3 Interelectronic repulsion parameters for different ligands

Ligands	h
6 Br ⁻	2.3
6 Cl ⁻	2.0
6 CN ⁻	2.0
3 en	1.5
6 NH ₃	1.4
6 H ₂ O	1.0
6 F ⁻	0.8

**Fig. 9** Schematic of the energy levels of $4f$ - $5d$ transitions (Shionoya and Yen 1998; Xie et al. 2008)

activator and ligand ions (i.e., the covalency of the activator and ligands is large because the repulsion force is small). Although the spectrochemical and nephelauxetic series are mainly adopted for TM ions, the trend of the interactions of different ligand anions with RE ions is the same. Figure 9 shows the overall effect of the host lattice on the activator, which influences the excitation and emission properties of $4f$ - $5d$ transitions (Shionoya and Yen 1998; Xie et al. 2008). The polarization and covalency of ligand anions affect the centroid shift, which results in the nephelauxetic effect. The geometry of the anion polyhedron, the coordination number (CN), and the type of coordinated anion are related to crystal field splitting. The shape of the excitation spectrum denotes the splitting of $5d$ orbitals. The emission spectrum shifts toward a long wavelength region because of the centroid shift from the nephelauxetic effect and the lowest $5d$ orbital position from crystal field splitting.

CN has an important effect on emission properties because different types of coordinated anion influence covalency (related to the nephelauxetic effect) and bond distance (related to the crystal field effect). The radial distribution function, which is an effective method to estimate CN, can be calculated in disordered systems. The first CN can be defined as follows (Waseda 1980; Vahvaselkä and Mangs 1988):

$$n_1 = 4\pi \int_{r_0}^{r_1} r^2 g(r) \rho dr \quad (9)$$

where r_0 is the rightmost position starting from $r = 0$ wherein $g(r)$ is approximately zero and r_1 is the first minimum. Therefore, r_1 denotes the area under the first peak of $g(r)$. The second CN is defined similarly as follows:

$$n_2 = 4\pi \int_{r_1}^{r_2} r^2 g(r) \rho dr \quad (10)$$

First, the $4f^1$ ground state of the Ce^{3+} ion is separated into ${}^2F_{5/2}$ and ${}^2F_{7/2}$ through spin-orbit coupling. The difference of both levels is approximately $2,000 \text{ cm}^{-1}$. The $5d^1$ excited state is split into five components by the crystal field, as mentioned previously. Maximum splitting reaches $15,000 \text{ cm}^{-1}$, with its typical emission double-band shape occurring from the lowest excited state ($5d^1$) to the two levels of the ground state (${}^2F_{5/2}$ and ${}^2F_{7/2}$). The $5d \rightarrow 4f$ transition is parity allowed and is unsuitable for spin selection. The lifetime of Ce^{3+} emission is short and proportional to the square of the emission wavelength ($\tau \sim \lambda^2$) (Di Bartolo 1968).

Furthermore, the $4f \rightarrow 5d$ transition of trivalent RE ions in the host lattice can be predicted by Dorenbos' expression (Dorenbos 2000a, b, c; Zhang et al. 2012; Yang et al. 2006), as follows:

$$D(\text{Ln}, \text{A}) = E(\text{Ln}, \text{free}) - E(\text{Ln}, \text{A}) \quad (11)$$

$$E(\text{Ln}, \text{A}) = E(\text{Ce}, \text{free}) - D(\text{Ln}, \text{A}) + \Delta E^{\text{Ln}, \text{Ce}} \quad (12)$$

where $D(\text{Ln}, \text{A})$ is the crystal field depression of the $4f^{n-1}5d$ levels of a lanthanide ion (Ln^{3+}) in compound A relative to the energy in the free ion, $E(\text{Ln}, \text{free})$ is the energy of the first $f-d$ transition of Ln^{3+} as a free ion (gaseous), $E(\text{Ln}, \text{A})$ is the $f-d$ energy difference of Ln^{3+} -doped compound A with $D(\text{Ln}, \text{A})$, and $\Delta E^{\text{Ln}, \text{Ce}}$ is the difference in $f-d$ energy of Ln^{3+} with that of the first electric dipole-allowed transition of Ce^{3+} . The effect of the crystal field and covalency of the host lattice on the red shift of the $5d$ levels is approximately equal for all RE ions.

Second, the well-known Eu^{2+} ($4f^7$) ion exhibits a $5d \rightarrow 4f$ transition that can vary from UV to yellow, depending on different host lattices, which is determined by the aforementioned factors, such as in the Ce^{3+} ion case. The high energy level of the $4f^65d$ configuration is observed when the crystal field is weak and covalency is low. The ${}^6P_{7/2}$ level of the $4f^7$ ground state is located below the lowest component of the excited configuration, such as $\text{SrB}_4\text{O}_7:\text{Eu}^{2+}$ (Meijerink

et al. 1989). At low temperatures, several line emission peaks can be ascribed to the parity-forbidden ${}^6P_{7/2} \rightarrow {}^8S_{7/2}$ transition. By contrast, a broad emission band of $5d \rightarrow 4f$ transition is observed at high temperatures. This finding indicates that environmental temperature is also a factor that influences the luminescent properties of Eu^{2+} , except for the previously discussed effects.

Fundamentals of Phosphor

Various properties of phosphors must be considered in practical and actual applications. For example, the excitation spectra of suitable phosphors combined with UV- or blue-chip sources have been developed and are extensively used. We list the excitation regions of several phosphors applied in blue- and UV light-emitting diodes (LEDs) in Tables 4 and 5, respectively. Moreover, other important factors, such as intensity (quantum efficiency), durability, thermal stability, and bandwidth, are discussed. We discuss these factors and the principle of the energy conversion process of phosphors in this section. In addition, the most representative applications of phosphors are also described.

Requirements for Phosphor-Converted LEDs (pc-LEDs)

Extreme phosphors are composed of a host (matrix) and an activator(s), except for special situations, such as vacancy defect-induced and ligand exchange-induced lighted compounds. The luminescent processes of phosphors will be affected by the intrinsic properties of the components, external impurities, the interaction between the matrix and the activators, and their specific effects. pc-LEDs typically comprise blue- or UV-excited LEDs and partially convert light into blue, yellow, green, and red by using single or multiple phosphors. The main requirements for pc-LED performance and the development of LED-excited phosphors are discussed in this section (as shown in Fig. 10) (Smet et al. 2011).

(A) *Emission property*: Figure 11 shows the emission and excitation spectra of $\text{Y}_3\text{Al}_5\text{O}_{12}:\text{Ce}^{3+}$ (0.05 mol%) as an example, with its crystal structure categorized as garnet symmetry, as shown in the inset of Fig. 11. The broad emission spectrum is ascribed to the excited electron ($5d^1$) released to the ground state ($4f$), which is split into ${}^2F_{5/2}$ and ${}^2F_{7/2}$ states by spin-orbit coupling. The relative shape of the emission spectra denotes the intrinsic properties of the activators and host lattice. We also present our theory in the subsequent paragraphs. Most commercial white LEDs (WLEDs) are based on the combination of blue LED chips and yellow-emitting YAG: Ce^{3+} phosphors. However, such approach limits the application range to cool white light (correlated color temperature (CCT) of 4,000–8,000 K) and limits the color rendering index (CRI < 75) because of the lack of red emission in the luminescent spectrum (Nguyen et al. 2014). The CRI is used to measure the capability of a light source

Table 4 Blue LED-excited phosphors for WLEDs (⊗: best, ○: good, Δ: poor)

LED	Color	Composition	Emission properties			Excitation bandwidth (nm)	
			Intensity ^a	Durability ^b	Thermal stability ^c		Bandwidth ^d
Blue LED	Green	β -SiAlON:Eu (Xie et al. 2007)	○	⊗	○	250–500	
		$\text{Lu}_3\text{Al}_5\text{O}_{12}$:Ce (Li et al. 2007)	Δ	○	⊗	300–500	
		(Sr,Ba) $_2$ SiO $_4$:Eu (Kim et al. 2005a)	⊗	Δ	Δ	250–500	
		(Sr,Ba)Si $_2$ O $_7$:Eu (Lei et al. 2011)	⊗	○	⊗	250–525	
		Y $_3$ (Al,Ga) $_3$ O $_7$:Ce (Hansel et al. 2009)	○	⊗	○	300–500	
		SrGa $_2$ S $_4$:Eu (Arai et al. 2005)	⊗	Δ	Δ	300–500	
		Ca $_3$ Sr $_2$ Si $_3$ O $_7$:Eu (Nagai 2011)	⊗	⊗	⊗	300–500	
		Ca- α -SiAlON:Eu (Li et al. 2008a)	Δ	○	○	250–550	
		(Y,Gd) $_3$ Al $_5$ O $_7$:Ce (Park et al. 2009)	⊗	○	○	400–500	
		Tb $_3$ Al $_5$ O $_7$:Ce (Zorenko et al. 2009)	○	○	○	400–520	
Yellow		(Ca,Sr,Ba) $_2$ SiO $_4$:Eu (Kim et al. 2005a)	○	Δ	Δ	250–500	
		La $_3$ Si $_6$ N $_7$:Ce (Seto et al. 2009)	○	○	○	300–500	
		CaGa $_2$ S $_4$:Eu (Najafov et al. 2002)	⊗	Δ	Δ	350–550	
		Sr $_2$ Si $_5$ N $_8$:Eu (Li et al. 2008b)	○	○	○	250–600	
		CaAlSiN $_3$:Eu (Piao et al. 2007)	⊗	⊗	⊗	200–600	
		(Sr,Ca)S:Eu (Van Haecke et al. 2007)	⊗	Δ	Δ	250–620	
		(Sr,Ba) $_3$ SiO $_5$:Eu (Jang et al. 2009)	⊗	Δ	⊗	300–550	
		Red					

^aThe emission intensity of phosphor is ⊗: better than, ○: equal to, or Δ: poorer than other same-color phosphors

^bThe emission intensity of phosphor after the durability task (at 80 °C and 80 % relative humidity for over 12 h) is ⊗: equal to, ○: 80 % compared with the original intensity, and Δ: 50 % compared with the original intensity

^cThe decay percentages of phosphor emission intensity compared with the original intensity after the 300 °C task is ⊗: < 10 %, ○: ~ 10 %–30 %, and Δ: > 30 %

^dThe bandwidth of the emission peak is narrow (<80 nm), middle (~80–120 nm), or broad (> 120 nm)

Table 5 UV LED-excited phosphors for WLEDs (⊙: best, ○: good, Δ: poor)

LED	Color	Composition	Emission properties			Excitation bandwidth (nm)	
			Intensity ^a	Durability ^b	Thermal stability ^c		Bandwidth ^d
UV LED	Blue	(Sr,Ba) ₃ MgSi ₂ O ₈ :Eu (Hana et al. 2014)	○	○	○	250–450	
		BaMgAl ₁₀ O ₁₇ :Eu (Yadav et al. 2010)	⊙	Δ	○	200–400	
Green		β-SrAlON:Eu (Xie et al. 2007)	○	⊙	○	250–500	
		(Sr,Ba)SiO ₄ :Eu (Kim et al. 2005a)	⊙	Δ	Δ	250–500	
		(SrBa)Si ₂ O ₂ N ₂ :Eu (Lei et al. 2011)	○	○	○	250–525	
		Ba ₃ Si ₆ O ₁₂ N ₂ :Eu (Tang et al. 2011)	○	○	○	200–500	
		SrAl ₂ O ₄ :Eu (Peng et al. 2004)	○	○	○	200–450	
		BaMgAl ₁₀ O ₁₇ :Eu,Mn (Zhou et al. 2010)	⊙	⊙	⊙	225–400	
		SrGa ₂ S ₄ :Eu (Do et al. 2000)	⊙	Δ	Δ	300–500	
	Red		Sr ₂ Si ₃ N ₈ :Eu (Li et al. 2008b)	○	○	○	250–600
			CaAlSiN ₃ :Eu (Piao et al. 2007)	⊙	⊙	⊙	200–600
			(Sr,Ca)S:Eu (Van Haecke et al. 2007)	⊙	Δ	Δ	250–620
		La ₂ O ₂ S:Eu (Yap et al. 2009)	○	○	○	250–400	
	Ba ₃ MgSi ₂ O ₈ :Eu,Mn (Kim et al. 2005b)	⊙	○	○	250–450		
	CaSc ₂ O ₄ :Eu (Hao et al. 2011)	⊙	⊙	⊙	350–410		

^aThe emission intensity of phosphor is ⊙: better than, ○: equal to, or Δ: poorer than the other same-color phosphors

^bThe emission intensity of phosphor after the durability task (at 80 °C and 80 % relative humidity for over 12 h) is ⊙: equal to, ○: 80 % compared with the original intensity, and Δ: 50 % compared with the original intensity

^cThe decay percentages of phosphor emission intensity compared with the original intensity after the 300 °C task is ⊙: < 10 %, ○: ~ 10 %–30 %, and Δ: > 30 %

^dThe bandwidth of the emission peak is narrow (<80 nm), middle (~80–120 nm), or broad (> 120 nm)

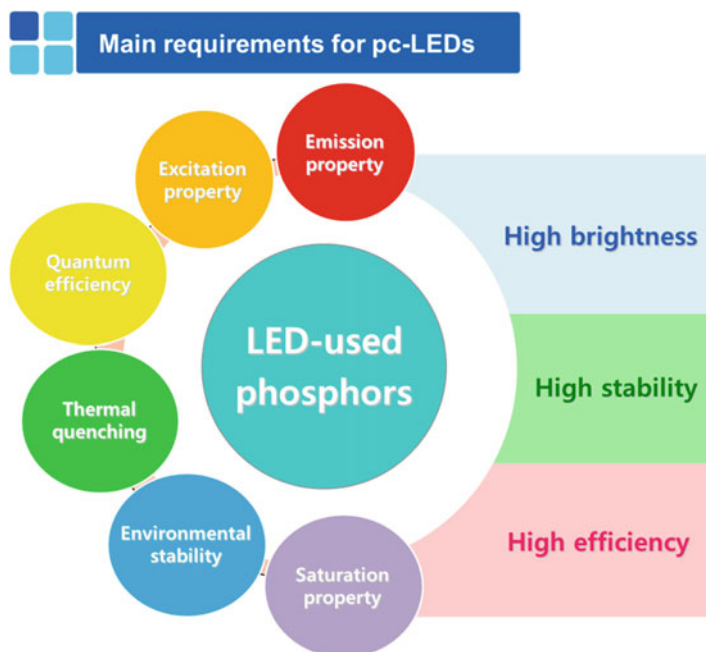


Fig. 10 Main requirements for LED-used phosphors

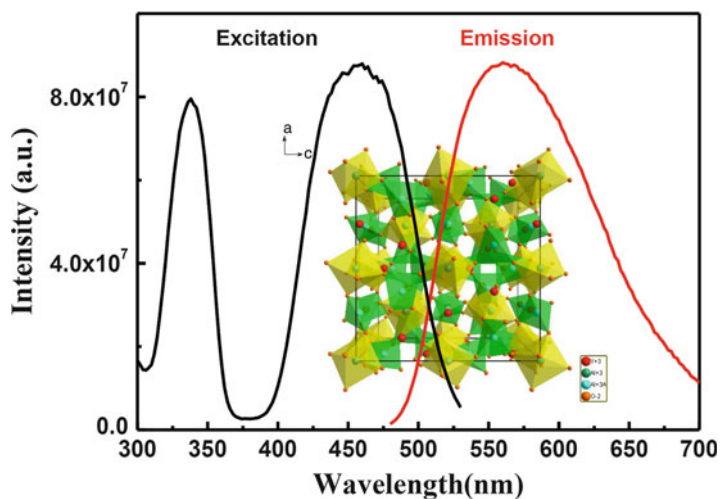


Fig. 11 Emission and excitation spectra of $\text{Y}_3\text{Al}_5\text{O}_{12}:\text{Ce}^{3+}$ compound. *Inset:* Crystal structure of $\text{Y}_3\text{Al}_5\text{O}_{12}:\text{Ce}^{3+}$ (red ball: Y^{3+} ; green ball: $\text{Al}^{3+}(1)$; blue ball: $\text{Al}^{3+}(2)$; orange ball: O^{2-})



Fig. 12 Three images with different CRI values (<http://pinstake.com/colour-rendering-index-cri/>)

to reveal the true colors of objects compared with an ideal or natural light source. Figure 12 (<http://pinstake.com/colour-rendering-index-cri/>) shows distinct images with different CRI values. In addition, CCT is also important for WLEDs because it directly influences human sense of sight. For example, an ideal black-body radiator radiates light with comparable hue to that of the light source at a certain temperature. The CCT of the white light source can be easily adjusted by changing the weight ratio of different phosphors. Figure 13 shows the CCT locus of a Planckian radiator in the International Commission on Illumination (CIE) chromaticity diagram (Yen et al. 2006). The two parameters (CRI and CCT) are significantly correlated with the emission of phosphors.

- (B) *Excitation property*: The wavelengths of normal UV chip and blue chip are 355–400 and 440–460 nm, respectively. The excitation region of phosphors is important for LED packages. Thus, the excitation spectrum should be broad to fit the flexible emission source of semiconductor-based LEDs easily. For example, the excitation spectrum ($4f \rightarrow 5d$ transition) of YAG:Ce³⁺ matches that of blue LED (460 nm), as shown in Fig. 11, which results in the excellent quantum efficiency (QE) of luminescent materials. Consequently, a suitable excitation spectrum should be present near the wavelength of the LED light source to maintain the white color stability of WLEDs. Moreover, the reabsorption phenomenon occurs in the overlap region between the excitation and emission spectra of phosphors. Sakuma et al. (2007) reported that the redshift of the emission wavelength of Ca- α -SiAlON:Eu²⁺ ceramic phosphors is caused by the reabsorption effect because the long excitation band overlaps with the yellow emission band of these phosphors. High activator concentration causes the energy transfer among Eu²⁺ ions, and the redshift can be explained by the classical configurational coordinate model.
- (C) *QE*: Fluorescence quantum yield, which is the ratio of the emitted to the absorbed photons, is generally determined experimentally through careful photoluminescence (PL) measurements using an integrated sphere. QE is an important factor in choosing excellent phosphors for WLEDs. Although determining the internal and external quantum efficiencies of phosphors for use in practical

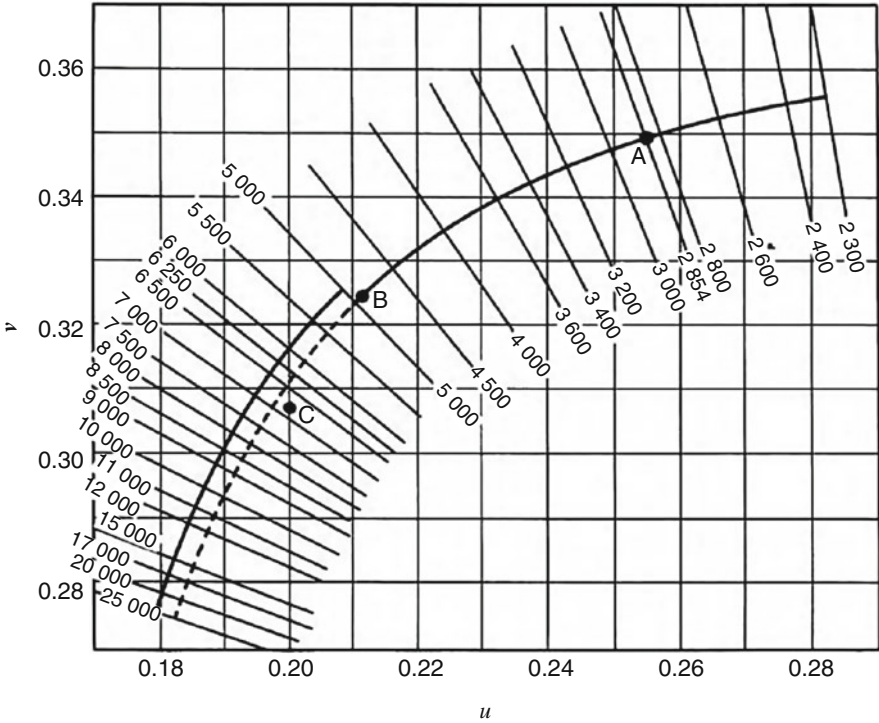


Fig. 13 Locus of Planckian color in the 1964 CIE diagram for a wide range of CCT values (Yen et al. 2006)

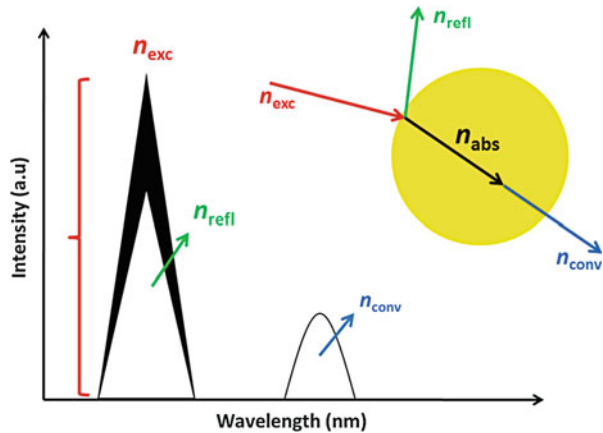
WLEDs is necessary, several studies did not mention the efficiency of phosphors. A general definition of relevant parameters is provided in the following formula:

$$\begin{aligned}
 n_{exc} &= n_{refl} + n_{conv} + n_{nr} \\
 &= n_{refl} + Q_{int}n_{abs} + (1 - Q_{int})n_{abs}
 \end{aligned}
 \tag{13}$$

where n_{exc} is the number of the photons moving toward the phosphors, n_{refl} is the number of photons reflected outside the phosphors, n_{conv} is the number of emitted photons from the phosphors, n_{nr} is the number of photons lost in nonradiative decays, Q_{int} corresponds to the relative number of absorbed photons that are converted into emitted photons, and n_{abs} is the number of excited photons that are quenched inside the phosphors. Internal QE is the ratio of the number of photons collected by the phosphors to the number of photons of a given energy that shines on the phosphors from an external source and is absorbed by the phosphors. Internal QE is illustrated in Fig. 14 and can be derived by the following formula (Smet et al. 2011):

$$Q_{int} = \frac{n_{conv}}{n_{abs}} = \frac{n_{conv}}{n_{exc} - n_{refl}}
 \tag{14}$$

Fig. 14 Scheme of QE (Smet et al. 2011)



The external QE of phosphors should be considered to convert the power of excited photons into optical output power. External QE is the ratio of the number of photons collected by the phosphors to the number of photons of a given energy that is shining on the phosphors from an external source (incident photons), as shown in the following formula (Smet et al. 2011):

$$Q_{ext} = Q_{int} \times \frac{n_{abs}}{n_{exc}} \tag{15}$$

where Q_{ext} is the external QE and n_{abs}/n_{exc} is the absorbance of phosphors. For a practical package, only a small amount of the high-external QE phosphors is required in WLEDs to help reduce energy loss caused by down conversion.

(D) *Thermal quenching*: In high-power semiconductor-based LEDs, approximately 60 % of the electrical input power is converted into heat, which increases the temperature of the entire LED device, including the phosphors. Local heat can reach up to 400–450 K; thus, the thermal quenching (TQ) property of phosphors is important for practical WLED devices. The emission peak of activator-doped phosphors does not only decrease intensity as temperature increases but also broadens the spectral bandwidth because excited electrons rise to higher vibrational microstates. In this study, four mechanisms of the thermal effects on phosphors are introduced. The perfect luminescence of dopants proceeds from the lowest position of the excited state to the ground state without thermal effects, as exhibited by the emission spectrum shown in Fig. 15a (solid green line) (Lin and Liu 2014). However, numerous lanthanide activators display emission spectra with spectral intensities and positions that are easily affected by environmental temperature. Heat is generally detrimental to phosphors, and phosphor efficiency decreases through nonradiative relaxation as device temperature increases. This phenomenon indicates TQ, and consequently, phosphors shift to emission peak wavelengths, which decreases luminescent intensity.

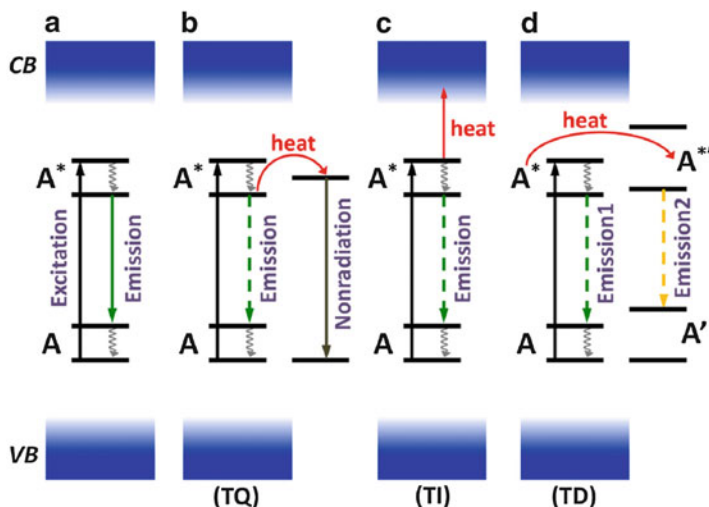


Fig. 15 Luminescent mechanisms of thermal effects on inorganic solids. (a) Emission from a luminescent activator upon excitation. (b) TQ results in a nonradiative pathway associated with heat. (c) TI excites electrons toward the conduction band through heat. (d) TD can lead to other emissions as a result of heat. A and A^* represent the ground and excited states of the activator, respectively. A' and A'^* represent the ground and excited states of the activator with different charges, respectively. VB and CB represent the valence and conduction bands of the host, respectively (Lin and Liu 2014)

Excited electrons can relax through radiative (Fig. 15b; *dashed green line*) and nonradiative (Fig. 15b; *gray line*) processes, such as photon emission and collisional quenching, respectively. The TQ property can also be elucidated using the configurational coordinate diagram shown in Fig. 3, in which thermal energy assists excited electrons across the intersection point between the excited and ground state parabola curves, thereby resulting in nonradiative decay. Figure 15c shows the relative positions between the localized $5d$ electron states of the activators and the delocalized conduction band states of the hosts. First, autoionization spontaneously occurs and no $5d-4f$ emission is observed when the lowest $5d$ state is above the bottom of the conduction band. Such cases include $\text{Ba}_{10}(\text{PO}_4)_4(\text{SiO}_4)_2:\text{Eu}^{2+}$ (Yu et al. 2012), $\text{Ln}_2\text{O}_3:\text{Ce}^{3+}$ (Hirosaki et al. 2003), $\text{LaAlO}_3:\text{Ce}^{3+}$ (van der Kolk et al. 2007), and the Eu^{2+} on trivalent RE sites in oxide compounds (Dorenbos 2003). Second, the $5d$ states of the activators are below the conduction band of the hosts in most $5d-4f$ emission situations. The $5d$ electrons are ionized to the conduction band through thermal ionization (TI), which depends on the energy E_{dC} between the $5d$ state (d) of the activator and the bottom of the conduction band (C) (Lyu and Hamilton 1991; Dorenbos 2005). The activator Eu^{2+} located in the fluffy structure is easily oxidized to the trivalent species at high temperatures. Therefore, the existence of Eu^{3+} can be observed in the PL and X-ray absorption spectra. This phenomenon is called the thermal degradation (TD) effect (Fig. 15d).

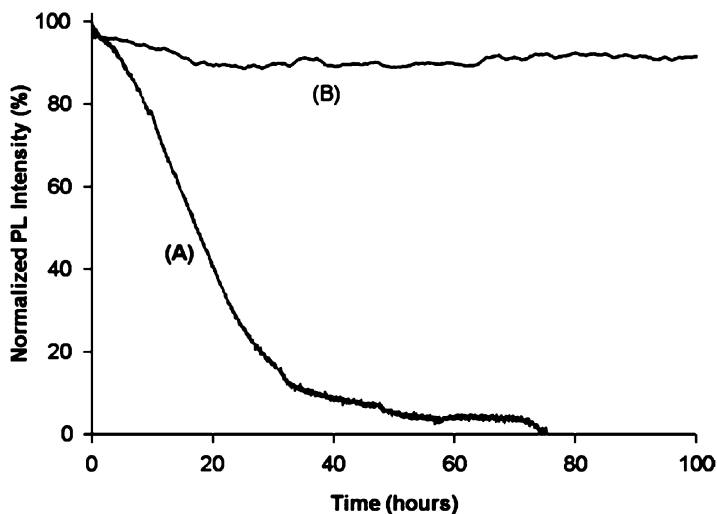


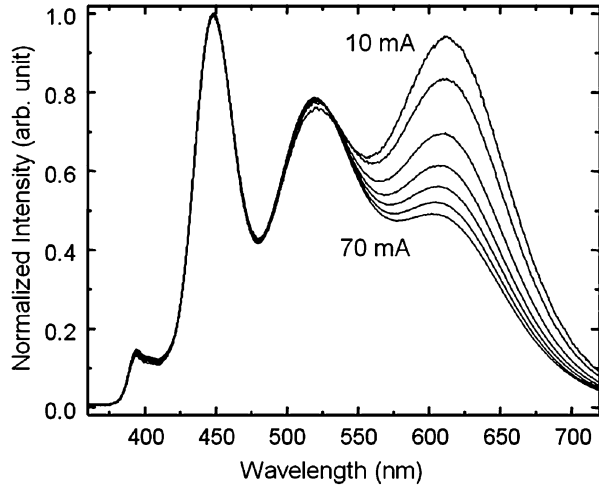
Fig. 16 PL intensities of (A) uncoated CaS:Eu^{2+} material and (B) CaS:Eu^{2+} material coated with Al_2O_3 against accelerated aging at 80 % relative humidity and 80 °C (Smet et al. 2011; Avci et al. 2011)

- (E) *Environmental stability*: The lifetime of commercial WLED products is currently 50,000 hours. This period presents a challenge to the packaging of LED devices and the durability of phosphors and driver circuits. Moreover, the color and intensity of phosphors should be maximally stable. However, phosphors, such as silicate $(\text{Sr,Ba})_2\text{SiO}_4:\text{Eu}$ or sulfide $(\text{Ca,Sr})\text{S:Eu}$ phosphors, are unsuitable for use in a humid environment, in which the emission of phosphors will lead to irreversible degradation. Therefore, a technique that coats a protective layer on the surface of phosphors is developed, as shown in Fig. 16 (Smet et al. 2011; Avci et al. 2011).
- (F) *Saturation property*: The saturation property of other dopant ions should not be neglected in WLEDs. At high-power LED excitation, a high flux leads to a saturation phenomenon in the excited states of the dopant ions with a long decay time (~ 10 ms), which is also the case for Mn^{2+} ion. If the flux of the excitation source is high, the electrons from the ground state will be depleted. Thus, emission intensity decreases, as shown in Fig. 17 (Smet et al. 2011; Kim et al. 2009). In addition, no saturation can be observed in Eu^{2+} and Ce^{3+} dopant ions because of a short decay time (10–1,000 ns).

Classification of Phosphors for pc-WLEDs

The 1-pc-WLEDs (one yellow phosphor embedded into blue LEDs that emit light at 420–480 nm) are commonly used in various applications, such as illumination and signage, because of their relatively simple fabrication process and low price.

Fig. 17 Normalized emission spectra of a WLED, which is composed of UV LED, blue and green Eu^{2+} phosphors, and red $\text{Eu}^{2+}\text{-Mn}^{2+}$ phosphor, as a function of the driving current (Smet et al. 2011; Kim et al. 2009)



However, the 3-pc-WLEDs (three kinds of phosphor embedded into UV LEDs that emit light at 360–410 nm) have gained more considerable attention than 1-pc-WLEDs because the former offers great flexibility in terms of fabrication design, higher CRI, and tunable CCT properties. According to the patent of Nichia (US 5,998,925), a light-emitting device comprises a light-emitting component ($\text{In}_i\text{Ga}_j\text{Al}_k\text{N}$ where $0 \leq i$, $0 \leq j$, $0 \leq k$ and $i + j + k = 1$) and a phosphor (garnet fluorescent material). The components of garnet phosphor are at least one element selected from the group of Y, Lu, Se, La, Gd, and Sm; at least another element selected from the group of Al, Ga, and In; and an activator with cerium (Shimizu et al. 1999). The patent infringement issue of 3-pc-LED development is not strictly limited compared with that of 1-pc-WLEDs, which easily violates the innovation of Nichia company. To date, phosphors, such as $\text{Y}_3\text{Al}_5\text{O}_{12}:\text{Ce}$ (yellow), $\text{Lu}_3\text{Al}_5\text{O}_{12}:\text{Ce}$ (greenish-yellow), $(\text{Sr},\text{Ba})_2\text{SiO}_4:\text{Eu}$ (yellowish-green), $\beta\text{-SiAlON}:\text{Eu}$ (green), $\text{Sr}_2\text{Si}_5\text{N}_8:\text{Eu}$ (orange-red), and $\text{CaAlSiN}_3:\text{Eu}$ (red), are commercialized and employed in either 1-pc-WLEDs or 3-pc-LEDs for illumination or for backlighting LCDs. Tables 4 and 5 summarize and classify popular phosphors for blue LEDs and UV LEDs, respectively. We have measured and compared these data ourselves. The other details of these phosphors are described in Xie et al. (2007), Li et al. (2007, 2008a, b), Kim et al. (2005a, b), Lei et al. (2011), Hansel et al. (2009), Arai et al. (2005), Nagai (2011), Park et al. (2009), Zorenko et al. (2009), Seto et al. (2009), Najafov et al. (2002), Piao et al. (2007), Van Haecke et al. (2007), Jang et al. (2009), Hana et al. (2014), Yadav et al. (2010), Tang et al. (2011), Peng et al. (2004), Zhou et al. (2010), Do et al. (2000), Yap et al. (2009), and Hao et al. (2011).

The five important properties of LED-used phosphors, namely, intensity, durability, thermal stability, emission bandwidth, and excitation bandwidth, are shown in Tables 4 and 5. From these tables, the individual superior properties of the phosphors can be easily determined. We have carefully described these parameters, except for excitation and emission bandwidths, in section “Requirements for Phosphor-Converted LEDs

Table 6 Key parameters for pc-WLEDs (Smet et al. 2011; Krames et al. 2007; Hu et al. 2005; Horikawa et al. 2009; Mueller-Mach et al. 2002, 2005; Wu et al. 2005; Shimomura et al. 2007; Yang et al. 2007; Setlur et al. 2010; Kimura et al. 2007; Huang and Chen 2010)

LED wavelength (nm)	Composition	CCT (K)	CRI
460	Y ₃ Al ₅ O ₁₂ :Ce	5600	71
460	Y ₃ Al ₅ O ₁₂ :Ce, CaS:Eu	5500	92
460	Y ₃ Al ₅ O ₁₂ :Ce, Sr ₂ Si ₅ N ₈ :Eu	2900	80
460	Sr ₂ GaS ₄ :Eu, SrS:Eu	3600	82
460	Sr ₂ GaS ₄ :Eu, (Ca,Sr)S:Eu	4800	92
450	Ca ₃ Sc ₂ Si ₃ O ₁₂ :Ce, CaAlSiN ₃ :Eu	6500	92
450	SrSi ₂ O ₂ N ₂ :Eu, Sr ₂ Si ₅ N ₈ :Eu	3200	89
455	SrSi ₂ O ₂ N ₂ :Eu, CaSiN ₂ :Ce	5200	91
450	(Sr,Ca) ₃ (Al,Si)O ₄ (O,F):Ce, K ₂ TiF ₆ :Mn	3200	90
455	BaSi ₂ O ₂ N ₂ :Eu, β-SiAlON:Eu, Ca-α-SiAlON:Eu, CaAlSiN ₃ :Eu	6400	96
455	BaSi ₂ O ₂ N ₂ :Eu, β-SiAlON:Eu, Ca-α-SiAlON:Eu, CaAlSiN ₃ :Eu	2900	98
365	BaMgAl ₁₀ O ₁₇ :Eu, Ca ₉ La(PO ₄) ₇ :Eu, Mn	4500	92

(pc-LEDs).” Many kinds of phosphor, such as nitride, oxynitride, or silicate compounds, are suitable for both UV LEDs and blue LEDs according to the excitation spectra of phosphors. Versatile high-QE systems, whose phosphors have suitable excitation spectra and accompanied by excited sources, have been investigated and are widely applied. Bandwidth is the full width at half maximum (FWHM) of the phosphor emission peak. FWHM is directly related to the color purity of phosphor emission. In general, a high CRI value of the broad emission is utilized in illumination. By contrast, the narrow emission band is suitable for LED TVs because the three fundamental colors require high color purity in such devices.

Based on the aforementioned characteristics of phosphors, the practical and economical quality control used in manufacturing LED devices is durability inspection, in which LED devices have to work at a high temperature and a humid environment for several hours. After which, LED devices can be released only if the emission properties satisfy the criteria without specifications. Table 6 summarizes the key parameters (phosphor composition, CCT, and CRI) for the selected 1-pc-LEDs and 3-pc-LEDs (Smet et al. 2011; Krames et al. 2007; Hu et al. 2005; Horikawa et al. 2009; Mueller-Mach et al. 2002; 2005; Wu et al. 2005; Shimomura et al. 2007; Yang et al. 2007; Setlur et al. 2010; Kimura et al. 2007; Huang and Chen 2010).

Applications of Phosphors

Phosphors generally exhibit luminescence phenomenon, which can be applied in various areas, such as illumination, backlighting, signals, and sensors. The applications

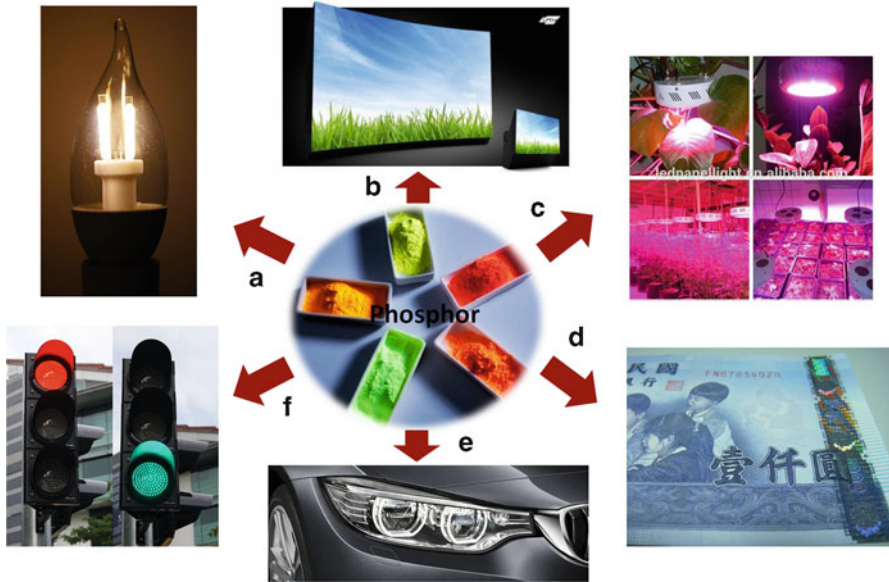


Fig. 18 Various applications of phosphors

of phosphor are shown in Fig. 18. Mercury-free illumination is produced by phosphor layers in fluorescent lamps, as shown in Fig. 18a. Such fluorescent lamps have attracted considerable attention because of their energy-saving, environment-friendly, and long-lifetime characteristics. Phosphorescent materials are used in radar screens and glow-in-the-dark toys, whereas phosphors are normally used in CRTs (Fig. 18b). In horticulture, phosphor-based grow lights are utilized in plant propagation, indoor gardening, and food production. Blue LEDs are suitable for vegetative growth because the light emitted has a wavelength in the mid-400 nm range (Fig. 18c). Most red LEDs (600–640 nm) are preferred for growing fruits or flowers. Phosphors are also used in security applications, in which items such as money (Fig. 18d) or concert tickets are invisibly marked. The authenticity of such materials is ensured by fluorescent marks under UV or blue light. In some concerts or nightclubs, the wrists of attendees are stamped with a fluorescent mark to allow them to leave and return without paying another admission fee. A challenge for phosphors, however, is high environmental temperature, which decreases QE and changes color. Car headlights are strictly required optical materials (Fig. 18e). Although the use of LED headlights has been accepted widely, the hope to reduce product cost remains. Phosphors are also used in traffic signals (Fig. 18f). They have wide-ranging applications, which make them ubiquitous.

Prospects of Phosphors

Developments in phosphor use and LEDs have continued for several decades. Efficiency, technology, design, and cost have distinctly improved according to human requirements. However, the CCT (4,000–8,000 K) and CRI (CRI < 75) of most WLED devices, which are composed of blue LED chips and yellow-emitting YAG:Ce³⁺ phosphors, remain limited by such combination because of the lack of red emission in the luminescent spectra (Nguyen et al. 2014). Original red phosphors, particularly silicon-free nitridoaluminates, group (III) nitrides, and fluorides, are being developed recently. Moreover, the thermal issue of luminescent materials is considerably ameliorated through novel designs in packaging technology, such as remote phosphor.

Red-Emitting Phosphor Materials

Pust et al. (2014) investigated new red-emitting nitridoaluminate phosphor Sr[LiAl₃N₄]:Eu²⁺ (SLA) with superior luminescent properties; it exhibited a narrow-band emission spectrum and a suitable excitation spectrum. A new high-performance GaN-based blue LED emerged from this phosphor (Fig. 19). The structural parameters of SLA are refined from single-crystal and powder X-ray diffraction (XRD). Its space group belongs to triclinic *P*-1 (no. 2) with $a = 5.86631(12)$ Å, $b = 7.51099(15)$ Å, $c = 9.96545(17)$ Å, $\alpha = 83.6028(12)^\circ$, $\beta = 76.7720(13)^\circ$, and $\gamma = 79.5650(14)^\circ$; it is isotypic to the oxoplumbate Cs[Na₃PbO₄] (Stoll and Hoppe 1987).

The crystal structure of the SLA host lattice has channels of vierer rings along [011], as shown in Fig. 19a. Its highly condensed, rigid framework is composed of face-sharing, cuboid-like SrN₈ polyhedra coordinated by AlN₄ and LiN₄ tetrahedra (Fig. 19b).

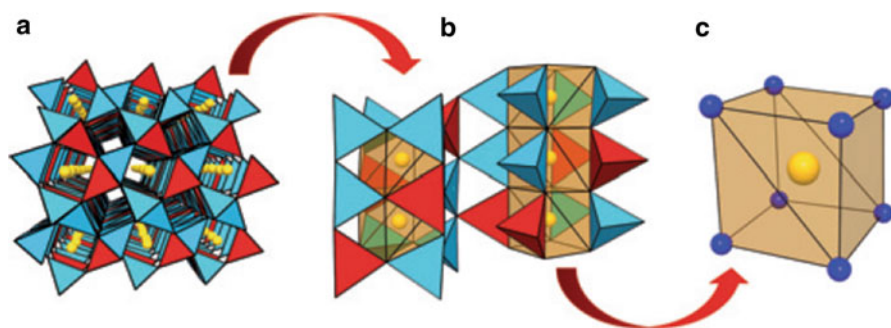


Fig. 19 Crystal structure of Sr[LiAl₃N₄]:Eu²⁺ red-emitting phosphor. (a) Perspective view with channels of vierer rings along [011]. (b) Strands of face-sharing, cuboid-like SrN₈ polyhedra coordinated by AlN₄ and LiN₄ tetrahedra. (c) Cuboid-like SrN₈ polyhedron. The yellow spheres are Sr, the blue spheres are N, the blue tetrahedra are AlN₄, and the red polyhedra are LiN₄ (Pust et al. 2014)

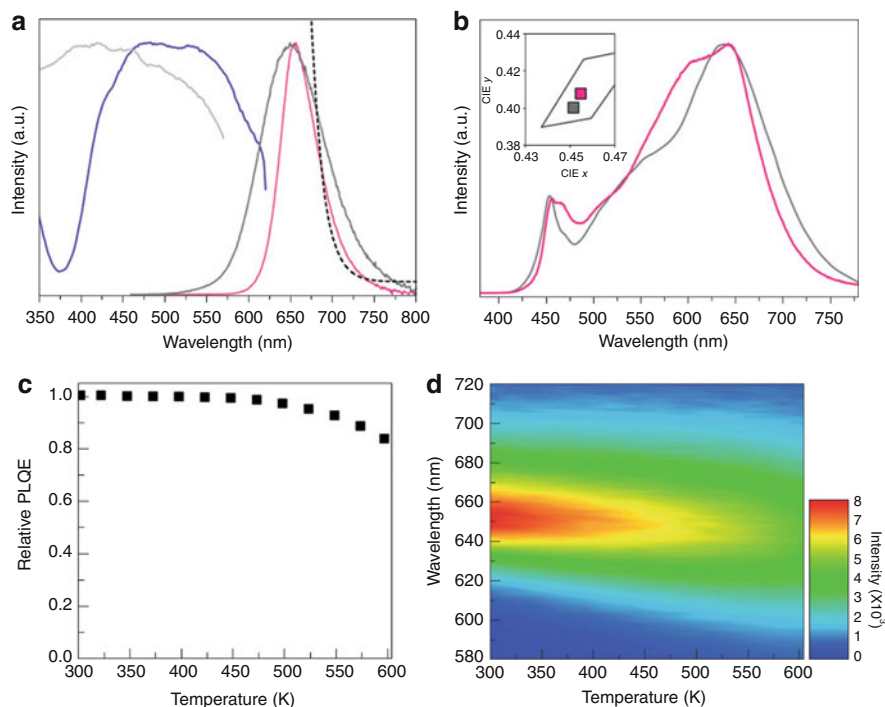


Fig. 20 PL properties of Sr[LiAl₃N₄]:Eu²⁺ red-emitting phosphor. **(a)** Excitation (SLA, blue; CaAlSiN₃:Eu²⁺, light gray) and emission spectra of SLA (pink) and of CaAlSiN₃:Eu²⁺ (dark gray). **(b)** Luminescent spectra of pc-LED with a CCT of 2,700 K. Pink curve: LED + LuAG + (Ba,Sr)₂Si₅N₈:Eu²⁺ and SLA mix. Gray curve: commercially available high-CRI LED. **(c)** Relative photoluminescence quantum efficiency (PLQE) of SLA at a function of temperature. **(d)** Temperature dependence of the spectral emission of SLA (Pust et al. 2014)

To charge balance, the net negative charge of the [LiAl₃N₄]²⁻ framework is compensated by incorporating Sr²⁺ ions in every second channel. The narrow emission band is ascribed to two Sr positions; each of which is coordinated by eight N atoms in a highly symmetric cuboid-like environment, as shown in Fig. 19c.

The broad excitation (blue) and narrow emission (pink) spectra of SLA:0.4%Eu are shown in Fig. 20a. The wavelength of the maximum emission is at $\lambda_{\text{em}} = 654 \text{ nm}$ with $\text{FWHM} = 1,180 \text{ cm}^{-1}$ when excitation is at $\lambda_{\text{ex}} = 440 \text{ nm}$. Such a particular narrow red emission is relative to the local environment of the activators, as shown in Fig. 19. In addition, a small part of the emission crosses the sensitivity of the human eye (dotted curve), which results in a slightly limited luminous efficiency of WLEDs. By contrast, the efficiency of WLEDs with commercial CaAlSiN₃:Eu²⁺ (dark gray) is significantly restricted because of its large emission spectrum over 700 nm. In Fig. 20b, the luminescent spectrum of a commercial high-CRI WLED (gray curve) is compared with the spectrum of a blue LED with LuAG, (Ba,Sr)₂Si₅N₈:Eu²⁺, and SLA (pink curve) at a correlated color temperature of 2,700 K (see inset). In

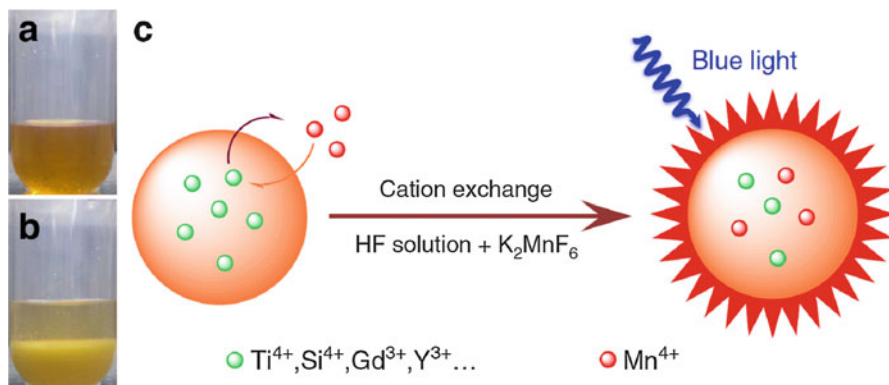


Fig. 21 Photographs of the HF solution (a) dissolved with K_2MnF_6 crystals and (b) the same solution containing K_2TiF_6 powder after cation exchange reaction for 3 min. (c) Schematic of the cation exchange procedure to synthesize Mn^{4+} -activated fluoride compounds (Zhu et al. 2014)

the results, the WLED with the red-emitting phosphor (SLA) exhibits a 14 % increase in luminous efficacy and excellent color rendering ($R_a8 = 91$, $R_9 = 57$). The thermal stability of SLA phosphor is shown in Fig. 20c, d. The integrated light output decreases by 5 %, which is comparable with the performance of highly efficient $\text{YAG}:\text{Ce}^{3+}$ phosphors. In chromaticity, only a small blue shift from $(x, y) = (0.692, 0.306)$ to $(x, y) = (0.668, 0.330)$ is observed at a function of temperature (from 303 to 465 K). Hence, the narrow-band red-emitting phosphor SLA is an excellent candidate for WLED devices.

Recently, Mn^{4+} -doped fluoride compounds have attracted considerable attention because they exhibit the most intense broadband excitation at 460 nm with a bandwidth of 50 nm and extremely sharp emission lines peaking at 630 nm as a result of a weak crystal field. Zhu et al. (2014) proposed a unique cation exchange approach for synthesizing micrometer-sized $\text{K}_2\text{TiF}_6:\text{Mn}^{4+}$ phosphor with an unusually fast reaction rate. The K_2TiF_6 powder was immersed in an HF solution that contained 5.50 % Mn^{4+} ions. After cation exchange reaction at room temperature for 3 min under stirring, the powder was isolated instantly through vacuum filtration. The color of the mixture changed from brown to light yellow, which indicated a remarkable decrease in the Mn^{4+} concentration of the solution (Fig. 21a, b). Accordingly, the color of the powder changed from white to yellow after cation exchange. Figure 21c illustrates the cation exchange procedure to prepare Mn^{4+} -activated fluoride compounds.

To evaluate device performance with the synthesized $\text{K}_2\text{TiF}_6:\text{Mn}^{4+}$ phosphor, WLEDs with various CCTs are fabricated by combining blue chips (455 nm), $\text{YAG}:\text{Ce}^{3+}$ yellow, and $\text{K}_2\text{TiF}_6:\text{Mn}^{4+}$ (5.50 at.%) red phosphors. The chromaticity coordinates of the three typical LEDs with CCTs of 2,783, 3,556, and 5,954 K under a drive current of 60 mA, (0.4569, 0.4158), (0.3997, 0.3821), and (0.3224, 0.3416), respectively, are marked in CIE 1931 color spaces; all three color points are laid on

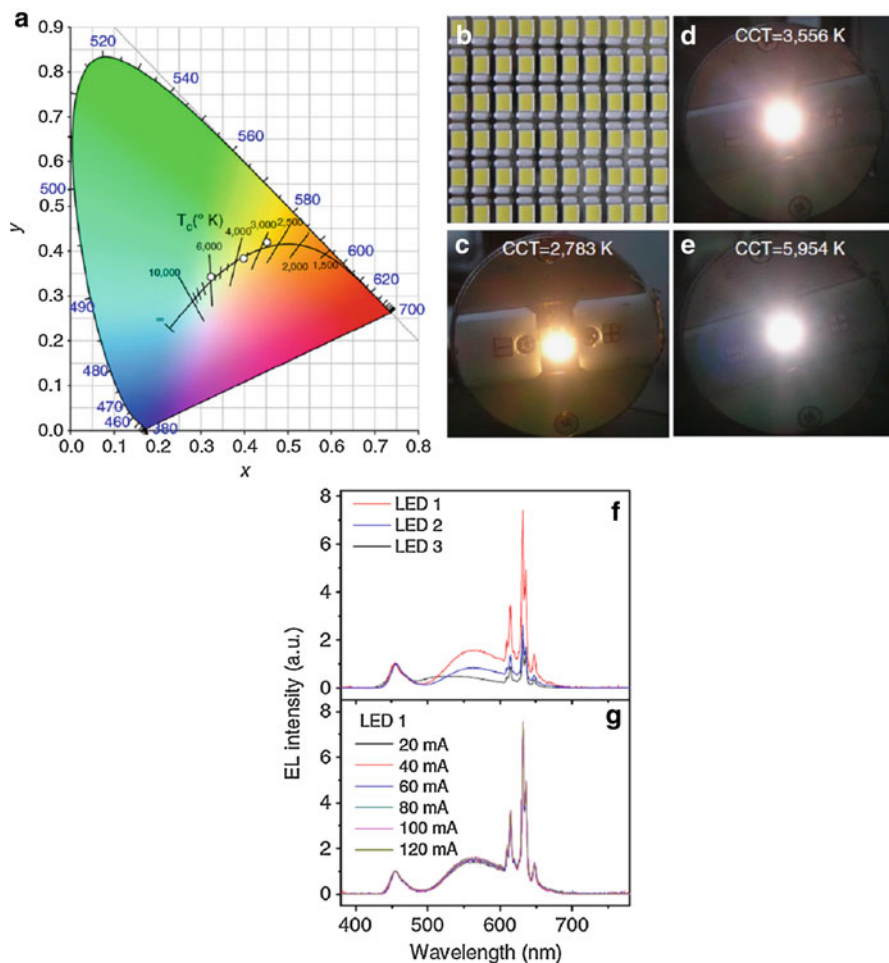


Fig. 22 (a) Chromaticity coordinates of three typical LEDs with CCTs of 2,783, 3,556, and 5,954 K under a drive current of 60 mA in CIE 1931 color spaces. Photographs of the (b) as-fabricated LEDs and lighted ones with CCTs of (c) 2,783, (d) 3,556, and (e) 5,954 K. The EL spectra of (f) the three LEDs under a drive current of 60 mA and (g) LED 1 under various drive currents. All spectra are normalized at the peak intensity at 455 nm (Zhu et al. 2014)

or close to the black-body locus (Fig. 22a). Photographs of the as-fabricated LEDs and the three lighted ones are provided in Fig. 22b–e. The electroluminescent (EL) spectra of the three LEDs reconfirm the sharp emission lines of Mn^{4+} in K_2TiF_6 phosphor and other red-emitting components with lower CCT (Fig. 22f). A high luminous efficacy of 116 lm/W is achieved under a drive current of 60 mA for the WLED with a CCT of 3,556 K. In particular, R_9 values are positive for all three LEDs, which suggests a good rendering of a strong red color. The LEDs also exhibit excellent chromaticity coordinate stability with the drive current varying from 20 to

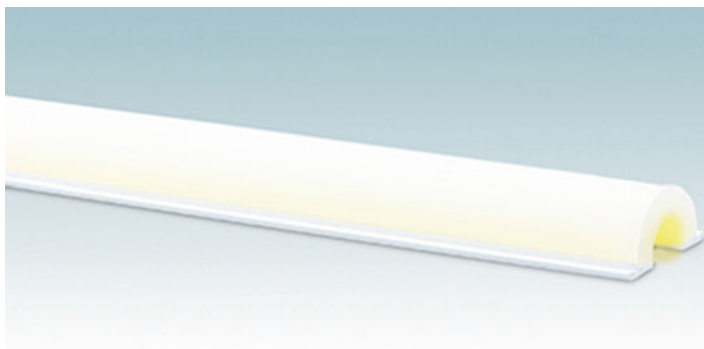


Fig. 23 Remote-phosphor product (Carey 2014)

120 mA, as shown in Fig. 22g. These findings indicate the significant potential of $\text{K}_2\text{TiF}_6:\text{Mn}^{4+}$ as a commercial red phosphor for warm WLEDs and open up new avenues for exploring novel non-RE red-emitting phosphors.

Packaging Technology for WLEDs

According to a report from McKinsey & Company (Carey 2014), the growth of LED-based lighting in North America is estimated to average 38 % between 2012 and 2016. The quality of light, packaging technology, and cost have become predominant requirements, which have, in turn, led to a demand for advanced phosphor solutions and a considerably complex mixing strategy. Intematix Corporation developed a remote-phosphor technology, which separated the phosphor materials from the LED plate (Fig. 23). High-quality white light is generated by illuminating blue LED behind the remote-phosphor plate. Individual-level thermal properties are also important; these properties can be improved and rapidly utilized for various applications that require uniform illumination and high efficiency.

Wang et al. (2014) reported a novel non-RE-based oxide red phosphor, i.e., $\text{CaMg}_2\text{Al}_{16}\text{O}_{27}:\text{Mn}^{4+}$ (CMA: Mn^{4+}), incorporated with YAG: Ce^{3+} phosphor microcrystals into the glass host via a “phosphor-in-glass” (PiG) approach. Figure 24 illustrates the PL excitation (PLE) and PL spectra of CMA: Mn^{4+} compound. The PLE spectrum shows several excitation bands (${}^4A_{2g} \rightarrow {}^4T_{2g}$; ${}^4A_{2g} \rightarrow {}^2T_{2g}$; ${}^4A_{2g} \rightarrow {}^4T_{1g}$; $\text{Mn}^{4+}\text{-O}^{2-}$ charge transfer) ranging from 250 to 550 nm. This red-emitting band is composed of four distinguishable Stokes/anti-Stokes sidebands at 642, 655, 665, and 671 nm, which are ascribed to the different vibrational modes of 2E_g , ${}^2T_{2g} \rightarrow {}^4A_{2g}$ transitions for the $3d^3$ electrons in the $[\text{MnO}_6]^{8-}$ octahedral complex. Moreover, the photon reabsorption effect can be considerably reduced because the spectral overlap between the excitation of CMA: Mn^{4+} and the emission of commercial YAG: Ce^{3+} is relatively small compared with the spectral overlap in the combination of YAG: Ce^{3+} and nitride phosphors.

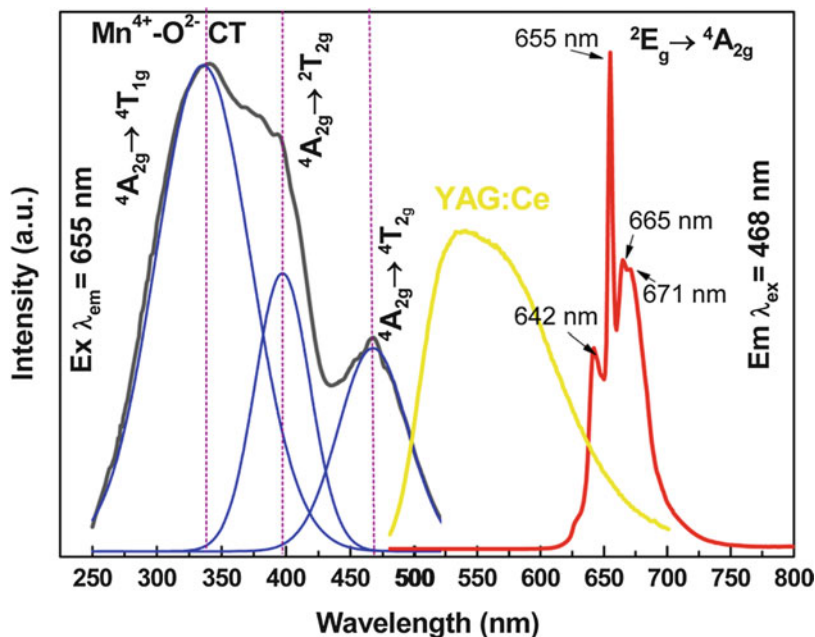


Fig. 24 Room temperature PLE ($\lambda_{em} = 655$ nm) and PL ($\lambda_{ex} = 468$ nm) spectra of the CMA:0.08Mn⁴⁺ sample. The solid blue line represents the data fit using a Gaussian function. The vertical dashed lines are a guide for the eye. The yellow line represents the referenced emission spectrum of commercial YAG:Ce³⁺ phosphor under a light excitation of 468 nm (Wang et al. 2014)

Then, a two-phosphor PiG color converter is fabricated by incorporating y wt% CMA:0.08Mn⁴⁺ ($y = 0, 1, 3, 5, 7, 9$) and 5 wt% commercial YAG:Ce³⁺ phosphors into the glass matrix, as shown in Fig. 25. The internal QE of the device (5 wt% CMA:0.08Mn⁴⁺ and 5 wt% YAG:Ce³⁺) is measured as 76.8 %. Upon the appearance of a white light with various CCTs (cool to warm), the CIE coordinates change from point a (0.312, 0.333) to point f (0.395, 0.416), CCT decreases monotonously from 6674 to 3896 K, and CRI increases from 70.0 to 85.5. These results indicate that CMA:Mn⁴⁺ red phosphor can improve the CCT and CRI values of high-powered WLEDs for indoor applications. However, the luminous efficiency of the samples declines from 124.6 to 58.3 lm/W as the content of CMA:Mn⁴⁺ phosphor increases. The balance between QE and CRI should be controlled depending on different requirements.

Furthermore, quantum dots (QDs) are extensively applied to color converters of WLEDs because of their quantum confinement effect, narrow FWHM, and pure color. Sohn et al. (2014) synthesized CuInS₂/ZnS QDs by a heating method and embedded QDs in silica (QDES) via the Stöber–Fink–Bohn method (Kim et al. 2011). The emission spectra of QDs are shown in Fig. 26a, and peaks emerge from 550 to 600 nm under an excitation of 450 nm. The quantum yield of the synthesized QDs in solution is approximately 60 %, with FWHM = 102 nm.

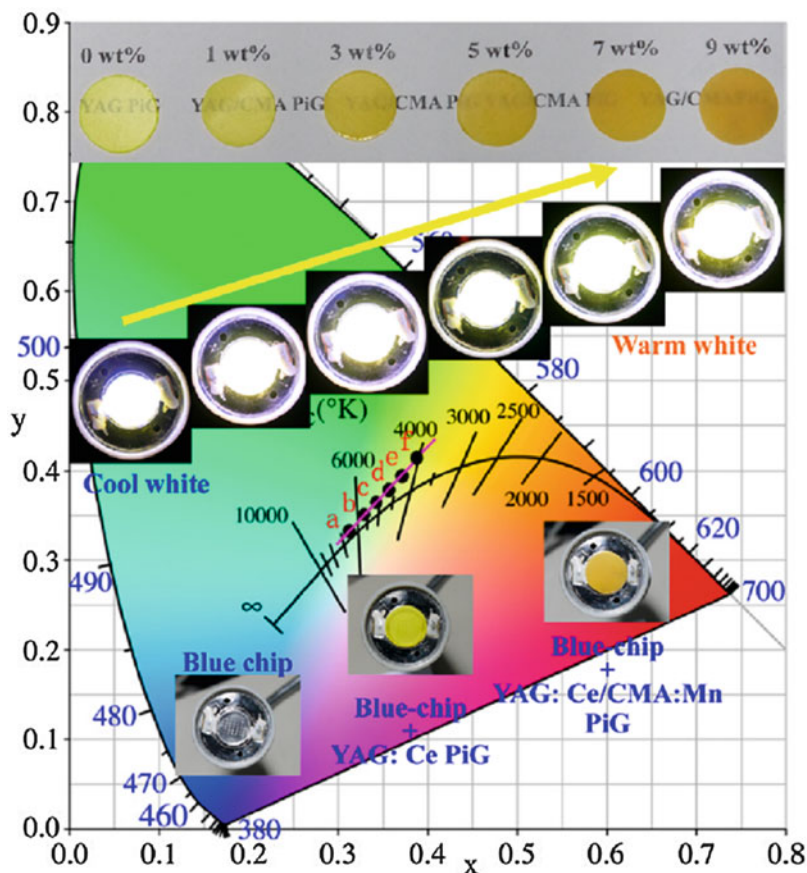


Fig. 25 CIE chromaticity diagram of WLEDs fabricated by coupling 5 wt% YAG:Ce³⁺ and y ($y = 0, 1, 3, 5, 7, 9$) wt% CMA:Mn⁴⁺ in glass (PiG) with blue chips. The insets show photographs of the PiG samples with various CMA:Mn⁴⁺ weighted contents, the corresponding LED packages, and their EL driven by a current of 350 mA (Wang et al. 2014)

Under a UV lamp, the photographs of the QD solutions with different colors are presented in Fig. 26b. Approximately 3 nm-sized QDs are obtained via transmission electron microscopy (TEM), as shown in the inset of Fig. 26d. By utilizing ligand exchange, the hydroxyl-functionalized QD ligands ensure effective trapping of QDs in the silica matrix, as shown in Fig. 26c, d. To demonstrate the reactions between CuInS₂/ZnS QDs and silica, QDES is measured by energy-dispersive X-ray spectroscopy (EDS), and the results are shown in Fig. 26d. The data indicate that all the elements exist at expected ratios and that a chemical reaction between CuInS₂/ZnS QDs and silica has occurred.

In general, phosphors are dispersed in silicon resin, and the matrix is encapsulated into a LED chip-based device. In this configuration, device efficiency is reduced

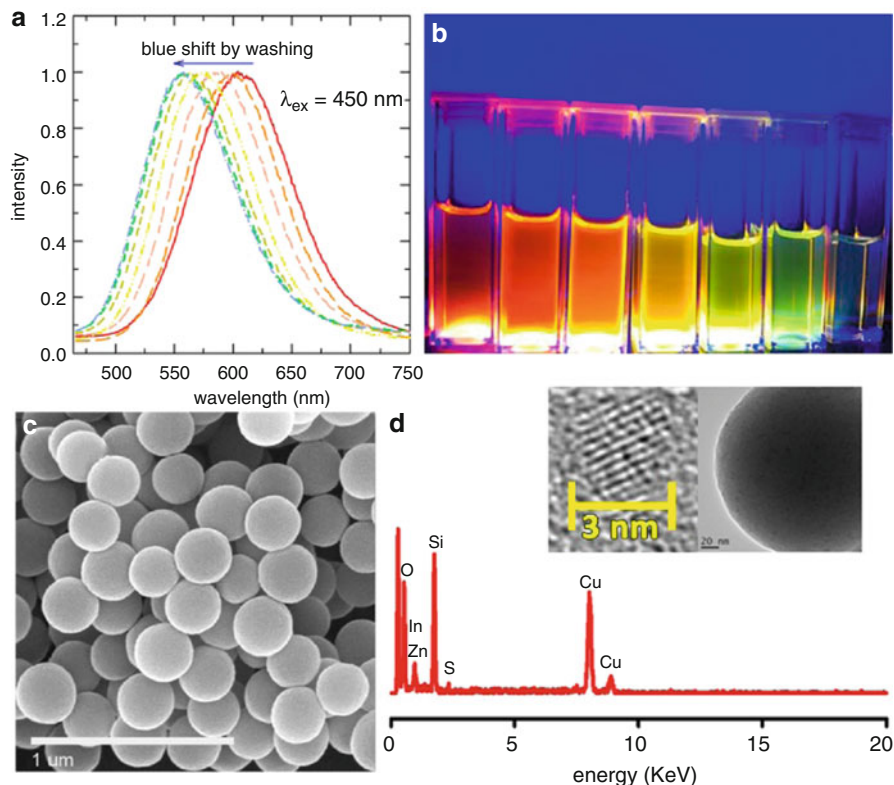


Fig. 26 (a) PL emission spectra of synthesized QDs. (b) QDs under UV excitation. (c) Scanning electron microscopy image of QDES. (d) EDS spectra of QDES. The *inset* of panel (d) shows the lattice fringes obtained by TEM (Sohn et al. 2014)

because the phosphor in silicon resin directly endures thermal stress. Given that the remote-phosphor configuration (Fig. 27c) maintains an appreciable distance between chips (UV/blue LED) and color converters (luminescent materials), it can resolve the thermal stress issue, as well as provide smooth chromaticity variation and sufficient diffusing effect. Figure 27a shows a picture of YAG:Ce³⁺ phosphor dispersed into the glass host via the PiG approach. The QDES is formed as a film by dispersing it in ethoxylated trimethylolpropane triacrylate (ETPTA) to improve the stability of optical materials. The weight ratio is 8:2 (QDES:ETPTA), and the thickness of a Q-ETPTA film is 0.4 mm. Figure 27d shows a Q-ETPTA film under UV illumination. Therefore, a novel WLED package is produced by placing a Q-ETPTA film on top of a robust phosphor plate, as shown in Fig. 27b. A WLED in operation is presented in Fig. 27e. In this configuration, the PiG plate, which has low thermal conductivity, maintains a high temperature gradient, blocks the thermal stress on the Q-ETPTA film from the LED chip, and maintains a suitable distance between the LED chip and the Q-ETPTA film. With the PiG plate as a conservator, the Q-ETPTA

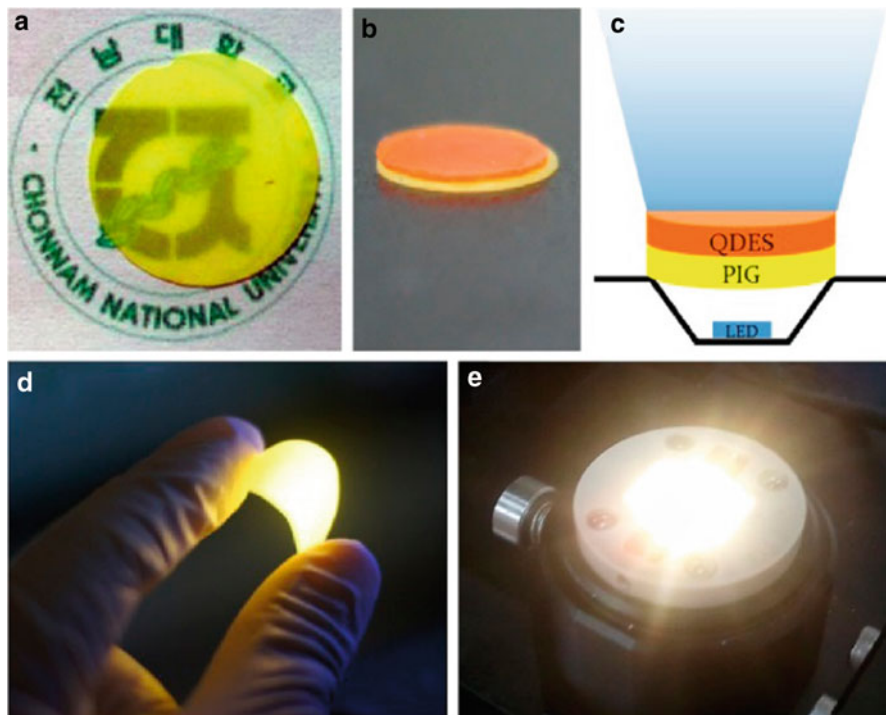


Fig. 27 (a) Photograph of the PiG plate. (b) Q-ETPTA film placed on top of the PiG plate. (c) Schematic of the novel WLED package. (d) Photograph of a Q-ETPTA film under UV illumination. (e) WLED in operation (Sohn et al. 2014)

film is protected from the usual photo-induced damages on packaging materials caused by unconverted high-energy radiations that are traversing through it. The superior spectral and mechanical stabilities of a WLED system can be achieved by considering all the aforementioned factors.

Conclusion

In summary, inorganic material-based WLED devices have recently played significant roles in modern life because of their high-efficiency, low-pollution, energy-saving, and environment-friendly features. Phosphor is an important component of WLEDs. In this study, we first describe the history and luminescent phenomena of optical materials. Next, we introduce the fundamental concepts of phosphors and the requirements for phosphor-converted LEDs. In addition, popular UV-excited or blue-excited phosphors are analyzed. Then, the prospects of phosphor are discussed by presenting novel phosphors and new designs for WLED packages. Although producing WLED devices with high luminescent properties and excellent CRI

presents a considerable challenge, its prospective applications in the field of phosphor-based LED are significant.

Acknowledgement This work was supported by the Ministry of Science and Technology, Taiwan (Contract Nos: MOST 104-2113-M-002-012-MY3, 104-2119-M-002-027-MY3 and 104-2923-M-002-007-MY3).

References

- Arai Y, Kominami H, Nakanishi Y, Hatanaka Y (2005) Luminescent properties of SrGa₂S₄:Eu thin film phosphors deposited by two electron beam evaporation. *Appl Surf Sci* 244:473
- Avci N, Cimieri I, Smet PF, Poelman D (2011) Stability improvement of moisture sensitive CaS:Eu²⁺ micro-particles by coating with sol-gel alumina. *Opt Mater* 33:1302
- Blasse G (1972) The ultraviolet absorption bands of Bi³⁺ and Eu³⁺ in oxides. *J Solid State Chem* 4:52
- Blasse G, Grabmaier BC (1994) *Luminescent materials*. Springer, Berlin Heidelberg
- Boudeile J, Didierjean J, Camy P, Doualan JL, Benayad A, Ménard V, Moncorgé R, Druon F, Balembois F, Georges P (2008) Thermal behavior of ytterbium-doped fluorite crystals under high power pumping. *Opt Express* 16:10098
- Caravaggio was early 'photographer' BBC News. 11 March 2009
- Carey J (2014) LED phosphor IP trends and licensing. Senior Director of Marketing, Intematix Corporation, <http://www.ecnmag.com/article/2014/05/led-phosphor-ip-trends-and-licensing>
- Chandross EA (1963) A new chemiluminescent system. *Tetrahedron Lett* 4(12):761
- Clark DE (2000) Peroxides and peroxide forming compounds. Chemical and Biological Safety Officer Texas A&M University
- Clegg W, Bourhill G, Sage I (2002) Hexakis(antipyrine-O)terbium(III) triiodide at 160 K: confirmation of a centrosymmetric structure for a brilliantly triboluminescent complex. *Acta Cryst* E58:m159
- Condon E (1926) A theory of intensity distribution in band systems. *Phys Rev* 28:1182
- Cotton FA (1990) *Chemical applications of group theory*, 3rd edn. Wiley, Chichester
- Curie D (1963) *Luminescence in crystals*. Methuen & Co., London, pp 31–68
- Di Bartolo B (1968) *Optical interactions in solids*. Wiley, New York, pp 420–427
- Dieke GH (1968) *Spectra and energy levels of rare earth ions in crystals*. Interscience, New York
- Do YR, Bae JW, Kim Y, Yang HG (2000) Preparation and optical properties of SrGa₂S₄:Eu phosphor. *Bull Korean Chem Soc* 21:295
- Dorenbos P (2000a) Predictability of 5d level positions of the triply ionized lanthanides in halogenides and chalcogenides. *J Lumin* 87–89:970
- Dorenbos P (2000b) The 4fⁿ to 4fⁿ⁻¹5d transitions of the trivalent lanthanides in halogenides and chalcogenides. *J Lumin* 91:91
- Dorenbos P (2000c) The 5d level positions of the trivalent lanthanides in inorganic compounds. *J Lumin* 91:155
- Dorenbos P (2003) Energy of the first 4f⁷-4f⁶5d transition of Eu²⁺ in inorganic compounds. *J Lumin* 104:239
- Dorenbos P (2005) Thermal quenching of Eu²⁺ 5d-4f luminescence in inorganic compounds. *J Phys Condens Matter* 17:8103
- Duffy JA (1990) *Bonding, energy levels and bands in inorganic solids*. Longman Scientific and Technical, Harlow
- Figgis BN, Hitchman MA (2000) *Ligand field theory and its applications*. Wiley-VCH, New York
- Hammond CR (2000) *The elements*. In: *Handbook of chemistry and physics*, 81st edn. CRC Press. ISBN 0-8493-0481-4

- Hana JK, Piquette A, Hannah ME, Hirata GA, Talbot JB, Mishra KC, McKittrick J (2014) Analysis of (Ba, Ca, Sr)₃MgSi₂O₈:Eu²⁺, Mn²⁺ phosphors for application in solid state lighting. *J Lumin* 148:1
- Hansel RA, Allison SW, Walker DG (2009) Temperature-dependent luminescence of Ce³⁺ in gallium-substituted garnets. *Appl Phys Lett* 95:114102
- Hao ZD, Zhang JH, Zhang X, Wang XJ (2011) CaSc₂O₄:Eu³⁺: a tunable full-color emitting phosphor for white light emitting diodes. *Opt Mater* 33:355
- Hirosaki N, Ogata S, Kocer C (2003) Ab initio calculation of the crystal structure of the lanthanide Ln₂O₃ sesquioxides. *J Alloys Comp* 351:31
- Horikawa T, Piao XQ, Fujitani M, Hanzawa H, Machida K (2009) Preparation of Sr₂Si₃N₈:Eu²⁺ phosphors using various novel reducing agents and their luminescent properties". *IOP Conf Ser Mater Sci Eng* 1:012024
- HowStuffWorks (2001) How do fireflies light up? [Science.howstuffworks.com](http://science.howstuffworks.com). Retrieved 22 June 2013
- Hu Y, Zhuang W, Ye H, Zhang S, Fang Y, Huang X (2005) Preparation and luminescent properties of (Ca_{1-x}, Sr_x)S:Eu²⁺ red-emitting phosphor for white LED. *J Lumin* 111:139
- Huang CH, Chen TM (2010) Ca₉La(PO₄)₇:Eu²⁺, Mn²⁺: an emission-tunable phosphor through efficient energy transfer for white light-emitting diodes. *Opt Express* 18:5089
- Jang HS, Won YH, Vaidyanathan S, Kim DH, Jeon DY (2009) Emission band change of (Sr_{1-x}M_x)₃SiO₅:Eu²⁺ (M = Ca, Ba) phosphor for white light sources using blue/near-ultraviolet LEDs. *J Electrochem Soc* 156:J138
- Jørgensen CK (1962a) Absorption spectra and chemical bonding in complexes. Pergamon, Oxford
- Jørgensen CK (1962b) Electron transfer spectra of lanthanide complexes. *Mol Phys* 5:271
- Jørgensen CK (1971) Modern aspects of ligand field theory. North-Holland, Amsterdam
- Jüstel T (2005) Luminescent materials for high brightness LEDs. FH Munster-Philips Research Aachen, https://www.fh-muenster.de/fb1/downloads/personal/juestel/juestel/Luminescent_Materials_for_High_Brightness_LEDs_September_2004_.pdf
- Kamimura A, Sugano S, Tanabe Y (1969) Ligand field theory and its applications, 1st edn. Shokabo, Tokyo, pp 269–321
- Kim JS, Park YH, Kim SM, Choi JC, Park HL (2005a) Temperature-dependent emission spectra of M₂SiO₄:Eu²⁺ (M = Ca, Sr, Ba) phosphors for green and greenish white LEDs. *Solid State Commun* 133:445
- Kim JS, Lim KT, Jeong YS, Jeon PE, Choi JC, Park HL (2005b) Full-color Ba₃MgSi₂O₈:Eu²⁺, Mn²⁺ phosphors for white-light-emitting diodes. *Solid State Commun* 135:21
- Kim TG, Kim YS, Im SJ (2009) Energy transfer and brightness saturation in (Sr, Ca)₂P₂O₇:Eu²⁺, Mn²⁺ phosphor for UV-LED lighting. *J Electrochem Soc* 156:J203
- Kim H, Jang HS, Kwon BH, Suh M, Youngsun K, Cheong SH, Jeon DY (2011) In situ synthesis of thiol-capped CuInS₂-ZnS quantum dots embedded in silica powder by sequential ligand-exchange and silanization. *Electrochem Solid-State Lett* 15:K16
- Kimura N, Sakuma K, Hirafune S, Asano K, Hirosaki N, Xie RJ (2007) Extrahigh color rendering white light-emitting diode lamps using oxynitride and nitride phosphors excited by blue light-emitting diode. *Appl Phys Lett* 90:051109
- Klick CC, Schulman JH (1997) Solid state physics. In: Seitz F, Turnbull D (eds), Academic, New York, vol 5, pp 97–116
- Krames MR, Shchekin OB, Mueller-Mach R, Mueller GO, Ling Z, Harbers G, Craford MG (2007) Status and future of high-power light-emitting diodes for solid-state lighting. *J Disp Technol* 3:160
- Lei BF, Machida KI, Horikawa T, Hanzawa H (2011) Preparation of (Sr_{0.5}Ba_{0.5})Si₂N₂O₂:Eu²⁺ phosphor and its luminescence properties. *Chem Lett* 40:140141
- Lever ABP (1984) Inorganic electronic spectroscopy, 2nd edn. Elsevier, Amsterdam
- Li HL, Liu XJ, Huang LP (2007) Luminescent properties of LuAG:Ce phosphors with different Ce contents prepared by a sol-gel combustion method. *Opt Mater* 29:1138
- Li HL, Xie RJ, Hirosaki N, Suehiro T, Yajima Y (2008a) Phase purity and luminescence properties of fine Ca-alpha-SiAlON:Eu phosphors synthesized by Gas reduction nitridation method. *J Electrochem Soc* 155:J175

- Li HL, Xie RJ, Hirosaki N, Yajima Y (2008b) Synthesis and photoluminescence properties of $\text{Sr}_2\text{Si}_5\text{N}_8:\text{Eu}^{2+}$ red phosphor by a gas-reduction and nitridation method. *J Electrochem Soc* 155: J378
- Li GG, Hou ZY, Peng C, Wang WX, Cheng ZY, Li CX, Lian HZ, Lin J (2010) Electrospinning derived one-dimensional $\text{LaOCl}:\text{Ln}^{3+}$ ($\text{Ln} = \text{Eu}/\text{Sm}, \text{Tb}, \text{Tm}$) nanofibers, nanotubes and microbelts with multicolor-tunable emission properties. *Adv Funct Mater* 20:3446
- Lin CC, Liu RS (2014) Thermal effects in (oxy)nitride phosphors. *J Solid State Light* 1:16
- Lyu LJ, Hamilton DS (1991) Radiative and nonradiative relaxation measurements in Ce^{3+} doped crystals. *J Lumin* 48–49:251
- Maeda K (1963) Luminescence. Maki Shoten, Japan, pp 6–10 and 37–48
- Meijerink A, Nuyten J, Blasse G (1989) Luminescence and energy migration in $(\text{Sr}, \text{Eu})\text{B}_4\text{O}_7$, a system with a $4f^7-4f^65d$ crossover in the excited state. *J Lumin* 44:19
- Mueller-Mach R, Mueller GO, Krames MR, Trottier T (2002) High-power phosphor-converted light-emitting diodes based on III-Nitrides. *IEEE J Sel Top Quantum Electron* 8:339
- Mueller-Mach R, Mueller G, Krames MR, Hoppe HA, Stadler F, Schnick W, Jüstel T, Schmidt P (2005) Highly efficient all-nitride phosphor-converted white light emitting diode. *Phys Status Solidi A* 202:1727
- Nagai H (2011) Light-emitting device. US Patent 08288790
- Najafov H, Kato A, Toyota H, Iwai K, Bayramov A, Iida S (2002) Effect of Ce co-doping on $\text{CaGa}_2\text{S}_4:\text{Eu}$ phosphor: I. Energy transfer from Ce to Eu ions. *Jpn J Appl Phys* 41:1424
- Nakazawa E (2006) Phosphor handbook: fundamentals of luminescence. CRC Press, Boca Raton, FL
- Newscientist.com Dial R for radioactive – 12 July 1997 – New Scientist. Retrieved 12 Sept 2008
- Nguyen HD, Lin CC, Fang MH, Liu RS (2014) Synthesis of $\text{Na}_2\text{SiF}_6:\text{Mn}^{4+}$ red phosphors for white LED applications by co-precipitation. *J Mater Chem C* 2:10268
- Park JY, Jung HC, Raju GSR, Moon BK, Jeong JH, Son SM, Kim JH (2009) Sintering temperature effect on structural and luminescence properties of 10 mol% Y substituted $\text{Gd}_3\text{Al}_5\text{O}_{12}:\text{Ce}$ phosphors. *Opt Mater* 32:293
- Peng TY, Liu HJ, Yang HP, Yan CH (2004) Synthesis of $\text{SrAl}_2\text{O}_4:\text{Eu}, \text{Dy}$ phosphor nanometer powders by sol–gel processes and its optical properties. *Mater Chem Phys* 85:68
- Piao XQ, Machida KI, Horikawa T, Hanzawa H (2007) Preparation of $\text{CaAlSiN}_3:\text{Eu}^{2+}$ phosphors by the self-propagating. *Chem Mater* 19:4592
- Pust P, Weiler V, Hecht C, Tücks A, Wochnik AS, Henß AK, Wiechert D, Scheu C, Schmidt PJ, Schnick W (2014) Narrow-band red-emitting $\text{Sr}[\text{LiAl}_3\text{N}_4]:\text{Eu}^{2+}$ as a next-generation LED-phosphor material. *Nat Mater* 13:891
- Rauhut MM (1969) Chemiluminescence from concerted peroxide decomposition reactions (science). *Acc Chem Res* 2(3):80
- Sakuma K, Hirosaki N, Xie RJ (2007) Red-shift of emission wavelength caused by reabsorption mechanism of europium activated $\text{Ca}-\alpha\text{-SiAlON}$ ceramic phosphors. *J Lumin* 126:843
- Setlur AA, Radkov EV, Henderson CS, Her JH, Srivastava AM, Karkada N, Kishore MS, Kumar NP, Aesram D, Deshpande A, Kolodin B, Grigorov LS, Happek U (2010) Energy-efficient, high-color-rendering LED lamps using oxyfluoride and fluoride phosphors. *Chem Mater* 22:4076
- Seto T, Kijima N, Hirosaki N (2009) A new yellow phosphor $\text{La}_3\text{Si}_6\text{N}_{11}:\text{Ce}^{3+}$ for white LEDs. *ESC Trans* 25:247
- Shimizu Y, Sakano K, Noguchi Y, Moriguchi T (1999) Light emitting device having a nitride compound semiconductor and a phosphor containing a garnet fluorescent material. US Patent 5,998,925
- Shimomura Y, Honma T, Shigeiwa M, Akai T, Okamoto K, Kijima N (2007) Photoluminescence and crystal structure of green-emitting $\text{Ca}_3\text{Sc}_2\text{Si}_3\text{O}_{12}:\text{Ce}^{3+}$ phosphor for white light emitting diodes. *J Electrochem Soc* 154:J35
- Shionoya S, Yen WM (1998) Phosphor handbook. CRC Press, Boca Raton, FL
- Shriver DF, Atkins PW, Langford CH (1990) Inorganic chemistry. Oxford University Press, Oxford

- Smet PF, Parmentier AB, Poelman D (2011) Selecting conversion phosphors for white light-emitting diodes. *J Electrochem Soc* 158:R37
- Sohn IS, Unithrattil S, Im WB (2014) Stacked quantum dot embedded silica film on a phosphor plate for superior performance of white light-emitting diodes. *ACS Appl Mater Interfaces* 6:5744
- Stanger-Hall KF, Lloyd JE, Hillis DM (2007) Phylogeny of North American fireflies (Coleoptera: Lampyridae): implications for the evolution of light signals. *Mol Phylogenet Evol* 45(1):33
- Stoll H, Hoppe R (1987) Ein neues Oxoplumbat (IV): $\text{CsNa}_3[\text{PbO}_4]$. *Rev Chim Min* 24:96
- Su Q (1991) Proceedings of the 2nd international conference on rare earth development and application. International Academic Publishers, Beijing, pp 765–769
- Tang JY, Chen JH, Hao LY, Xu X, Xie WJ, Li QX (2011) Green Eu^{2+} -doped $\text{Ba}_3\text{Si}_6\text{O}_{12}\text{N}_2$ phosphor for white light-emitting diodes: synthesis, characterization and theoretical simulation. *J Lumin* 131:1101
- The Homer Laughlin China Company (2011) Statement regarding the Good Morning America broadcast. Accessed 25 Mar 2012
- U.S. Environmental Protection Agency. EPA facts about uranium. Retrieved 20 Sept 2014
- Uranium containing dentures (ca. 1960s, 1970s). In Health physics historical instrumentation museum collection. Oak Ridge Associated Universities. Retrieved 10 Oct 2013
- Vahvaselkä KS, Mangs JM (1988) X-ray diffraction study of liquid sulfur. *Phys Scr* 38(5):737
- van der Kolk E, de Haas JTM, Bos AJJ, van Eijk CWE, Dorenbos P (2007) Luminescence quenching by photoionization and electron transport in a $\text{LaAlO}_3:\text{Ce}^{3+}$ crystal. *J Appl Phys* 101:083703
- Van Haecke JE, Smet PF, De Keyser K, Poelman D (2007) Single crystal CaS:Eu and SrS:Eu luminescent particles obtained by solvothermal synthesis. *J Electrochem Soc* 154:J278
- Vecht A, Gibbons C, Davies D, Jing X, Marsh P, Reland T, Silver J, Nowport A, Barber D (1999) Engineering phosphors for field emission displays. *J Vac Sci Technol B* 17:750
- Wang B, Lin H, Xu J, Chen H, Wang YS (2014) $\text{CaMg}_2\text{Al}_{16}\text{O}_{27}:\text{Mn}^{4+}$ -based red phosphor: a potential color converter for high-powered warm W-LED. *ACS Appl Mater Interfaces* 6:22905
- Waseda Y (1980) The structure of non-crystalline materials—liquids and amorphous solids. McGraw-Hill, New York, pp 48–51
- Wilson E (1999) What's that stuff? Light sticks. *Chem Eng News* 77(3):65
- Wu H, Zhang XM, Guo CF, Xu R, Wu MM, Su Q (2005) Three-band white light from InGaN-based blue LED chip precoated with green/red phosphors. *IEEE Photonics Technol Lett* 17:1160
- Xie RJ, Hirotsaki N, Li HL, Li YQ, Mitomo M (2007) Synthesis and photoluminescence properties of beta-sialon: Eu^{2+} ($\text{Si}_{6-z}\text{Al}_z\text{O}_2\text{N}_{8-z}:\text{Eu}^{2+}$). *J Electrochem Soc* 154:J314
- Xie RJ, Hirotsaki N, Sakuma K, Kimura N (2008) White light-emitting diodes (LEDs) using (oxy) nitride phosphors. *J Phys D Appl Phys* 41:144013
- Yadav RS, Pandey SK, Pandey AC (2010) Blue-shift and enhanced photoluminescence in $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}$ nanophosphor under VUV excitation for PDPs application. *J Rare Earths Mater Sci Appl* 1:25
- Yang HC, Li CY, He H, Tao Y, Xu JH, Su Q (2006) VUV-UV excited luminescent properties of $\text{LnCa}_4\text{O}(\text{BO}_3)_3:\text{RE}^{3+}$ ($\text{Ln} = \text{Y, La, Gd}$; $\text{Re} = \text{Eu, Tb, Dy, Ce}$). *J Lumin* 118:61
- Yang CC, Lin CM, Chen YJ, Wu YT, Chuang SR, Liu RS, Hu SF (2007) Highly stable three-band white light from an InGaN-based blue light-emitting diode chip precoated with (oxy)nitride green/red phosphors. *Appl Phys Lett* 90:123503
- Yap SV, Ranson RM, Cranton WM, Koutsogeorgis DC, Hix GB (2009) Temperature dependent characteristics of $\text{La}_2\text{O}_2\text{S}:\text{Ln}$ [$\text{Ln} = \text{Eu, Tb}$] with various Ln concentrations over 5–60 °C. *J Lumin* 129:416
- Yen WM, Shionoya S, Yamamoto H (2006) Measurement of phosphor properties. CRC Press, Boca Raton, FL
- Yu JJ, Gong WT, Xiao ZG, Ning GL (2012) Spectral structure of barium–phosphate–silicate phosphor $\text{Ba}_{10}(\text{PO}_4)_4(\text{SiO}_4)_2:\text{Eu}^{\text{M}+}$. *J Lumin* 132:2957
- Zhang FL, Yang S, Stoffers C, Penczek J, Yocom PN, Zaremba D, Wagner BK, Summers CJ (1998) Low voltage cathodoluminescence properties of blue emitting $\text{SrGa}_2\text{S}_4:\text{Ce}^{3+}$ and ZnS:Ag . *Cl phosphors. Appl Phys Lett* 72:2226

- Zhang Z, ten Kate OM, Delsing A, Kolk EVD, Notten PHL, Dorenbos P, Zhao J, Hintzen HT (2012) Photoluminescence properties and energy level locations of RE³⁺ (RE = Pr, Sm, Tb, Tb/Ce) in CaAlSiN₃ phosphors. *J Mater Chem* 22:9813
- Zhou J, Wang YH, Liu BT, Li F (2010) Energy transfer between Eu-Mn and photoluminescence properties of Ba_{0.75}Al₁₁O_{17.25}-BaMgAl₁₀O₁₇:Eu²⁺, Mn²⁺ solid solution. *J Appl Phys* 108:033106
- Zhu HM, Lin CC, Luo WQ, Shu ST, Liu ZG, Liu YS, Kong JT, Ma E, Cao YG, Liu RS, Chen XY (2014) Highly efficient non-rare-earth red emitting phosphor for warm white light-emitting diodes. *Nat Commun* 5:4312
- Zorenko Y, Gorbenko V, Voznyak T, Zorenko T, Kuklinski B, Turos-Matysyak R, Grinberg M (2009) Luminescence properties of phosphors based on Tb₃Al₅O₁₂ (TbAG) terbium-aluminum garnet. *Opt Spectrosc* 106:365

Component-Level Reliability: Physical Models and Testing Regulations

Cher Ming Tan

Contents

Introduction	224
Phosphor Degradation	227
Die Attach	229
Silicone Encapsulant and Plastic Lens	230
Bonding Wire	232
Reliability Prediction and Measurement Methods	234
Conclusion	235
References	236

Abstract

We have witnessed the incredible progress in Phosphor-converted LEDs technology from around 10 lm/W efficacy in 1996 to now that is in excess of 170 lm/W. This efficacy value make the white LEDs an excellent candidates for the realization of the next-generation illumination devices. Also, LEDs are having small dimension, high robustness to atmospheric agents and shocks, and fast modulation speed that can be effectively exploited to achieve a linear control of the luminous output of the devices, by means of pulse width modulation (PWM) as well as visible light communication, an alternative to Wi-Fi; their applications have been extended over simple illumination. All these applications require the LEDs to operate reliably.

To ensure the reliability of LEDs under operation, it is important to understand the degradation mechanisms of LEDs under the three external operating conditions, namely, drive current, temperature, and moisture. The degradation of LEDs can be attributed to either the GaN chip itself or the LED package. In this chapter, we will focus only on the package degradation due to the operating conditions.

C.M. Tan (✉)

Department of Electronic Engineering, Chang Gung University, Taoyuan, Taiwan

e-mail: cherming@ieee.org

In view of the general packaging structure of high power LEDs, this chapter will examine the degradation of the several components in the packaging, namely the phosphor coating, die attach, silicone encapsulant, plastic lens, and bonding wire under the above mentioned operating conditions. Reliability prediction of the LEDs and the measurement methods to quantify their reliability will also be discussed, including LM-80 and TM-21.

Introduction

Phosphor-converted LEDs were commercialized for the first time in 1996, and we have witnessed the incredible progress in LED technology from around 10 lm/W efficacy to now that is in excess of 170 lm/W. This efficacy value is significantly higher than those of conventional light sources, such as incandescent lamps (~10 lm/W) and fluorescent lamps (~90 lm/W). White LEDs can therefore be considered as excellent candidates for the realization of the next-generation illumination devices. Also, LEDs are having small dimension, high robustness to atmospheric agents and shocks, and fast modulation speed that can be effectively exploited to achieve a linear control of the luminous output of the devices, by means of pulse width modulation (PWM) as well as visible light communication, an alternative to Wi-Fi; their applications have been extended over simple illumination.

An important feature of white LEDs is their high expected reliability which is claimed to be in excess of 50,000 h, and this can be a strong argument to convince customers to adopt LEDs instead of incandescent or fluorescent lamps for lighting applications. However, over the last few years, several authors (Yanagisawa 1997; Manyakhin et al. 1998; Barton et al. 1999; Mueller-Mach et al. 2002; Polyakov et al. 2002; Cao et al. 2003; Yanagisawa and Kojima 2003; Chen et al. 2004; Narendran et al. 2004; Uddin et al. 2005; Bychikhin et al. 2005; Rossi et al. 2006; Meneghini et al. 2006, 2007, 2008a, b, 2010; Yu et al. 2007; Trevisanello et al. 2007; Hu et al. 2008; Buso et al. 2008; Jeon et al. 2009) have demonstrated that the lifetime of white LEDs is shorter than expected, due to the existence of a number of degradation mechanisms, when they are submitted to high-current or high-temperature or high-humidity operating conditions.

State-of-the-art high-power LEDs usually come with an area of 1 mm² and operate at current levels in the range between 350 mA and 1 A, depending on the specific model and manufacturer. These current levels correspond to current densities in the range of 35–100 A/cm², which were demonstrated to be sufficiently high to induce a degradation of the blue semiconductor chip (Rossi et al. 2006; Trevisanello et al. 2007; Meneghini et al. 2008b).

They are also rated for maximum junction temperatures in excess of 130–150 °C (Meneghini et al. 2010). During operation, the actual junction temperature is much higher than ambient temperature due to low luminous efficiency and thus produces a lot of heat (Panahi 2012). Unlike incandescent lighting where most of the heat generated from the filament is dissipated through radiation and is transferred directly to the ambient environment (Scheepers and Visser 2009), the major heat transfer

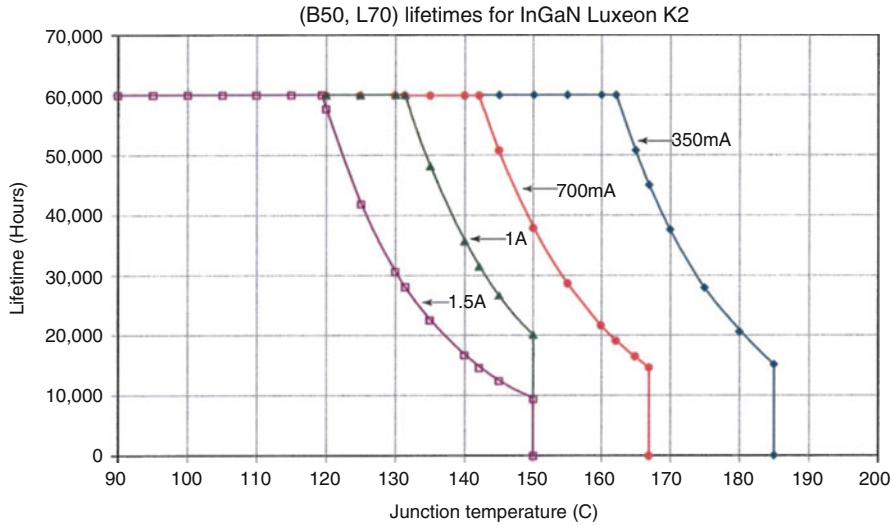


Fig. 1 Expected LED life based on drive current indicated by each colored line and target (LED junction temperatures) (Source: Philips Lumileds [USA DoE])

process in high-power LED is conduction, and the heat generated by the PN junction has to be dissipated primarily through conduction (Kückmann 2006). Therefore, the high-power LEDs have high junction temperature, and the temperature level can in principle be so high to induce severe degradations of the package materials (Manyakhin et al. 1998; Chen et al. 2004; Meneghini et al. 2007, 2008b). Figure 1 shows the effect of drive current and temperature on the lifetime of LEDs.

Besides affecting the useful lifetime, the high junction temperature and forward current may also cause the wavelength shifts. When the drive current increases, the junction temperature will increase, and this will cause a wavelength shift (Chen et al. 2004). This can be a serious failure in some applications.

As the applications of LEDs are also extended to outdoor and marine applications, humidity stress on the LEDs' lifetime is also important. Figure 2 shows the lumen output degradation of LEDs subjected to various RHs at 85 °C without bias (Tan et al. 2014). If the LEDs are operating, the high junction temperature will produce a temperature gradient that enhances moisture diffusion into the package (Fan and Yuan 2013). Fan and Yuan (2013) showed that the time to reach moisture saturation inside the package of 3 mm × 5.3 mm × 0.9 mm is 348 h for ambient temperature of 30 °C and 60 % RH without temperature gradient, and it is shortened to 82 h with temperature gradient even with RH decreased to 12.8 %.

To ensure the reliability of LEDs under operation, it is important to understand the degradation mechanisms of LEDs under the three external operating conditions, namely, drive current, temperature, and moisture. The degradation of LEDs can be attributed to either the GaN chip itself or the LED package. In this chapter, we will focus only on the package degradation due to the operating conditions, and hence only temperature and humidity are considered. The effect of drive current on

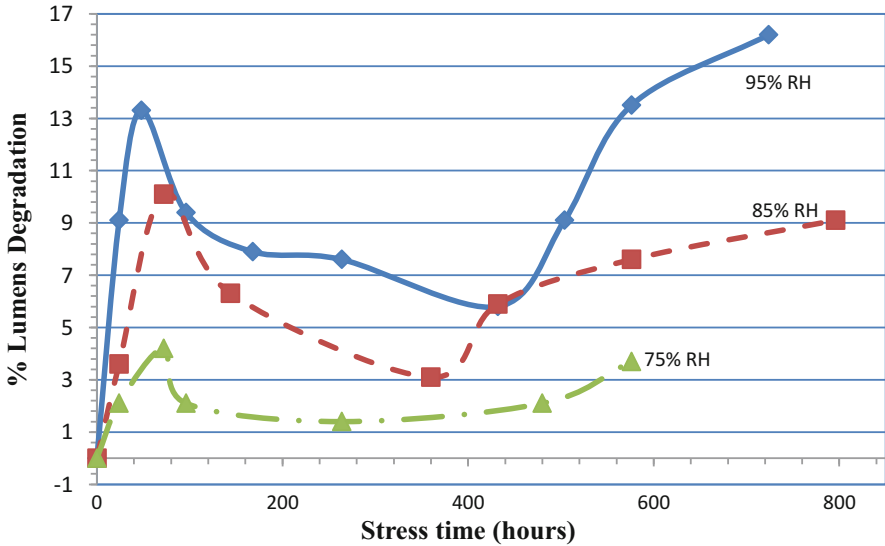


Fig. 2 Lumen degradation (in percent) of white GaN-based LEDs after prolonged accelerated humidity test at 85 °C and 85 % RH (Tan, 2014, “unpublished data”)

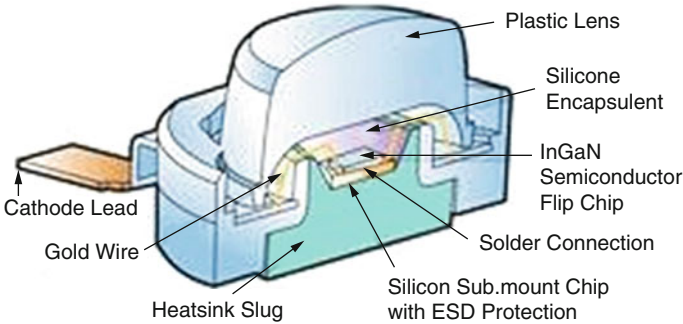


Fig. 3 Cross-sectional view of high-power LED package (www.btfsolar.com)

package degradation is mainly via temperature or temperature gradient that enhances moisture diffusion. The only effect drive current has on package degradation is that higher drive current will increase the intensity of short-wavelength emission and degrade the lens, as will be discussed later.

Figure 3 shows a typical LED structure. There are several components in the packaging of LED, namely, the phosphor coating, die attach, silicone encapsulant, plastic lens, and bonding wire. Let us examine the degradation of each component under the temperature and humidity conditions, and the degradation of the lens due to drive current will also be included.

Phosphor Degradation

Over 50 % of the electrical input power is converted into heat in the PN junction of LED chips due to non-radiative recombination of electrons and holes and light trapping inside the chip. Heat is also generated in the phosphor materials due to the non-unity quantum efficiency of phosphors, the Stokes shift loss, and the self-absorption of yellow light by the phosphor in the fluorescence events (Yan et al. 2013). It is found that phosphor temperature is critical in the thermal management of high-power white LEDs (Yan et al. 2011). An elevated phosphor temperature can result in color shift and a reduction in luminous efficacy (Zachau et al. 2007; Keppens et al. 2010).

Therefore, there are three models of phosphor coating on the LED chips. The cerium-doped Ce:YAG phosphor is either applied on the top surface of LED chips (“conformal coating”), distributed uniformly in the silicone (“phosphor in cup”), or separated from LED chips by silicone (“remote phosphor”). The thickness of phosphor layer fixed at 200 μm in the conformal coating and remote phosphor and the concentration of Ce in YAG are varied to produce a neutral white light with a correlated color temperature (CCT) of 4500 K; Fig. 4 shows the maximum temperature in the junction of LED array (“junction temperature”), the maximum temperature in the phosphor layer (“phosphor temperature”), and the optical power output in all the three models (Schubert 2006).

From Fig. 4, we can see that as the phosphor layer is moving away from the chip, i.e., from conformal phosphor to phosphor in cup to remote phosphor models, the following can be observed: (1) the optical power output increases from 12.8 W to 21 W to 22.2 W, (2) the junction temperature decreases from 98.3 $^{\circ}\text{C}$ to 91.6 $^{\circ}\text{C}$ to 90 $^{\circ}\text{C}$ to

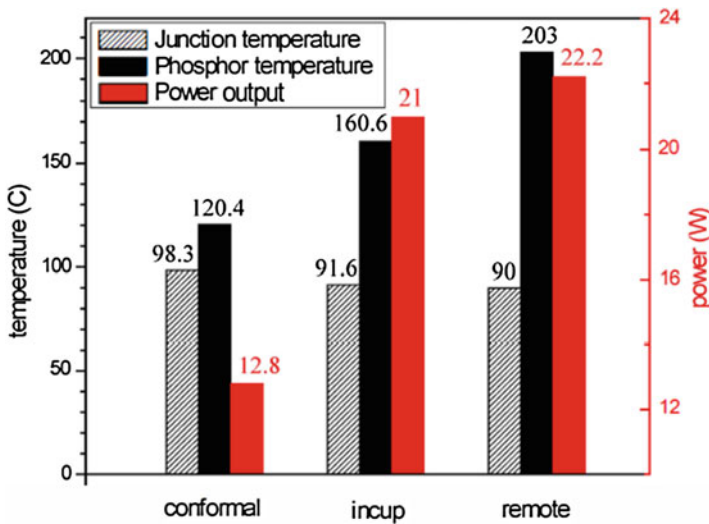


Fig. 4 Junction temperature, phosphor temperature, and power output for models with different phosphor configurations (Yan et al. 2013)

90 °C, but (3) the phosphor temperature increases dramatically from 120.4 °C to 160.6 °C to 203 °C (Yan et al. 2013).

As heat can cause irreversible damage to the phosphor (Tamura et al. 2000; Narendran et al. 2004) and with the information from Fig. 4, the different models of phosphor coating can affect the lumen output of an LED and its lifetime. Figure 5 shows the test results comparing the lifetime of conformal phosphor versus remote phosphor layer (Hwang et al. 2010). LEDs with the remote phosphor layer show rapid drop of the luminous flux at the beginning of the test, and the drop rate is still higher than that of the conformal phosphor case after the beginning period (until 200 h). This was verified to be due to the high phosphor temperature. After 1500 h test, the degradation rate of the conformal phosphor case is about 5 % but that of the remote case is 23 % (Hwang et al. 2010).

When the phosphor material is directly coated on the LED chips in the conformal phosphor model, a large amount of backscattered light from the phosphor particles is absorbed by the chips; thus light extraction from the phosphor coating as well as from the chip array package is significantly reduced, and the junction temperature in the conformal phosphor model is elevated by 6.7 °C as compared to the phosphor-in-cup model, while the optical output power is significantly reduced by 40 %. A large decrease of 40.2 °C in the phosphor temperature is observed in the conformal phosphor model compared to the phosphor-in-cup model because the heat sources, i.e., the phosphor particles, are much closer to the chips in the conformal phosphor model (within 200 μm), and the heat generated in the phosphors is effectively conducted through the chips to the heat sink (Yan et al. 2013).

In the remote phosphor model, backscattered light from the phosphor particles cannot be absorbed by the chips directly due to the separation of silicone encapsulant. Hence the remote phosphor model produces a slightly lower junction temperature and a higher optical output power compared to the in-cup model, but the phosphor temperature is the highest (Yan et al. 2013). This high phosphor temperature renders rapid lumen degradation as shown in Fig. 5 (Hwang et al. 2010).

Fig. 5 Degradation rate for different models of phosphor coating (Hwang et al. 2010)

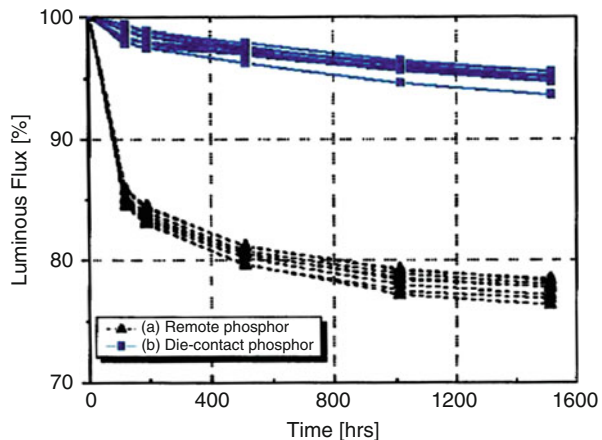
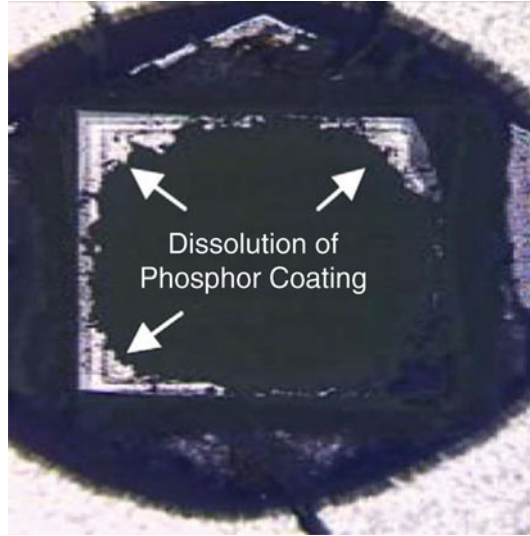


Fig. 6 Optical image demonstrating the extent of the dissolution of the phosphor coating for degraded LED (Tan et al. 2009)



For the phosphor-in-cup model, Tsai et al. (2009) showed that the thermal aging of the Ce:YAG-doped silicone causes the performance degradation of LED, and the degradation increases with the concentration of the Ce:YAG in silicone. The thermal aging can be categorized into three parts: silicone aging, Ce:YAG radioactive quenching, and degradation due to the Ce:YAG-silicone interaction. The main reason could be due to the mismatching of the two materials, such as the refractive index, the thermal property, and a chemical incompatibility. Therefore, minimizing any mismatch of the refractive index and thermal expansion between the phosphor and silicone during thermal aging is a direction of addressing thermal reliability for high-power LEDs (Tsai et al. 2009).

Phosphor coating can also be dissolved due to the inception of moisture from the ambient. Tan et al. (2009) performed unbiased 85 °C and 85 % RH on commercial LEDs and found that one of the main degradation mechanisms is the outdiffusion of the activator in the phosphor along the bondwire to the external package and the dissolution of the phosphor coating as shown in Fig. 6.

Die Attach

As temperature can affect the LED performance and reliability significantly as seen earlier, thermal management for LEDs is continuous to be a hot topic for LED luminaries. The dominant thermal path for LED chip is through the die attach to the heat sink. Any degradation on the die attach will cause an increase in the thermal resistance and impede the heat dissipation from the chip, rendering an increase in the junction temperature.

During the operation of LEDs, alternating shear strain, even thermal fatigue failure, will occur due to thermal expansion coefficient mismatch of chip and substrate. Chen et al. (2011) found that, after illuminating 1000 h, the light output power of LED module with low thermal conductivity adhesive ($K = 15\text{W/mK}$) decreased 35 %, while for highly thermal conductivity adhesive ($K = 85\text{W/mK}$), the value was only 20 %. The reason was that low thermal conductivity adhesive had low flexural modulus and low adhesion with adjacent substrate, so it cannot withstand the larger thermomechanical stress when the packages are subjected to the thermal condition variation. In serious cases, delamination would occur at the interface between die attach and substrate or heat sink, reducing the heat dissipation and degrading the performance of the devices (Chen et al. 2011).

Solder is a common die attach material for LEDs; Deng et al. (2010) found experimentally that void formation occurred in solder die attach after power cycling of the LEDs, resulting in three times increase in the thermal resistance.

Another common die attach for LEDs is nano-silver paste, and it has higher thermal conductivity than solder die attach, and thus the light output is higher for a given drive current because of lower junction temperature (Wang et al. 2009). For such die attach, Li et al. (2010) found that the average shear strength decreases with temperature cycling due to the generation and enlargement of crack in the paste. The shear strength also drops with prolonged hygrothermal aging, but the effect is a lot less than that due to thermal cycling. If one can ensure the shear force is below 11 N, then the fatigue lifetime of the paste will be very long, above 112465 cycles (Li et al. 2010).

LED packages are usually molded with polymer plastic materials. Mismatching coefficients of moisture expansion (CMEs) induce hygro-mechanical stress in LED packages and cause the LED packages to swell after absorbing moisture. Different levels of swelling occur between polymeric and non-polymeric materials as well as among the polymeric materials. This differential swelling induces hygroscopic stress in the package, adding thermal stress at high-reflow temperatures, thereby inducing delamination (Chang et al. 2012). The moisture presence in packages can also reduce interfacial adhesion strength by 40–60 % and lead to delamination (Chang et al. 2012). In most cases, the delamination begins from the corners (where the highest stress occurs) and then expands to other areas (Chang et al. 2012).

The curing of epoxy resins involves the repetition of shrinkage and the development of internal stress, which may also cause delamination (Chang et al. 2012).

Silicone Encapsulant and Plastic Lens

LEDs are encapsulated to prevent mechanical and thermal stress shock and humidity-induced corrosion. Transparent epoxy resins are generally used as an LED encapsulant. However, cured epoxy resins are usually hard and brittle owing to rigid cross-linked networks, and it degrades under exposure to radiation (blue/UV radiation) and high temperatures, resulting in chain scission (which results in radical

formation) and discoloration (due to the formation of thermo-oxidative cross-links). This is called encapsulant yellowing (Chang et al. 2012).

Modification with silicone materials has been used as an efficient method to increase the toughness and thermal stability of transparent epoxy encapsulant resin. However, silicone compound has lower glass transition temperature (T_g), larger CTE, and poor adhesion to housing (Chang et al. 2012).

The photodegradation of the encapsulant that causes yellowing depends on exposure time and the amount of radiation. Thus, even long-term exposure to visible light can cause polymer and epoxy materials to be degraded (Chang et al. 2012). It is well known that many epoxies can turn yellow when subjected to prolonged exposure to ultraviolet (UV) light as well as levels of blue light (Chang et al. 2012). Discoloration results in a reduction in the transparency of the encapsulant and causes a decrease in LED light output. Thus, with the increase in the drive current, the intensity of the short-wavelength radiation will also increase, and together with the associated increase in the junction temperature, the yellowing of the encapsulant will be accelerated (Chang et al. 2012).

It is demonstrated that degradation and the associated yellowing increase exponentially with exposure energy (amount of the light illuminating the encapsulant). The thermal effects associated with excessive junction temperature also play a role in encapsulant yellowing (Chang et al. 2012). Narendran et al. (2004) reported that the degradation rate of 5 mm epoxy-encapsulated YAG:Ce low-power white LEDs was mainly affected by junction heat and the amount of short-wavelength emissions with greater influence from the thermal effect. They demonstrated that a portion of the light circulated between the phosphor layer and the reflector cup with increased temperature can cause epoxy yellowing. Barton and Osinski (Barton et al. 1999) found that 150 °C was sufficient to change the transparency of the epoxy, causing the attenuation of the light output of LEDs. Arik et al. (2003) used finite element analysis to show that localized heating of the phosphor particles occurs during wavelength conversion because of low quantum efficiency. They reported that as little as 3 mW heat generation on a 20 μm -diameter spherical phosphor particle can lead to excessive temperatures sufficient to degrade light output.

A high-temperature gradient can cause delamination between the LED chip and the encapsulant, which forms a thin chip-air-silicone interface inside the LED package (Chang et al. 2012), and this will allow moisture to diffuse to the surface of the chip as also shown experimentally by Tan et al. (2014).

Chang et al. 2012 found that yellowing is not significantly affected by a high-humidity test environment. However, Tan et al. (2012) found that the silicone encapsulant is permeable to water vapor and can entrap moisture. This entrapped moisture scatters the white light as it travels through the encapsulant and reduces the light output. This is found to be responsible for the rapid light output degradation observed in the humidity test of LEDs. They also found that once the moisture is penetrated deep into the encapsulant, the heat generated during the normal operation of the LED is unlikely to drive out the moisture, resulting in a permanent degradation of the LED light output.

Since standard silicone retains mechanical softness in its cured state, the silicone encapsulant is enclosed in a plastic cover that serves as a lens to give mechanical protection. The plastic lenses also serve to increase the amount of light emitted from the LEDs into the free space. The failure mode of lens degradation is a number of small hairline cracks that decrease light output due to increased internal reflection in LEDs. The degradation appears due to thermomechanical stresses and hygro-mechanical stresses (Chang et al. 2012).

Chang et al. 2012 found a number of cracks introduced from thermal expansion in the center of the lens surface and on the inside of the polymer encapsulation when high-power LED samples with three different lens shapes were aged at 80 °C, 100 °C, and 120 °C under a constant voltage of 3.2 V. The hemispherical lens LEDs had longer lives than the cylindrical and elliptical-shaped plastic lenses due to a more uniform thermal dissipation along the thermal path from the LED chip to the lens (Chang et al. 2012). Extreme thermal shock can also cause crack in an epoxy lens due to the induced mechanical stress (Chang et al. 2012).

Tsai et al. (2009) found that the LED chip placement can affect the crack generation on the lens when subjected to high temperature with a constant current of 350 mA. Three types of chip location were used for their experiments. Type I has the chip in the cavity of heat sink; type II has the chip on top of heat sink; and type III has the chip in the shallow cavity of heat sink. They observed that type I degraded fastest, followed by type II, and type III is the least. The differences in the degradation rates are larger at higher temperature. Rapid degradations were observed for types I and II at 95 °C due to the yellowing of the lens at high temperature, in addition to the crack on the lens (Tsai et al. 2009).

Once crack lines occur on the lens, moisture and contamination from ambient will be able to penetrate into the encapsulant and speed up further degradation.

Long-term exposure to high condensing moisture can also cause cloudiness of the epoxy lenses in a plastic LED lamp due to hygro-mechanical stresses (Chang et al. 2012).

Bonding Wire

The degradation of the gold bonding wire is mostly similar to the standard integrated circuit packaging, and extensive review on the gold bonding wire degradation is readily available. However, there is one unique degradation of gold bonding wire in LED which is due to its bonding onto silver mirror on the chip.

Tan et al. (2014) studied the degradation mechanisms of LEDs under various relative humidities at 85 °C. They found that the initial degradation under humid conditions occurs at the bond pad, and the mechanism is silver atom migration into gold wire. It is known that Ag is a rapid diffuser in Au and Au is a slow diffuser in Ag; thus Kirkendall voids can easily be formed at the interface between the Au wires bonding on Ag. At elevated temperature of 85 °C, such void formation will be further enhanced. When Ag is diffused into Au, it is completely dissolved into Au as the Ag-Au binary system has perfect solid solubility.

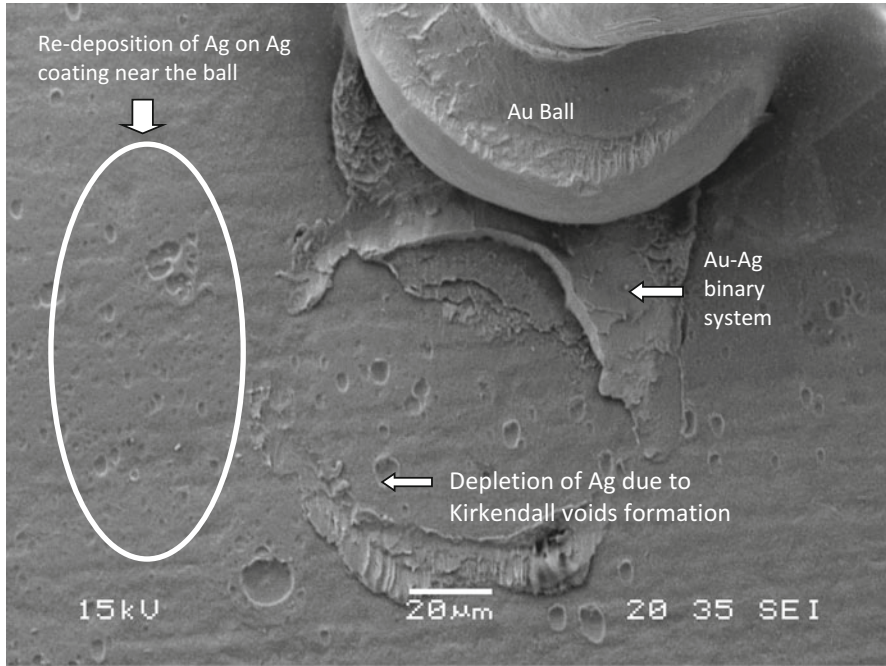


Fig. 7 SEM micrograph of the Ag coating near the Au bonding ball of a degraded LED at 70 % RH and 85 °C after 184.6 h. The movement of the gold ball away from its original place could be an artifact of the decapsulation, but it clearly indicates the weakening of the bond (Tan et al. 2014)

Under the temperature-humidity test, moisture will enter into the package through the interface between the lead and the package and also the interface between the lens and the package. Thus layers of water will be present at the Au-Ag interface. As the electrode potential of Au is 1.52 V which is much higher than that of Ag ($= 0.8$ V) at 25 °C, the Au-Ag contacts at the bond pad can easily form a galvanic cell during the temperature-humidity test, with Ag as anode, and anodic dissolution of Ag from the coating will occur with a removal rate that can be as high as 35.6 nm/s. The mechanism of Ag migration proceeds in three steps: electro-dissolution, ion transport, and electrodeposition elsewhere. The activation energy of Ag^+ diffusion is found to be 0.2 eV, and the electrodeposition will result in either Ag clouds or dendrites.

Combining the evidence of fast diffusion of Ag into Au ball and the anodic dissolution of Ag, Tan et al. (2014) discovered that the moisture that enters into the LED package during the temperature-humidity test can penetrate through the voids at the Au ball-Ag coating interface and remove the Ag from underneath the ball, increase the contact resistance of the Au ball on Ag, and thus weaken the bonding and increase the series resistance. The dissolved Ag can then be redeposited on the Ag coating around the ball area. All these are observed in Fig. 7 which shows the SEM micrograph of Ag coating near the bond of a degraded LED under 70 % RH and 85 °C and the movement of the Au bonding ball after decapsulation.

Reliability Prediction and Measurement Methods

The lifetime prediction of high-power LEDs is now formulated with TM-21 that uses the data obtained using LM-80 standard.

The creation of the LM-80 standard by the Illuminating Engineering Society of North America (IESNA) is to provide a consistent method for manufacturers to measure lumen maintenance of LED light sources such as LED package, array, or module alone, i.e., at component level. Here, lumen maintenance refers to the number of hours that a light source remains “useful” before its output diminishes to 70 %. The measurement requires at least 6000 h of testing at case temperature of 55 °C and 85 °C and a third temperature to be selected by the manufacturer.

Before the advent of LM-80, LED component manufacturers each reported lumen maintenance data using their own systems. To avoid customer confusion, members of IESNA came together to create a standard methodology that would allow customers to evaluate and compare the lumen maintenance of LED components from different companies.

LM-80 can be a useful tool for lighting professionals who are looking to analyze LED products; however, it is not a measure of LED system performance or reliability. Unfortunately, LED component, lamp, and luminaire manufacturers, among others, often use the data found in an LM-80 report to substantiate the “lifetime” claims of an LED product, even though that data alone cannot be used to predict the useful life of an LED product or system.

TM-21, developed by IESNA, takes the LM-80 data and makes useful LED lifetime projections. The results can then be used to interpolate the lifetime of an LED source within a system (luminaire or integrated lamp) using the in-situ LED source case temperature.

Before the arrival of TM-21, there was no standard basis for the extrapolation, and calculation varied among manufacturers. TM-21 dictates which values can be used in the calculation based on the sample size, number of hours and interval tested, and test suite temperature. It also puts a cap on the extrapolation, a maximum of six times the hour tested. The method of extrapolation uses exponential least square curve fit.

Accurate lumen and temperature measurements of LED are important for LM-80 and TM-21 to predict its lifetime. In fact, LM-80 only specific case temperature, but the reliability of a LED depends both on the case and junction temperature, and thus both temperatures should be obtained for accurate reliability prediction of LEDs. The junction temperature can be obtained either through measurement or computation via thermal resistance of the package.

There are three methods to measure LED temperature, namely, the infrared thermal imaging method, spectra method, and electrical method.

Infrared thermal imaging method is a common method to measure the thermal field distribution of semiconductor devices, and it can realize real-time dynamic measurement of thermal resistance. However, the temperature which is directly read from the infrared thermograph image is the surface radiation temperature of objects other than the true junction temperature. Also, it is limited by the accuracy of spatial resolution and temperature calibration.

High-power LED radiates light of different frequencies when it is under working conditions, and the radiation intensity of spectral lines and the wavelength (or the main frequency) of radiant light of PN junction are related to the temperature of PN junction (Chen et al. 2009), and the change of the peak wavelength due to the increase in junction temperature is linear (Ge et al. 2010) regardless of the manufacturers. One can thus use the spectra method to obtain the change in the peak wavelength and hence the change in the junction temperature from time zero (i.e., ambient temperature) directly and contactless.

When there is a lower forward current through high-power LED, there is a good linear relation between the variation of PN junction temperature and the variation of forward voltage drop of the LED; hence one is able to obtain the junction temperature also directly from its low-current forward voltage drop (Wang et al. 2011), and the measurement method is simple and direct without expensive equipment.

However, regardless of which method is used for temperature measurement, one has to ensure that there is no self-heating of the PN junction by the measurement current. This also applies to the lumen measurement from the LEDs. Chen et al. (2012) used the response surface method to determine the optimal setting of pulse width modulation of the drive current so that the heat generated by the pulse current can be dissipated quickly to the ambient, ensuring no self-heating during the measurement. They found that the duty cycle of the pulse width modulation is to be 4 % and the pulse width is to be 30 ms.

For the determination of the junction temperature from the case temperature, the thermal resistance from the chip to the case must be determined, and the heat generated from the chip must also be determined. T3Ster method (Poppe 2012) is widely used for the thermal resistance measurement and is now adopted as part of JEDS 50. The heat generated from the chip can only be estimated by assuming all the non-radiative energy loss is converted into heat. For accurate determination of the junction temperature, one has to perform finite element method to model the heat flow from the chip to the case in 3D, with the thermal resistances of the various paths determined.

Although temperature is a main factor to the degradation of LEDs, drive current and humidity can also affect the degradation rate of LEDs as we have seen earlier, and in combination with the temperature, the degradation rate of LEDs will be accelerated. LM-80 plus TM-21 covers only the lifetime of LEDs due to temperature effect, and the actual lifetime of LEDs should be smaller than that determined using TM-21, and one should use the results with cautions.

Conclusion

High-power LED is a complex device with many components, and each component has several degradation mechanisms which are affected by the temperature of the chip, temperature gradient within the LED package, as well as ambient humidity. In this chapter, we focus ourselves only on LED packaging degradation, and the various known degradation mechanisms are described. From these mechanisms,

one can see that material selection, phosphor location, and Ce doping concentration, as well as lens design, are to be considered in totality in order to ensure LED reliability.

The lifetime estimation of LEDs and the associated measurement methods are also discussed. While it is necessary to have standardized methods for lifetime estimation, one needs to be aware that the current method is limited to the temperature degradation-related lifetime only. Also, one needs to note that LED lifetime will not be equal to the LED luminaries' lifetime as the luminaries consist of many other devices.

References

- Arik M, Weaver S, Becker CA, Hsing M, Srivastava A (2003) Effects of localized heat generations due to the color conversion in phosphor conversion in phosphor particles and layers of high brightness light emitting diodes. In: ASME international electronic packaging technical conference and exhibition, pp 1–9
- Barton DL, Osinski M, Perlin P, Eliseev PG, Lee J (1999) Single-quantum well InGaN green light emitting diode degradation under high electrical stress. *Microelectron Reliab* 39:1219
- Buso S, Spiazzi G, Meneghini M, Meneghesso G (2008) Performance degradation of high brightness light emitting diodes under DC and pulsed bias. *IEEE Trans Device Mater Reliab* 8:312
- Bychikhin S, Pogany D, Vandamme LJK, Meneghesso G, Zanoni E (2005) Low frequency noise sources in as-prepared and aged GaN-based light-emitting diodes. *J Appl Phys* 97:123714
- Cao XA, Sandvik PM, LeBoeuf SF, Arthur SD (2003) Defect generation in InGaN/GaN light-emitting diodes under forward and reverse electrical stresses. *Microelectron Reliab* 43:1987
- Chang M-H, Das D, Varde PV, Pecht M (2012) Light emitting diodes reliability review. *Microelectron Reliab* 52:762–782
- Chen ZZ, Zhao J, Qin ZX, Hu XD, Yu TJ, Tong YZ (2004) Study on the stability of the high-brightness white LED. *Phys Status Solidi B* 241:2664
- Chen H, Liu Y, Chen Z, Zhang H (2009) Analysis of thermal spreading boards for high-power AlGaInP red LEOs. *Acta Optica Sinica* 29:805–810
- Chen M, Xu T, Liu S, Wong CP (2011) Study on thermal conductive adhesives for high power LEDs packaging. In: Proceedings of the international symposium on advanced packaging materials, pp 104–108
- Chen S, Tan CM, Chen BK, Chua ZY (2012) Ensuring accuracy in optical and electrical measurement of ultra-bright LEDs during reliability test. *Microelectron Reliab* 52:1632–1635
- Deng H, Feng S, Guo C, Qiao Y, Zhang G (2010) Reliability of solder joints in high-power LED package in power cycling tests. In: Proceedings of the 10th IEEE international conference solid-state and integrated circuit tech, pp 1683–1685
- Fan X, Yuan C (2013) Effect of temperature gradient on moisture diffusion in high power devices and the applications in LED packages. In: IEEE ECTC, pp 1466–1470
- Ge C, Feng S, Zhang G, Ding K (2010) Enhanced thermal measurements of high power LEOs by junction characteristic. In: Inter workshop on junction tech, pp 202–205
<http://www.btf-solar.com/brightledlight.htm>
- Hu J, Yang L, Shin MW (2008) Electrical, optical and thermal degradation of high power GaN/InGaN light-emitting diodes. *J Phys D Appl Phys* 41:035107
- Hwang JH, Kim YD, Kim JW, Jung SJ, Kwon HK, Oh TH (2010) Study on the effect of the relative position of the phosphor layer in LED package on the high power LED lifetime. *Phys Status Solidi C* 7:2157–2161

- Jeon S-K, Lee J-G, Park E-H, Jang J, Lim J-G, Kim S-K (2009) The effect of the internal capacitance of InGaN-light emitting diode on the electrostatic discharge properties. *Appl Phys Lett* 94:131106
- Keppens A, Hansellaer P, Zong Y, Ohno Y (2010) Characterization of remote phosphor type of LEDs. In: CORM 2010, Las Vegas
- Kückmann O (2006) High power LED arrays Special requirements on packaging technology. In: Proceedings of SPIE, pp 613404–613406
- Li X, Chen X, Yu D, Lu GQ (2010) Study on adhesive reliability of low temperature sintered high power LED modulus. In: Proceedings of the international conference electronic packaging tech & high density packaging, pp 1371–1376
- Manyakhin F, Kovalev A, Yunovich AE (1998) Aging mechanisms of InGaN/AlGaIn/GaN light-emitting diodes operating at high currents. *MRS Internet J Nitride Semicond Res* 3:1998
- Meneghini M, Trevisanello L-R, Zehnder U, Zahner T, Strauss U, Meneghesso G (2006) High-temperature degradation of GaN LEDs related to passivation. *IEEE Trans Electron Devices* 53:2981
- Meneghini M, Trevisanello L-R, Zehnder U, Meneghesso G, Zanoni E (2007) Reversible degradation of ohmic contacts on p-GaN for application in high brightness LEDs. *IEEE Trans Electron Devices* 54:3245
- Meneghini M, Rigutti L, Trevisanello L, Cavallini A, Meneghesso G, Zanoni E (2008a) A model for the thermal degradation of metal/(p-GaN) interface in GaN-based LEDs. *J Appl Phys* 103:063703
- Meneghini M, Trevisanello L, Meneghesso G, Zanoni E (2008b) A review on the reliability of GaN-based LEDs. *IEEE Trans Device Mater Reliab* 8:323
- Meneghini M, Tazzoli A, Mura G, Meneghesso G, Zanoni E (2010) A review on the physical mechanisms that limit the reliability of GaN-based LEDs. *IEEE Trans Electron Devices* 57:108
- Mueller-Mach R, Mueller GO, Krames MR, Trottier T (2002) High-power phosphor converted light-emitting diodes based on III-nitrides. *IEEE J Sel Top Quantum Electron* 8:339
- Narendran N, Gu Y, Freyssonier JP, Yu H, Deng L (2004) Solid-state lighting: failure analysis of white LEDs. *J Cryst Growth* 268:449–456
- Panahi AS (2012) Review of recent technological advances in High Power LED Packaging. In: Proceedings of SPIE, pp 83680T
- Philips L (2007) Understanding power LED lifetime analysis. Technology White Paper, Philips Lumileds Lighting Company, San Jose
- Polyakov AY, Smirnov NB, Govorkov AV, Kim J, Luo B, Mehandru R (2002) Enhanced tunneling in GaN/InGaN multi-quantum-well heterojunction diodes after short-term injection annealing. *J Appl Phys* 91:5203
- Poppe A (2012) Thermal test methodology and standards for lighting LEDs. www.mentor.com/micred
- Rossi F, Pavesi M, Meneghini M, Salviati G, Manfredi M, Meneghesso G (2006) Influence of short-term low current dc aging on the electrical and optical properties of InGaN blue light-emitting diodes. *J Appl Phys* 99:053104
- Scheepers G, Visser JA (2009) Detailed thermal modeling of high powered LEDs. In: 25th IEEE SEMI-THERM symposium, San Jose, pp 87–91
- Schubert EF (2006) Light-emitting diodes, 2nd edn. Cambridge University Press, New York, Chapter 11
- Tamura T, Setomoto T, Taguchi T (2000) Illumination characteristics of lighting array using 10 candela-class white LEDs under AC 100 V operation. *J Lumin* 87–89:1180–1182
- Tan CM, Singh P (2014) Time evolution degradation physics in high power white LEDs under high temperature-humidity conditions. *Device and Materials Reliability*, IEEE Transactions on 14.2, pp 742–750
- Tan CM, Chen BK, Xu G, Liu Y (2009) Analysis of humidity effects on the degradation of high-power white LEDs. *Microelectron Reliab* 49:1226–1230

- Tan CM, Chen E, Li X, Chen SJ (2012) Rapid light output degradation of GaN-based packaged LED in the early stage of humidity test. *IEEE Trans Device Mater Reliab* 12:44–48
- Trevisanello L-R, Meneghini, Mura G, Sanna C, Buso S, Spiazzi G, Vanzi M, Meneghesso G, Zanoni E (2007) Thermal stability analysis of high brightness LED during high temperature and electrical aging. In: *Proceedings of SPIE*, vol 6669
- Tsai CC (2009) Decay mechanisms of radiation pattern and optical spectrum of high power LED modules in aging test. *IEEE J Sel Top Quantum Electron* 15:1156–1162
- Tsai CC, Wang J, Chen MH, Hsu YC, Lin YJ, Lee CW, Huang SB, Hu HL, Cheng WH (2009) Investigation of Ce:YAG doping effect on thermal aging for high power phosphor-converted white light emitting diodes. *IEEE Trans Device Mater Reliab* 9:367–371
- Uddin A, Wei AC, Andersson TG (2005) Study of degradation mechanism of blue light emitting diodes. *Thin Solid Films* 483:378
- USA Department of Energy, LED measurement series: LED luminaries reliability. http://cool.conservation-us.org/byorg/us-doe/luminaire_reliability.pdf
- Wang T, Lei G, Chen X, Guido L, Khai N, Lu GQ (2009) Improved thermal performance of high power LED by using low-temperature sintered chip attachment. In: *Proceedings of the international conference on electronic packaging technology & high density packaging*, pp 581–584
- Wang H, Dong J, Liu Z, Liang B (2011) The analysis of measurement methods for high power LED thermal resistance. In: *6th international forum on strategic technology*, pp 867–870
- Yan B, Tran NT, You J-P, Shi FG (2011) Can junction temperature alone characterize thermal performance of white LED emitters? In: *IEEE photonics technology letters*, pp 555–557
- Yan B, You PY, Tran NT, Shi FG (2013) Influence of phosphor configuration on thermal performance of high power white LED array. In: *Proceedings of the international symposium on advanced packaging materials*, pp 274–289
- Yanagisawa T (1997) Estimation of the degradation of InGaN/AlGaIn blue light emitting diodes. *Microelectron Reliab* 37:1239
- Yanagisawa T, Kojima T (2003) Degradation of InGaN blue light-emitting diodes under continuous and low-speed pulse operations. *Microelectron Reliab* 43:977
- Yu T, Shang S, Chen Z, Qin Z, Lin L, Yang Z (2007) Luminescence degradation of InGaN/GaN violet LEDs. *J Lumin* 122:696
- Zachau M, Fiedler T, Jermann F (2007) Phosphors – key materials for solid-state lighting. presented at the LEDs in general lighting workshop, EU JRC, ISPRA, Italy

Thermal Management: Component to Systems Level

Te-Yuan Chung

Contents

Basic Heat Transfer Phenomena	240
Thermal Energy and Temperature	240
Thermal Energy Generation	240
Heat Dissipation and Thermal Insulation	241
Thermal Resistance	241
Heat Capacity	244
Conduction Shape Factor of a Circular Surface Heat Source on a Disk	244
General Consideration of LED Lamp in Thermal Perspective	246
Heat Sources Within a LED Lamp	247
LED Chip	248
Heat Generation Within Phosphor	249
The Temperature-Dependent Properties of LED	252
Temperature-Dependent Physical Quantities of LED Chip and Phosphor	252
Temperature-Related Phosphor Properties	254
Silicone	255
Transient Behavior of LED and Light Decay	255
Thermal Management of LED	256
Conductive Part	257
Convection Part	260
Estimation of Thermal Resistance of an LED Lamp	261
References	266

Abstract

Thermal management is one of the most essential issues for LED applications. The output power, efficiency, emission spectrum and reliability of the LED chip and phosphor are functions of temperature. Without proper thermal management, the output of the LED could depart from the desired performance. Basic heat

T.-Y. Chung (✉)

Department of Optics and Photonics, National Central University, Zhongli City, Taiwan

e-mail: tychung@dop.ncu.edu.tw

transfer phenomena and corresponding calculations are introduced as well as the thermal behaviors and concerns of the essential components of an LED. A quick method to estimate the thermal resistance of an LED package is provided based on the shape factor of the geometry at the end of this chapter.

Basic Heat Transfer Phenomena

Heat transfer is a knowledge system that studies the exchange of thermal energy between systems. When the thermal energy transfers, various physical quantities change accordingly. The corresponding physical may include temperature, pressure, internal energy, etc. Usually, thermal energy transfers from high-temperature region into low-temperature region. Heat transfer phenomena can be categorized into conduction, convection, and radiation. Conduction is to describe the heat transfer between contacting solid materials. Convection is to describe the heat transfer between contacting solid and fluid. Radiation is to describe the heat exchange via EM wave emission and absorption; therefore, even if two objects are not contacting, heat exchange can still be achieved. Detail theory and analysis of the three basic heat transfer phenomena can be found in textbooks of heat transfer (Incropera et al. 2007). Only the crucial concepts will be discussed in the following paragraphs.

Thermal Energy and Temperature

Thermal energy is the total kinetic energy of all particles in a system considered. Temperature is the measurement of the average kinetic speed of all particles in the system considered (Huang 1987). Therefore, a system that has higher temperature comparing with other systems does not necessary indicate this system have higher thermal energy than any other systems. However, giving energy to a thermally insulated system will rise the system temperature. Theoretically speaking, if the thermal insulation is perfect, continuously supplying heat flux to the system will lead to infinite high temperature regardless how small the heat flux is.

Thermal Energy Generation

Based on the second law of thermodynamics, thermal energy can be generated in various physical processes from other form of energies with great efficiency (Schroeder 2000). For LED, electric energy is transferred into light emission and thermal energy. Phosphor-converted LED (pc-LED) uses phosphor that absorbs the higher energy photons that emit from the LED and converts them into lower energy photons. Therefore, the energy difference between the higher and lower energy photon turns into heat. More details regarding the thermal energy generation within LED will be described in the next session.

Heat Dissipation and Thermal Insulation

When thermal energy is generated in a certain region, local temperature will rise. Larger thermal energy gives higher temperature. As described above, thermal energy transfers from high-temperature region to low-temperature region following thermal equation (Incropera et al. 2007). Therefore, the temperature of the higher-temperature region will become lower. Many practical systems including many LED lamps generate thermal energy continuously. Thermal energy generation and dissipation rates are in the unit of power; therefore, thermal power is the common term to be used instead of the thermal energy generation/dissipation rate. When the thermal generation power equals to the dissipation power, the system is known as at the thermal equilibrium condition which indicates the temperature distribution is not a function of time. If the thermal generation power is larger than the dissipation power, the temperature will rise. On the contrary, if the thermal generation power is lower than the dissipation power, the temperature will reduce.

Thermal Resistance

To evaluate the temperature distribution of a thermal system, thermal resistance is a useful tool to give a quick estimation (Incropera et al. 2007). The way of utilizing thermal resistance can be easily found in any heat transfer textbook. The most fundamental definition of the thermal resistance is the temperature difference between two isotherm (contour) surfaces divided by the thermal power passing through the isotherm surfaces or

$$R = \frac{\Delta T}{P_{th}}, \quad (1)$$

where ΔT is the temperature difference between the isotherm surfaces and P_{th} is the thermal power. Thermal resistance is in the unit of K/W in MKS. Thermal resistance can describe how easy and how efficiently heat transfer can be achieved. Smaller thermal resistance indicates better exchange. On the contrary, larger thermal resistance indicates better thermal insulation. By following the basic definition of conduction, convection, and radiation, thermal resistance can be derived (Incropera et al. 2007). For 1D conduction, the thermal resistance can be written as

$$R_{\text{cond}} = \frac{L}{kA}, \quad (2)$$

where L is the thickness of the conductive material, k is the thermal conductivity of the material, and A is the cross section of the isotherm surface area. To reach smaller thermal resistance, material with large thermal conductivity has to be used. On the geometric concern, the material is better to be thin and has large cross section. In other words, the material has to have large area to the next thermal stage, and the

thermal propagation distance has to be short. For 2D and 3D cases, thermal resistance are usually hard to get a close form since the temperature distribution on the surface of a 2D or 3D object is usually not uniform. However, if the material used has very high thermal conductive, the surface of the object may be approximated as an isotherm. Therefore, the 1D thermal resistance may be employed for a coarse estimation for the thermal performance. Thermal resistance values may be listed in the specification of commercially available LED chip or module. Such thermal resistance is estimated using the temperature difference between the chip and the temperature at the surface position on the major cooling path where it is closest to the chip mounting area (2005). Although the LED chip or module should be considered as 2D or even 3D structure for thermal management, such measurement method can still provide a way to evaluate the thermal resistance. For convection, the thermal resistance can be written as

$$R_{\text{conv}} = \frac{1}{hA}, \quad (3)$$

where h is the heat transfer coefficient of the contacting fluid and solid interface and A is the contacting area of the fluid and solid interface. To reduce the convection thermal resistance, larger heat transfer coefficient and large heat exchange area are preferred. In general, the heat transfer coefficient for free convection is about 10–100 W/m²K and depends on different temperature, geometry, and fluid properties (Incropera et al. 2007). For radiation, the thermal resistance can be written as

$$R_{\text{rad}} = \frac{1}{h_{\text{rad}}A} = \frac{1}{\varepsilon\sigma(T_1^2 + T_2^2)(T_1 + T_2)A}. \quad (4)$$

where ε is the emissivity of the radiation surface, σ is the Stefan–Boltzmann constant ($\sigma = 5.67 \times 10^{-8}$ [W/m² · K]), A is the radiation surface area, and T_1 and T_2 are the radiation surface temperatures. Surface with emissivity of 1 is referred as a black body which has the highest possible radiation heat transfer capability. On the contrary, a surface having 0 emissivity cannot perform radiation heat transfer and is known as a white body. Realistic objects have emissivity between 0 and 1. Please note that the Stefan–Boltzmann constant is relatively small which results in poor heat transfer if the temperatures of the surfaces are low.

Conduction thermal resistance as in Eq. 2 is determined by the thermal conductivity and geometry of the object considered. Thermal conductivity is a material property which is independent from the geometry of the object. Two major phenomena contribute to thermal conductivity. One is the free electron gas and the other is lattice vibration or phonon in the material considered. Free electrons behave like gas in electric conductive material and effectively carry energy away from the heat source when heating up (Huang 1987). Materials with better electric conductivity have better thermal conductivity in general. The electric resistance comes from the scattering of free electrons by the defect or impurity in the material. Larger electric conductivity indicates that the free electron collides less with the atoms and/or loses

less energy in each collision [modern physics]. Similarly, when electron gas expands due to the high temperature, less collision gives better thermal conductivity and leads to smaller thermal resistance. For nonconductive material, thermal energy transfers mainly by lattice vibration which is known as phonon in quantum mechanics (Kittel 1996). Better crystal lattice structure allows phonons moving more freely without scattering; therefore, single crystalline has the best thermal conductive, and amorphous one always has the worst thermal conductivity for the same material. The bond strength between the atoms is also crucial. Higher bond strength gives better mechanical hardness and higher thermal conductivity (Kittel 1996).

On the contact surfaces of different conductive parts, contact thermal resistance should be carefully considered. Due to the surface roughness, two pieces of material may only have very small and sparse contact surface area in a microscopy point of view. Small and sparse contact surface can also be described as voids in between the contacting surfaces. Such small contact surface area becomes the bottleneck of heat exchange. Therefore, considerable temperature drop usually can't be neglected between the contacting surfaces. Commonly, a contact thermal resistance can be defined in a similar manner using the temperature drop divided by the thermal power passing through the nominal macroscopic contact area (Incropera et al. 2007). Three major ways can be used to reduce the contact thermal resistance. One is to polish the contact surfaces to reduce the surface roughness. Another is to apply pressure on the materials and increase the contact surface area. The other is to insert another material which fills the void in between the contacting surfaces. Thermal conductive adhesive, solder, soft metal, and thermal paste serve the purpose in general. In practice, contact thermal resistance cannot be easily estimated unless a preliminary experiment is performed using the same material with the same surface finishing and pressure condition. The thermal resistance of typical contacted surfaces can easily reach 10^{-5} K/W per unit contact area in mm^2 (Chung et al. 2012).

Heat transfer coefficient is a rather complicated function which depends on temperature, viscosity, flow direction, gravitation, flow speed, Rayleigh number, etc. (Incropera et al. 2007). Therefore, heat transfer coefficient can't be easily described. Typically, convection can be categorized into free convection and forced convection. Free convection requires no additional input energy and is preferred for LED cooling. Table 1 gives the range of heat transfer coefficients of air and water. Clearly, air has much lower heat transfer coefficient compared with water. Therefore, a large area of heat transfer is required to achieve effective heat transfer for free convection air cooling. Fin structure is the most common heat sink solution to achieve large heat transfer area. Typically, fin exchanger heat sinks are made of high thermal-conductive material such as aluminum or copper. High thermal

Table 1 Range of the heat transfer coefficient of air and water in the unit of $[\text{W}/\text{K}\cdot\text{m}^2]$ (Incropera et al. 2007)

Heat transfer coefficient range	Air	Water
Free convection	5–25	20–100
Forced convection	10–200	50–10000
Boiling water		3000–100000
Condensing water vapor		5000–100000

conductivity material can ensure that the heat exchange surfaces have more uniform temperature distribution and allow utilizing the largest possible area to achieve efficient heat transfer.

Radiation heat transfer coefficient itself is a temperature-dependent function as Eq. 4 suggested. Qualitatively speaking, radiation heat transfer coefficient increases as the absolute temperatures of the radiation surfaces and the temperature difference between the radiation surface and the heat reservoir are higher. Typical LED is operated at temperature lower than 120 °C. However, the environment temperature is usually about 25 °C. Therefore, heat transfer via radiation for LED lamp is relatively less important than conduction and convection.

Heat Capacity

Heat capacity is a physical quantity which describes the temperature change of an object after a given amount of thermal energy is given. Heat capacity depends on the mass of the object and the specific heat capacity of the composition material of the object. The specific heat capacity has the same definition as heat capacity except it is normalized to the mass of the material and is a basic material property. A larger piece of object requires more thermal energy to rise temperature. Such behavior is similar to the capacitor in circuitry. An object with larger heat capacity rises its temperature slower than an object with smaller heat capacity if the heat flux is identical. Namely, a thermal system with larger heat capacity will take longer time to reach thermal equilibrium. Nevertheless, heat capacity does not directly influence the thermal equilibrium temperature of a thermal system. Practically, each object has both thermal resistance and heat capacity. When considering the transient behavior of a thermal system, each object can be considered as parallel-connected resistor and capacitor. By analyzing the temporal behavior of a thermal system, the thermal resistance and capacity of each component may be evaluated (Szekely 1997; Kim et al. 2008).

Conduction Shape Factor of a Circular Surface Heat Source on a Disk

In 1D conduction case, thermal resistance can be written as Eq. 1. Clearly, k is material dependent. L and A are geometry of the material considered. In other words, thermal resistance can be rewritten as

$$R_{\text{cond}} = \frac{1}{kS}, \quad (5)$$

where S is known as the shape factor with unit in length. The shape factor is trivial in 1D case. When dealing with much more complex shape, the shape factor usually does not have a close form except some simple geometrical cases (Incropera et al. 2007). LED is usually mounted on a submount. Therefore, the shape factor

Fig. 1 A circular flat heat source with radius of r_s is located at the center of the top surface of a disk shape material. The thickness of the disk is L and the radius is d , and the material has thermal conductivity of k . The bottom surface of the material is kept at a constant temperature. The rest of the surfaces are set to be thermal insulation

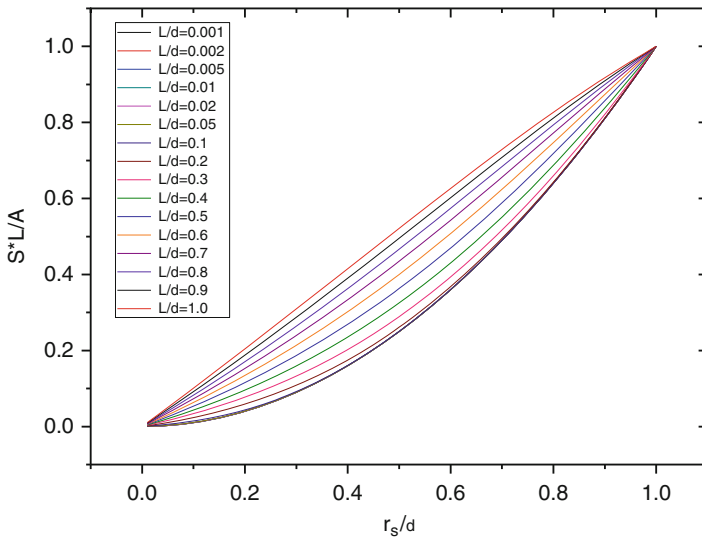
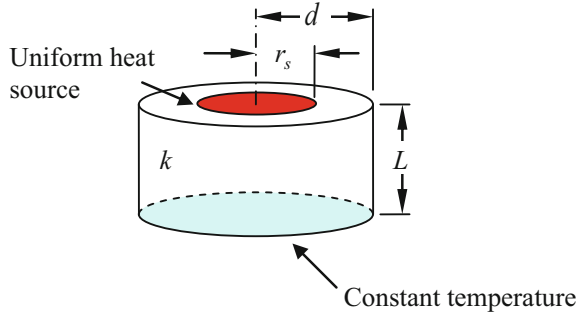


Fig. 2 The relation of the shape factor and geometry parameters of the thermal system as shown in Fig. 1

of such geometry is useful for estimating the corresponding thermal resistance. For simplicity, an axial symmetry case is considered which is shown as in Fig. 1. The heat source is assumed to be uniform.

The shape factor of such geometry can be obtained using numerical method such as finite element analysis (FEA). Figure 2 shows the obtained relation of the shape factor and the geometrical parameters for the geometry shown in Fig. 1. The vertical axis is $S \cdot L/A$ where A is the area of the heat source and equals to πd^2 . The horizontal axis is r_s/d or the heat source radius and disk radius ratio. Different curves indicate different L/d ratios. Smaller L/d ratio implies the disk is thinner. Regardless of the L/d ratio, these curves all intersect at point (0,0) and point (1,1). Point (1,1) refers that the heat source cover the disk's upper surface; therefore, this case is essentially the 1D

conduction and the shape factor is exactly A/L . For small source or L/d is less than 0,1, this curve can be approximated as a quadratic curve or

$$\frac{S \cdot L}{A} = \left(\frac{r_s}{d}\right)^2. \quad (6)$$

From Eq. 6, the shape factor can be obtained as

$$S = \frac{A}{L} \left(\frac{r_s}{d}\right)^2 = \frac{\pi r_s^2}{L} = \frac{A_s}{L}, \quad (7)$$

where A_s is the heat source area. When L/d equals 1, the curve in Fig. 2 can be approximated by

$$\frac{S \cdot L}{A} \approx \frac{r_s}{d}. \quad (8)$$

From Eq. 8, the shape factor can be obtained as

$$S \approx \frac{A}{L} \frac{r_s}{d} = \frac{\pi r_s d}{L}. \quad (9)$$

Similarly, as long as the geometric parameters are given, the shape factor can be obtained as well as the thermal resistance. In practice, LEDs are square in shape, and submounts may not be circular either. However, the method can be a well first-order approximation by considering the area of the LED and submount as circular in shape.

General Consideration of LED Lamp in Thermal Perspective

A LED lamp cooled by free convection usually contains several parts as shown in Fig. 3. The parts can be categorized into conduction section and convection section. Most parts belong to the conduction section except the finned heat sink is the only convection section in the entire structure. The heat source is the LED chip itself. The thermal energy is conducted through various parts and eventually dumped into the air by the finned heat sink. To ensure the LED lamp can be operated properly, the LED chip temperature should be lower than 120 °C. Therefore, the temperature difference between the LED junction temperature and the ambient air temperature should be about 95 °C if the ambient air temperature is 25 °C. The total thermal resistance of the lamp is the sum of the thermal resistances of each part in series. Therefore, the higher the LED power is, the smaller the total thermal resistance of the lamp is required. For example, if the LED chip generates thermal power of 5 W, the total thermal resistance of the lamp should be lower than 19 K/W according to Eq. 1. As described above, thermal resistance of conduction is determined by the thermal conductivity and the geometry of the material used. Therefore, thermal resistance of

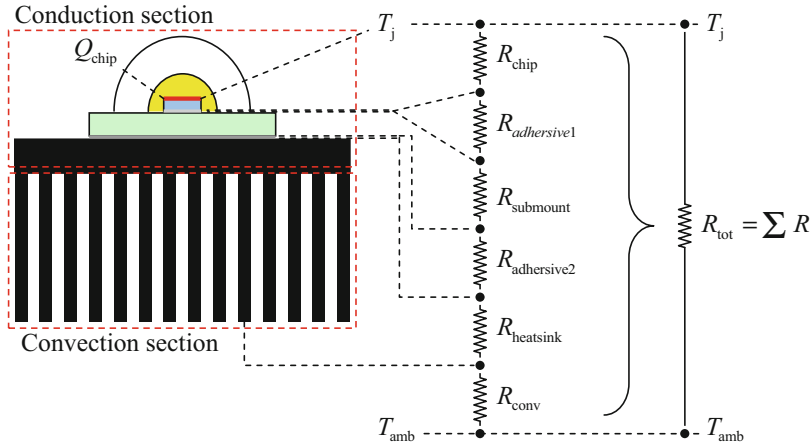


Fig. 3 A typical LED lamp contains three major parts, LED chip, submount, and heat sink. Adhesives are used to achieve good thermal contact between different parts. Each part or adhesive contributes thermal resistance which is labeled as R with subscript of each corresponding part. Convection also gives a thermal resistance since the heat sink temperature is always higher than the ambient temperature

the conduction parts can be reduced by using material with higher thermal conductivity or improving the geometry of the material. The geometrical way of reducing thermal resistance includes enlarging the heat exchange area and reducing the conduction path as Eq. 2 implies. For convection part, increasing the heat transfer coefficient and enlarging the heat exchange area are the two major ways to reduce thermal resistance of the convection part as Eq. 3 implies.

Due to practical specific requirement of LED lamp such as circuitry, handling, or thermal concerns, more parts may be added in an LED lamp; however, similar concept can be applied.

Heat Sources Within a LED Lamp

LED chip is the major heat source in an LED lamp. Usually, more than half of the input power to an LED becomes heat and distributed within the LED chip. Typically, the thermal management of LED usually refers to manage the heat generation in the LED chip. In order to cover wider spectral range without adding more LEDs, phosphors are commonly used in LED package with simple geometry. Such LEDs are known as phosphor-converted LED (pc-LED). For pc-LED which generates white light is known as phosphor-converted white LED (pc-WLED). Energy conservation and imperfection of the phosphor conversion efficiency imply phosphor also can generate heat. Therefore, LED chip and phosphor are the two heat sources to be discussed.

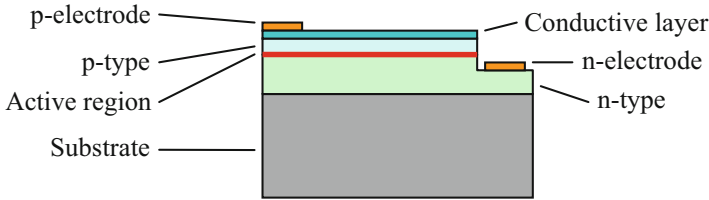


Fig. 4 Basic epitaxial structure of an LED

LED Chip

A simplified LED epitaxial structure is shown as in Fig. 4. The semiconductor materials grow on a substrate which matches the lattice of the material grown on it (Schubert 2006). LED is essentially a p–n junction diode formed by n-type and p-type semiconductors. The active region or the depletion region is located in between the n-type and p-type semiconductors and is where the electrons and holes recombine and generate light. A conductive layer may be added to make the current distribution over the active region more uniform. P-electrode is placed on top of the conductive layer, and n-electrode is directly contacted with the n-type material. Typically, the total thickness of the semiconductor layers is about 1 μm . The thickness of the substrate is usually larger than 100 μm . The thermal energy is generated in each layer which conducts electrons and holes. Similar to any other semiconductor devices, heat generation within LED chip is inevitable. However, the light generation nature of LED results in less heat generation comparing with other semiconductor devices. Non-radiative recombination of carriers and ohm heating are the two mechanisms responsible for heat generation within LED chip. From the microscopic point of view, the heat generation distribution is coarsely matching with the current distribution within the LED chip. A well-designed LED chip should have more or less uniform current distribution to ensure uniform light emission and less concern about current crowding (Schubert 2006). Semiconductors are generally having good thermal conductivity. Therefore, the heat generation within an LED can be considered as a volume heat source within the thickness of the epitaxy layers. The substrate usually does not contribute to the conduction of the current nor does the light emission. Therefore, the substrate contributes to extra thermal resistance when LED operates. Several methods have been developed to reduce the influence of the thermal resistance of the substrate such as thin-GaN, flip-chip, and using substrate with higher thermal conductivity.

In general, the injected energy is either generating light or heat for LED chip. Therefore, LED power efficiency can be written as

$$\eta_{\text{LED}}(T_j, j) + \eta_{\text{heat}}(T_j, j) = 1, \quad (10)$$

where η_{LED} is the light efficiency, η_{heat} is the heat efficiency, T_j is the junction temperature, and j is the injection current density of the LED chip. Both higher

Table 2 Efficacy and light efficiency of commercially available visible LED (Schubert 2006)

Color	Wavelength range (nm)	Material	Typical efficacy (lm/W)	Typical efficiency (W/W)
Red	620–645	AlGaAs GaAsP AlGaInP GaP	72	0.39
Red-orange	610–620	GaAsP AlGaInP GaP	98	0.29
Green	520–550	GaP AlGaInP AlGaP InGaN GaN	93	0.15
Cyan	490–520	InGaN	75	0.25
Blue	450–490	ZnSe InGaN	37	0.35

junction temperature and higher injection current density lead to higher heat efficiency regardless of the LED material and emission wavelength.

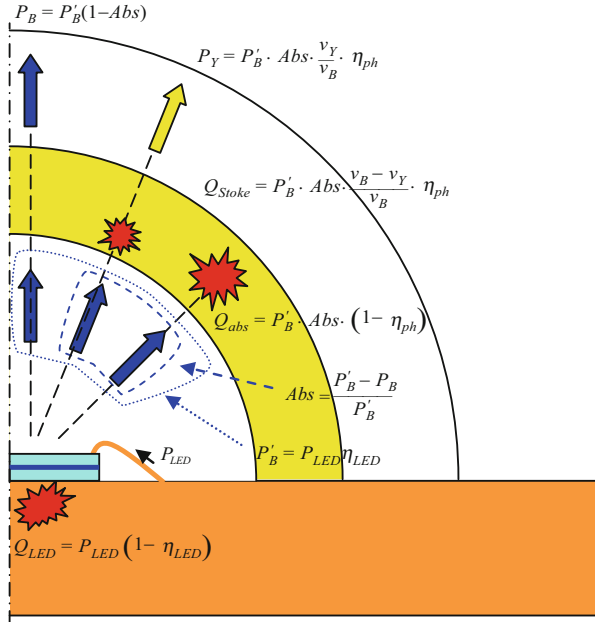
Typical LEDs emitting visible light have optical power efficiency and are listed in Table 2. The optical power efficiency is mainly limited by the material properties and epitaxy quality. Although blue LED has relatively good efficiency, more than half of the incident power is still converted into heat.

Comparing with other electronic devices or other thermal systems, LED chip does not generate large amount of thermal energy as others thanks to the light emission that carries away considerable energy. However, the small area of the LED chip makes the heat flux reach about 2×10^6 W/m² which is comparable with the heat flux of the heat shield of a spacecraft performing atmospheric entry (Regan 1993). Nevertheless, the highest LED operation junction temperature is about 120 °C which is only about 90–100 °C higher than the heat reservoir or the environment. With such temperature, thermal radiation does not help much on heat transfer. Therefore, the thermal management of LED mainly relies on conduction and convection. Since LED is considered to be an energy-saving device, passive cooling is preferred. Passive cooling only allows conduction and free convection which is less efficient compared with forced convection (Incropera et al. 2007). Consequently, thermal management becomes an essential issue for practical applications of LED.

Heat Generation Within Phosphor

Ideal phosphor absorbs one higher energy photon and reemits one and only one lower energy photon. Therefore, the energy difference between the photons turns into heat. This process is known as the down conversion, the Stoke shift, or quantum defect (Demtröder 2002). However, some absorbed photon may be directly

Fig. 5 The power conversion in a phosphor-converted pc-WLED



transferred its energy to thermal energy without emitting another photon. Such process originated from the collision of the atoms within the phosphor and result in non-radiation transition. The probability of an absorbed high energy photon can be converted into lower energy photon and is referred as the conversion quantum efficiency, η_{ph} , which is ranging from 0 to 1. Therefore, phosphor serves as another heat source in a pc-WLED due to the Stoke shift and the imperfection of the conversion quantum efficiency.

Figure 5 shows the power conversion channels in a pc-WLED using blue LED chip and yellow phosphor. P_{LED} indicates the injected electric power to the LED chip. Other symbol P indicates optical power. The subscript B and Y indicate the blue and yellow light-related physical quantities. Q_{LED} is the thermal power generated by the LED chip. Q_{Stoke} and Q_{abs} are the thermal powers given by Stoke shift and the imperfection of the conversion quantum efficiency of the phosphor, respectively. Symbol ν indicates the photon frequency. Abs is the absorbance of the phosphor.

Many phosphors have been developed for LED applications. Most of the phosphors used for pc-LED has emission wavelength covering from red to green. Depending on applications and required spectral range, multiple phosphors may be mixed to be pumped by the LED emission.

The most commonly used pc-WLED utilizes InGaN-based blue LED chip to pump Ce^{3+} :YAG phosphor which generates yellow light emission. The blue light emission of InGaN LED can be around 450 nm. The average emission wavelength of Ce^{3+} :YAG can be calculated around 565 nm. Therefore, the Stoke shift converts

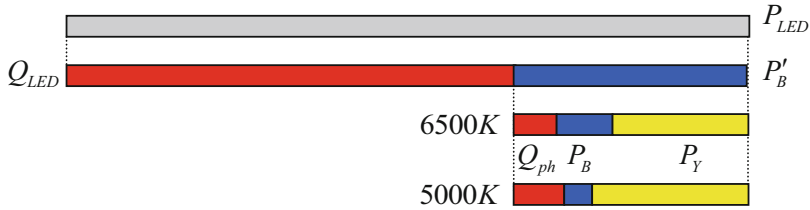


Fig. 6 Power budget of a phosphor-converted pc-WLED with color temperature of 6500 K and 5000 K

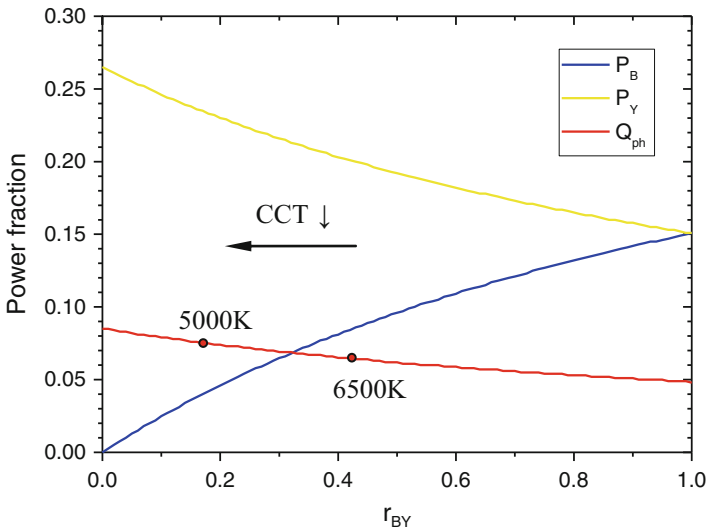


Fig. 7 The heat generated in phosphor, blue light output and yellow light output as function of the blue light and yellow light power ratio r_{BY}

about 20 % of the blue photon energy into heat. Ce^{3+} :YAG phosphor has conversion quantum efficiency η_{ph} around 0.90–0.95 nowadays. By adjusting the ratio of the residual blue light and emitted yellow light, the output white light with different correlated color temperature (CCT) can be achieved. For a pc-WLED with color temperature of 6500 K, the blue and yellow light power ratio is about 0.41. Assuming the Ce^{3+} :YAG phosphor has conversion quantum efficiency of 0.95, and the LED chip has blue light power efficiency of 0.35. The power budget of such LED is shown in Fig. 6. 65 % and 6.5 % of the input electric power are converted into heat within the LED chip and the phosphor, respectively. To reach lower CCT, more blue photons need to be converted into yellow photons and results in more heat as Fig. 7 shows. About 7.5 % of the input electric power will be converted into heat within the phosphor for a 5000 K pc-WLED.

The Temperature-Dependent Properties of LED

Many physical quantities of LED chip and phosphor are functions of temperature (Schubert 2006). Each temperature-dependent physical quantity can be measured relatively easy. However, the combined effect is extremely difficult to predict since these effect are cross-related nonlinearly. A complete and detailed model combining all temperature-dependent effects is currently unavailable. Therefore, a more practical way on operating an LED is to make sure the LED has a constant operation temperature which relies on good thermal management.

Temperature-Dependent Physical Quantities of LED Chip and Phosphor

The i - V relation of an LED can be described by Shockley equation with a series resistance

$$i = i_s \left[\exp\left(\frac{V - iR_s}{\eta kT}\right) - 1 \right], \quad (11)$$

where i_s is the saturation current, R_s is the series resistance, η is the ideality factor, k is the Boltzmann constant, and T is the LED junction temperature. If the injection current is a constant, the voltage drop will be proportional to the temperature of the LED. This property can be used to evaluate the junction temperature of LED. Such method is also known as the forward voltage method (Schubert 2006).

When the temperature of a semiconductor material increases, the lattice constant becomes larger due to larger vibration amplitude of the atoms and can be described by the linear expansion of the material. Therefore, the wavelength of the electron wave function becomes larger which is corresponding to smaller potential energy (Kittel 1996). Therefore, the effective bandgap energy becomes smaller. In other words, semiconductor bandgap energy is also a function of temperature which can be described by Varshni's empirical form (Schubert 2006) as

$$E_g(T) = E_g(0) - \frac{AT^2}{T + B}, \quad (12)$$

where A and B are the Varshni parameters listed as in Table 3.

The emission spectrum of an LED chip depends on the above parameters and can solve approximately. The emission wavelength corresponding to the bandgap energy can be written as

$$\lambda_g(T) = \frac{hc}{E_g(T)}, \quad (13)$$

Table 3 The Varshni parameters of selected semiconductor materials (Schubert 2006; Vurgaftman and Meyer 2003)

Material	$E_g(0)$ (eV)	A (meV/K)	$B(K)$	Remark
AlN	6.25	1.799	1462	Wurtzite
GaN	3.510	0.909	830	Wurtzite
InN	0.78	0.245	624	Wurtzite
AlN	5.4	0.593	600	Zincblende
GaN	3.299	0.593	600	Zincblende
InN	0.78	0.245	624	Zincblende
AlAs	2.239	0.6	408	
GaP	2.338	0.5771	372	
GaAs	1.519	0.5405	204	
GaSb	0.810	0.378	94	
InP	1.421	0.363	162	
InAs	0.420	0.25	75	
InSb	0.236	0.299	140	
Si	1.17	0.437	636	
Ge	0.66	0.477	235	

where h is the Planck’s constant and c is speed of light. This wavelength is the longest possible emission wavelength of the semiconductor used. The peak emission wavelength can be written as (Schubert 2006)

$$\lambda_p(T) = \frac{hc}{E_g(T) - \frac{1}{2}kT}, \tag{14}$$

where k is Boltzmann constant. The emission spectral width can be written as (Schubert 2006)

$$\Delta\lambda(T) \approx \frac{1.8kT\lambda_g^2}{hc}. \tag{15}$$

Figure 8 shows the longest possible emission wavelength, peak emission wavelength, and emission spectral width of GaN. Clearly, the entire emission spectrum has a redshift, and the spectral width increases as the temperature increases. By taking the derivate of Eq. 14, the temperature coefficient of spectral shift can be obtained. GaN has about 0.065 nm/°C at 100 °C. In comparison, GaAs has this temperature coefficient of spectral shift about 0.34 nm/°C at 100 °C. Since the emission spectra of semiconductor material change as temperature changes, CCT, CRI, and other spectral-related performance of LED are all functions of temperature. Since the spectral shift is temperature dependent, the peak wavelength is proposed to measure the junction temperature (Keppens et al. 2010).

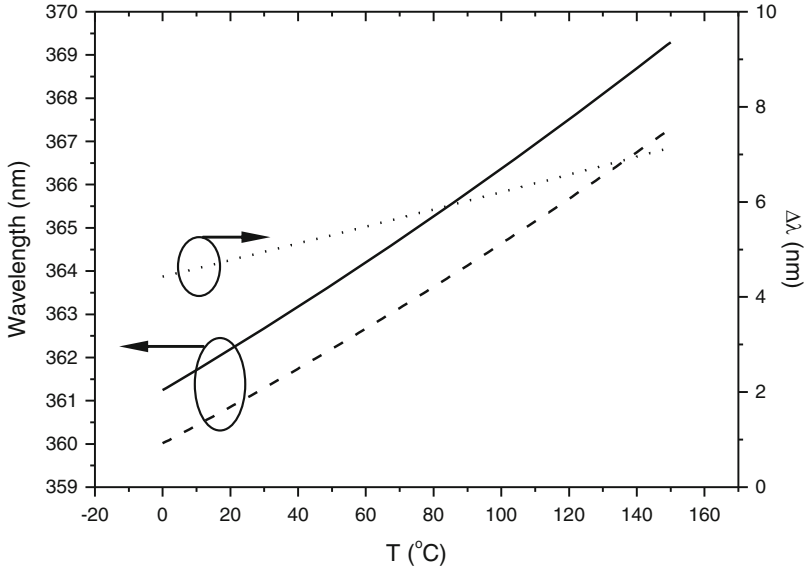


Fig. 8 The spectral parameters of GaN as functions of temperature. The *dashed curve* indicates the longest possible emission wavelength which corresponds to the bandgap energy. The *solid curve* indicates the peak emission wavelength. The *dotted curve* indicates the emission spectral width

The probability of non-radiative electron–hole recombination in the semiconductor becomes larger as the temperature increases. Namely, the internal quantum efficiency is smaller at higher junction temperature and results in lower light output. Therefore, as the LED turns on, the temperature of the LED will increase and leads to the reduction of the output optical power which is also known as the light decay. If the thermal management of LED is not properly done, light decay will be more significant. This effect is more significant in AlGaInP/GaAs than GaInN/GaN (Schubert 2006). High operation temperature will also seriously degrade the lifetime of the LED (Narendran and Gu 2005; Narendran et al. 2004).

Temperature-Related Phosphor Properties

Phosphor conversion quantum efficiency is also a function of temperature. As the phosphor temperature increases, the conversion quantum efficiency also decreases. This is because higher temperature leads to more frequent collision between the atoms in the phosphor and transfer of energy which leads to the increase of non-radiative transition probability (Demtröder 2002; Burshtein 2009). The collision also increases the uncertainty of the outer shell electron energy level which makes the emission spectrum of the phosphor becomes slightly wider accompanied with slightly redshift of the peak emission wavelength. Practically speaking, when phosphor temperature increases, the total output power of phosphor decreases and has its

output spectrum slightly deformed which results in the increase of the CCT even if the LED chip output remains unchanged. Since the conversion quantum efficiency of the phosphor decreases as temperature increases, more heat is generated within the phosphor region of the LED. More heat generation leads to even higher temperature in the phosphor region. Therefore, the CCT of a phosphor-converted LED is different under different driving currents. Under high-current driving, the CCT of a phosphor-converted LED will also increase till it reaches thermal equilibrium. Also, the total light power output will decrease at the same time.

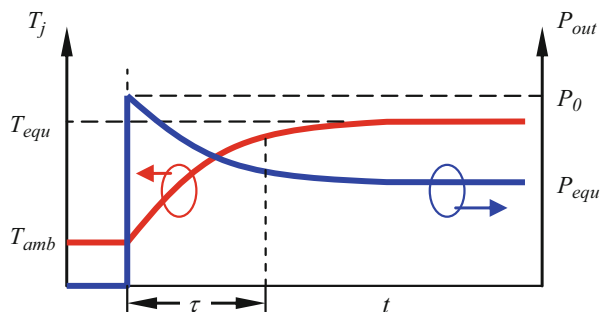
Silicone

Silicone is the most common material which can be molded into the optics for LED luminaries since silicone can sustain higher temperature than other common transparent organic polymers such as polycarbonate and acrylic. Silicone has maximum stable temperature of about 250 °C which is roughly 50 °C higher than epoxy (Anderson 2011; Petrie 2007). Therefore, silicone is more preferred than epoxy and other organic polymers. When the temperature of silicone exceeds 300 °C, silicone resin starts cracking. When cracking happens, the optical behavior of the luminaries will change and reduce the light output. Also, water and other organic compounds may reach the chip and reduce the lifetime of LED chip.

Transient Behavior of LED and Light Decay

When an LED turns on, the temperature of the LED chip takes some time to reach the equilibrium temperature. In other words, the LED chip temperature is a function of time. Therefore, the abovementioned temperature-dependent properties become a function of time. The most noticeable behavior is known as the light decay which describes the light output is higher at the instant when an LED is switched on comparing with the light output after the LED reaching thermal equilibrium as shown in Fig. 9. Light decay is usually defined as $(P_0 - P_{equ})/P_0$. At the instant when LED turns on, the LED junction temperature is equal to the ambient

Fig. 9 Temperature and light output as a function of time as a LED lamp turns on



temperature; therefore, the light output power is the highest. The LED junction temperature eventually reaches T_{equ} , the thermal equilibrium temperature. In the mean time, the light output also reaches equilibrium light output or P_{equ} . Clearly, light decay is larger if $T_{\text{equ}} - T_{\text{amb}}$ is larger. By definition, if the LED lamp has smaller total thermal resistance, the light decay will be smaller. Since light decay depends on the temperature difference, the LED junction temperature can be estimated accordingly. Besides the light output power, the CCT of the LED also will change with time due to the same reason. A way to avoid the observable light decay is to utilize time-dependent driving current based on the thermal behavior of the LED lamp.

The time constant τ of the transient behavior depends on the cross-interaction of the heat capacities and thermal resistance of different parts in the LED lamp. In general, larger heat capacitance gives longer time constant. In practice, the temporal behavior of LED lamp could be very complicated due to the cross interaction between many temperature-dependent parameters in an LED. For pc-WLED, the time constant of the phosphor region could be very different from the LED chip which leads to even more complicated behavior.

Thermal Management of LED

LED is considered to be an energy-saving device. Therefore, passive cooling is preferred, and the waste heat is dump into the air serving as the heat reservoir. The purpose of LED thermal management is to keep the LED chip temperature less than 120 °C, while it operates at its full power. The thermal management of LED mostly relies on conduction that spreads the thermal energy into larger area before the final stage of the thermal management which is to dump the waste heat into air by convection. Since passive cooling is preferred, effective free convection cooling should be achieved. To reach low thermal resistance of conduction, high thermal conductivity materials should be used. From the geometry point of view, the conduction path should be short, and the isotherm area should be large within each piece of material used. On the convection part of the cooling, the only practical rule is to enlarge the surface area contacting the cool air which can flow easily. A LED lamp contains three major parts, LED chip, submount, and heat sink. Certain adhesives have to be applied to achieve good thermal contact between the parts. Such configuration has been shown in Fig. 3. Each part or adhesive contributes a certain thermal resistance. Under thermal equivalent condition, the junction temperature of the LED chip can be obtained as

$$\begin{aligned} T_j &= T_{\text{amb}} + Q_{\text{chip}} \cdot \sum R \\ &= T_{\text{amb}} + Q_{\text{chip}} \cdot (R_{\text{chip}} + R_{\text{adhesive1}} + R_{\text{submount}} + R_{\text{adhesive2}} + R_{\text{heatsink}} + R_{\text{conv}}), \end{aligned} \quad (16)$$

where T_{amb} is the temperature of the ambient convection fluid, Q_{chip} is the thermal power generated by the chip, and R_x are the thermal resistance of each part in the LED lamp. In the following paragraph, conductive and convection part will be discussed separately.

Conductive Part

Material thermal conductivity and geometry are the two factors to influence the conductive thermal resistance as discussed above. Material used should have as large thermal conductivity as possible. However, due to other practical reasons, the choice of material is usually very limited. Contact thermal resistance should also be reduced as much as possible. Good geometry of the conductive parts has a simple rule which is to enlarge the isotherm surface area within each single material used.

LED Chip

Once the LED material and emission wavelength are determined, nothing much about the thermal property of the chip itself can be done. Fortunately, most semiconductor materials have good thermal conductivities since good crystal structure is also a required property to achieve good light-emitting efficiency. The free electrons and holes offer good electric conductivity as well as thermal conductivity which will be addressed more in the following paragraph. Moreover, the thicknesses of the epitaxy layers is very thin which makes the thermal resistance of the epitaxy layers almost negligible. However, the thermal resistance of the substrates may not be neglected since their thermal conductivity is usually higher, and thickness could be 100 times more than the epitaxy layers. Therefore, several methods have been utilized to reduce the thermal resistance of the LED chip itself. Taking GaN-based LED as an example, sapphire is the most common substrate. However, sapphire has thermal conductivity of about 33–35 W/m·K which is about one order of magnitude smaller than GaN which has thermal conductivity of 230 W/m·K. Thin-GaN and flip-chip are the most common solutions which effectively reduce the distance between the junction and submount (Schubert 2006). High thermal conductivity substrates such as GaN, SiC, and Si are also proposed and utilized to reduce the thermal resistance of LED chips (Nakamura et al. 1998; Lin et al. 1993; Tran et al. 1999).

Phosphor

Phosphors are usually in powder form. Phosphor powders are commonly mixed with transparent silicone resin and dispensed on top of the LED chip. Transparent silicone resin usually has thermal conductivity below 0.2 W/m·K which serves well for thermal isolator. With the powder dispersing within the resin, the generated heat within the phosphor powder cannot be removed efficiently. There are two main cooling paths for phosphor region. One is conducting the waste heat through the LED chip and submount. The other is conducting through the silicone exterior surface and cooled by air convection. However, these cooling mechanisms are not very efficient. Therefore, the phosphor temperature can easily reach 100 °C or more than the LED chip for embedded remote phosphor package configuration. Among various phosphor region geometries, conformal-coated phosphor has the lowest temperature from preliminary study (Chung et al. 2014). Silicate glass usually has thermal conductivity about 1 W/m·K. As phosphor glass has been proposed (Lee et al. 2012; Fujita et al. 2008), the high-temperature issue of the phosphor region may be able to have an alternative solution.

Table 4 Some common ceramic materials and the corresponding thermal conductivities and thermal expansion coefficients (Touloukian et al. 1979; Miyashiro et al. 1990; Akishin et al. 2009)

Material	Thermal conductivity (W/m·K)	Thermal expansion coefficient (K ⁻¹) @ 293 K
BeO	250–300	5.5×10^{-6}
AlN	70–270	4.6×10^{-6}
SiC	60–270	3.3×10^{-6}
Al ₂ O ₃	20–32	5.5×10^{-6}

Submount

Since LED chips are electronic devices, the submount for packaging LED chips needs to consider how to wire bond the electrodes and how to achieve electric insulation between electrodes. From the thermal management point of view, the submount material needs to be thermally conductive. Traditional FR4 PCB (printed circuit board) can serve the electronic issue well. However, the poor thermal conductivity of the FR4 PCB can only provide very limited thermal management capability. MCPCB (metal core-printed circuit board) utilizes aluminum alloy as the substrate. The aluminum alloy has much better thermal conductivity comparing with FR4. However, MCPCB requires a layer of polymer with thickness about 100 μm to achieve electric insulation between the aluminum alloy substrate and the laminated copper circuit tracks and pads. The polymer layer usually has very poor thermal conductivity below 1 W/m·K. Therefore, the MCPCB submount still can't fully benefit from the high thermal conductivity of the aluminum alloy. Another solution is to use ceramic with good thermal conductivity as the substrate material of the PCB. Ceramics are made by sintering process which fuses the powder of the raw materials together under heat and pressure. Voids among the fused particles make the ceramics porous and gives the diffusive look of the ceramic. The effective thermal conductivity of the ceramics also depends on the density of the voids and the quality of the fused quality of the powder particles. Therefore, the thermal conductivity of ceramic varies over a wide range depending on the sintering condition. Table 4 shows four different ceramic materials which have been used as thermally conductive but electrically insulated substrate materials for LED. When bonding the LED on the ceramic, one or more layers of metal may be applied to “wet” the surface of the ceramic (Peteves 1996). The wetting metal layer gives extra contact thermal resistance and increases the cost of such substrate. Metal layer usually has larger thermal expansion coefficient than the ceramic material and the metal layer bonding with the ceramic require higher temperature. For example, copper has about three times larger thermal expansion coefficients than these ceramic materials. Therefore, when temperature changes, the thermal expansion coefficient mismatch between the metal layer and ceramic substrate may cause the bending of the entire substrate. This issue can be reduced by laminating another metal layer on the other side of the substrate to balance the stress. Beryllium oxide (BeO) has the highest thermal conductivity than any other nonmetal except diamond. However, the pathogenicity of Beryllium oxide (BeO) dust is a major drawback for practical applications.

Multiple Chips Package

Commercially available single LED chip can be driven up to 10 W nowadays. To achieve even higher optical power output, multiple chips can be packed on one common submount. Chip-on-board (COB) package directly bonds multiple chips on one common circuit board. COB package allows large optical output in a relatively smaller area. Therefore, a larger dimension, less precise optics can be applied directly on top of the COB light source instead of smaller optics on each LED. However, thermal management of COB requires more concern. By choosing MCPCB or ceramic substrates, thermal resistance from the chip to heat sink can be reduced. However, the detail calculation of the thermal resistance of COB package cannot be easily obtained since each LED chip may not share identical effective substrate area. Therefore, different LED chips may have different thermal resistances which results in different temperatures. Since LED performance is temperature dependent, the output power, spectrum, CCT, and driving voltage of each LED on COB package may be all different.

Adhesive

With different parts of different materials in a complete LED lamp, proper adhesive should be used between parts to ensure proper thermal contact with physical strength. The adhesives usually serve thermal, mechanical, and electrical purposes at the same time. From the thermal point of view, the adhesive should always have good thermal conductivity to reduce the contact surface thermal resistance. From the mechanical point of view, the adhesive has to offer sufficient mechanical bonding strength to ensure the parts are properly connected and can withstand shear stress (Chung et al. 2012). From electrical point of view, the adhesive may be electrically conductive or insulated depending on the location.

To bond LED chip with the submount, electrically conductive adhesive should be used. Eutectic bonding such as Au–Sn provides thin ($<10\ \mu\text{m}$) and great thermal and electric conductivity along with good mechanical strength (Chung et al. 2014). However, eutectic bonding usually requires complicated process and precious metals. Solder is less expensive yet provides good thermal conductivity exceeding $50\ \text{W/m}\cdot\text{K}$ in general. However, the thickness of the solder is usually about tens of micron which gives additional thermal resistance. Also, residual flux for soldering may cause blackening of the silicone and reduce the lifetime of LED (CREE 2014).

Thermally conductive adhesives or thermal grease using polymerizable liquid matrix mixing with high thermal conductivity particles, aka the fillers, can be electrically insulated or conductive depending on the filler properties (Petrie 2007). The higher fraction of the filler is in the adhesive, the higher the thermal conductivity of the adhesive will be. The liquid matrix can be silicones, epoxies, urethanes, and acrylates. Electrically conductive adhesive uses metal fillers such as silver or aluminum particles. The conductive silver paste is easy to use and can have thermal conductivity exceeds $20\ \text{W/m}\cdot\text{K}$. Electrically insulated adhesives use oxide, nitrite, graphite, or diamond particle as the filler. The thermal conductivity of such electrically insulated adhesives

can range from 0.5 to 10 W/m·K. However, the adhesive layer thickness is usually above 100 μm which gives nonnegligible thermal resistance.

Heat Pipe and Vapor Chamber

Heat pipe and vapor chamber are relatively new concept for thermal management (Faghri 1995; Reay et al. 2006). Strictly speaking, they should not be considered as conductive materials. The basic concept of heat pipe and vapor chamber is using the heat source to heat up and vaporize liquid enclosed in a low-pressure container made of copper or high thermal-conductive material. Since vapor can move rather freely, hot vapor can quickly reach cooler place and condense back to liquid. If the condensed liquid is guided back to the hot spot, a complete cycle can be achieved. The thermal energy is efficiently moved away from the heat source. The shape of the container confined the heat transfer direction. Heat pipe is for 1D case, and vapor chamber is for 2D and 3D cases. A heat pipe/vapor chamber can be equivalent as a piece of material with extremely high thermal conductivity even though its structure is rather complicated. The effective thermal conductivity of a heat pipe depends on temperature, geometry, gravity, and many physical properties and operation conditions; therefore, it's difficult to give a detail and precise model of a heat pipe. To make sure a heat pipe is working, the heat source temperature should exceed the boiling point of the liquid within the heat pipe. As long as the liquid within the heat pipe is not totally vaporized, the heat pipe can work properly. The heat source should directly contact the heat pipe, and the far end of the heat pipe should be attached to a heat sink serving as the cold end for condensation. Since the effective thermal conductivity of a heat pipe is so high, the surface of a good working heat pipe can be considered as an isotherm surface.

Convection Part

In general, a passively cooled LED lamp requires a heat sink to dump the waste heat to the air. A finned heat sink is the most common solution which relies on free convection cooling. As discussed earlier, to achieve better convection cooling, large heat transfer coefficient and large heat sink surface area are required.

Finned Heat Sink

The most cost-effective heat exchangers between solid and air are finned heat sinks. To reach the best efficient heat exchange, the surface of the heat sink should be the isotherm surface. In other words, the thermal conductivity of the finned heat sink should be highly thermal-conductive material. In general, aluminum and copper are the most common materials used. Several methods can be used to make finned heat sink including extruding, die casting, forging, machining, and bonding. Machining is only suitable for small quantity or prototype. The performance of bonded fin depends highly on the quality of the process since it introduces additional contact thermal resistance. Die casting, extruding, and forging are all suitable for mass production. During the cooling process of die casting, the material gradually reduces

the volume and creates voids which lead to the reduction of thermal conductivity. However, die casting provides great flexibility on the design of the fin structure. Extrusion has relatively low manufacture cost and provides good thermal conductivity of finned heat sinks. Yet, the nature of extrusion limits the design of the fin structure. Forging gives better thermal conductivity among these three methods thanks to the reduction of number and volume of voids within the material during the forging process. Forging also provides better flexibility on the design of the fin structure. The only concern is the manufacture price is higher than the extrusion.

The discussion about geometry design of the fin structure can be found in heat transfer textbooks (Incropera et al. 2007; Spalding 1963). To achieve better convection heat exchange or smaller thermal resistance, larger heat exchange area and larger flow speed are preferred. Larger heat exchange area indicates the thermal resistance is smaller which can be achieved by increasing the fin number and reducing the thickness of each fin. To increase the flow speed essentially increases the heat transfer coefficient. Since surface friction of the fin will reduce the flow speed. Large separation of fin is preferred. However, increase fin number comes with the reduction of the cross section of the flow path in between neighboring fins. As the air flow resistance increases, the flow speed is reduced. Consequently, an optimized fin design may be found. Unfortunately, the analysis is way beyond the scope of this chapter. Experiment is probably a faster way to find the optimum fin design in general. In general, the open gap between neighboring fins should be around 5–8 mm calculated by the equation given in (Spalding 1963). Since free convection relies on the buoyancy of the hot air, fins should be aligned vertically which the air flow will follow. Otherwise, the air resistance or the drag will reduce the flow speed and reduce the heat transfer coefficient significantly.

Other Surfaces That Contact Air

Besides the finned heat sink, all the surfaces of an LED that exposes to the ambient air do exchange heat via convection. Again, heat transfer coefficient should be used on each surface. A coarse yet acceptable way to estimate the heat transfer on these surfaces is directly apply free convection heat transfer coefficient on all the exposed surface area of the lamp.

Estimation of Thermal Resistance of an LED Lamp

As discussed in previous paragraphs, LED performance is highly depending on the junction temperature. Therefore, proper thermal management should be applied to ensure the temperature change of LED will not increase dramatically when turn on. Therefore, the total thermal resistance of the lamp should be better to be small. The total thermal resistance of an LED lamp is the sum of the thermal resistance of each component. The junction temperature of an LED should be no more than 120 °C. If the air temperature is 25 °C, the relation between total system thermal resistance and the LED thermal power can be written as

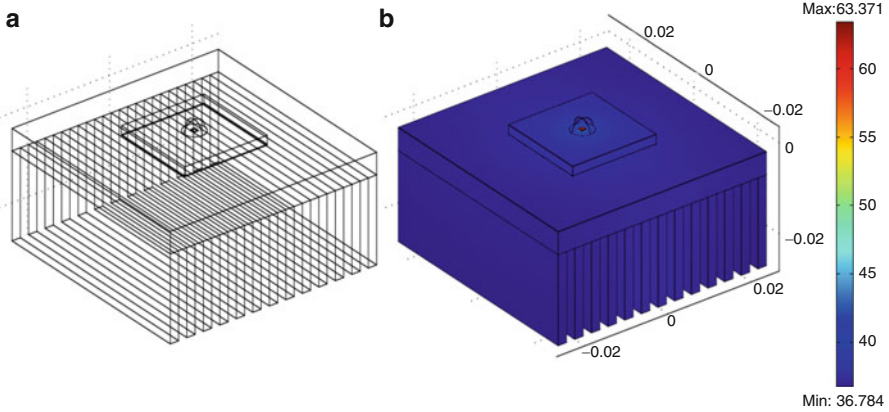


Fig. 10 (a) is the geometry of an LED lamp for simulation. (b) is the FEA simulation result of the temperature distribution of the LED lamp. The length is in the unit of meter and the temperature is in the unit of °C

$$R_{\text{tot}} = \frac{95}{P_{th}} \quad [K/W], \quad (17)$$

where R_{tot} is the total system thermal resistance and P_{th} is the thermal power of the LED in the unit of watt. Obviously, LED lamp with larger power requires smaller total system thermal resistance.

To obtain the total thermal resistance, the thermal resistance of each part in the lamp has to be obtained first. However, only 1D and a few highly symmetry 2D and 3D geometries have analytical solution of thermal resistance. In a typical LED lamp, the adhesive layers can be considered as 1D thermal-conductive problem. Therefore, Eq. 2 can be utilized to evaluate the corresponding thermal resistance. Taking a 70 μm thick silver paste with thermal conductivity of 13 W/m-K as an example, if the bonding area is $5 \times 5 \text{ mm}^2$, the corresponding thermal resistance is about 0.215 K/W.

For the object geometry departed from the cases with analytical solution of thermal resistance, several numerical methods can be applied to evaluate the thermal resistance of the object interested by solving the thermal equation. Finite element analysis (FEA) and finite difference are probably the most efficient and effective ways to evaluate the thermal resistance. Many commercially available softwares can serve this purpose well. As long as the geometry of the LED lamp design and the material used is given, the temperature distribution in any location within the LED lamp can be obtained. Figure 10 shows a FEA thermal simulation of an LED lamp. The LED chip dissipates 4 W thermal power over a $1 \times 1 \times 0.1 \text{ mm}^3$ GaAs chip. The chip is encapsulated in a silicone dorm with radius of 2.5 mm. The chip is bonded with a 6061 aluminum alloy submount by a solder layer with thickness of 20 μm and thermal conductivity of 10 W/m-K. The submount has an area of $20 \times 20 \text{ mm}^2$ and its thickness is 2 mm. A layer of adhesive with thickness of 0.2 mm and thermal conductivity of 10 W/m-K joins the submount and the heat sink.

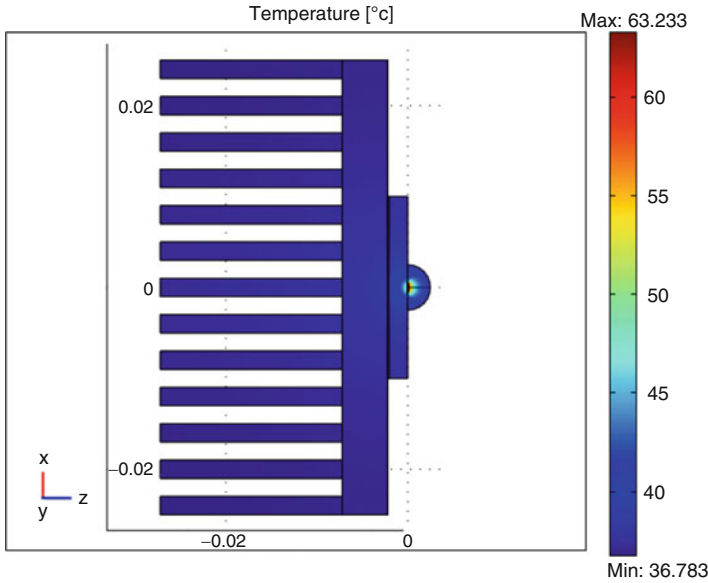


Fig. 11 Temperature distribution along the symmetry surface of the LED lamp

The heat transfer coefficient of all exterior surfaces is set to be 10 W/K·m². The finned heat sink has an area of 50 × 50 mm² and the base thickness of 5 mm. The fin length is 20 mm and the thickness is 2 mm. The spacing between the fins is also 2 mm. There are totally 13 fins. The geometry is of the simulated LED lamp, and the simulated temperature distribution is shown in Fig. 10a, b, respectively. Figure 11 shows the temperature distribution along one of the symmetry planes of the LED lamp. The maximum temperature is located at the chip surface contacting with the silicone. The enlarged temperature distribution and isotherm surfaces in the vicinity of the LED chip are shown in Fig. 12. The ambient temperature is set to be 25 °C. Therefore, the total thermal resistance of the LED lamp is

$$R_{\text{tot}} = \frac{63.371 - 25}{4} = 9.593 \text{ [K/W]}. \tag{18}$$

Several things can be noticed. The adhesive materials usually have about one order of magnitude lower thermal conductivity comparing with the submount and heat sink material. Therefore, the adhesive surfaces where have inward heat flux are almost isotherm surfaces as the arrows indicating in Fig. 12a, b. The isotherm surface away from the heat source in one medium is approximately a hemisphere as the long dashed curve in Fig. 12a.

Base on these observed facts in the LED lamp, one can perform a quick estimation of thermal resistance of the entire LED lamp by adding the thermal resistance of each part. Therefore, the thermal resistance of each part has to be evaluated. The following discussion provides a quick way to estimate the thermal resistance of each

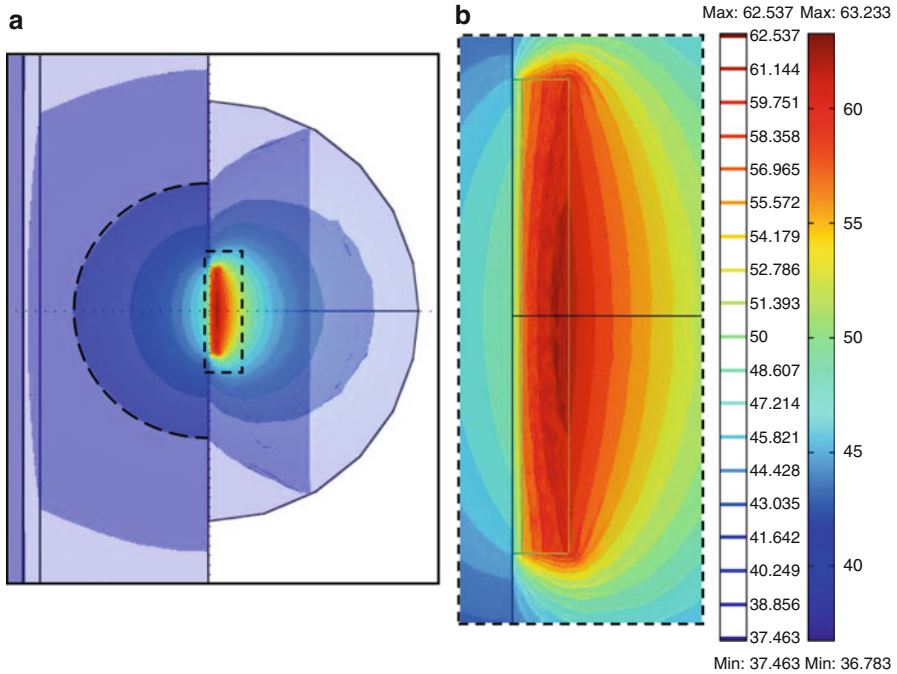


Fig. 12 Temperature distribution and the isotherm surfaces in the vicinity of the LED chip

part in an LED lamp. Several assumptions are needed to perform the analysis. (1) Adhesive material serves as 1D thermal-conductive material. (2) The surface which is away from the heat source of a high thermal conductivity material has constant temperature. (3) The thermal conductivity of the adhesive material is small compared with the contacted material.

The chip itself has thermal resistance about 1.068 K/W. The adhesive layers provide 1D thermal resistance. The bonding layer (adhesive 1) of the chip has thermal resistance of 2.0 K/W. The adhesive layer (adhesive 2) between the submount and heat sink has thermal resistance of 0.05 K/W. The shape factor can be used to evaluate the thermal resistance in the submount. The chip area is $1 \times 1 \text{ mm}^2$; therefore, the equivalent circular area has radius of 0.564 mm. The submount has area of $20 \times 20 \text{ mm}^2$. The equivalent circular area has radius of 11.284 mm. The thickness of the submount is 2 mm. In other words, r_s is 0.564 mm, d is 11.284 mm, and L is 2 mm. $L/d = 0.177$, $r_s/d = 0.050$,

$$\frac{S \cdot L}{A} \approx 0.00984 \quad (19)$$

Or

$$S \approx 0.00984 \frac{A}{L} = 0.00197 \quad [m] \quad (20)$$

Table 5 Simulated and estimated thermal resistance of the LED lamp in Fig. 10

Thermal resistance (K/W)	FEA simulation ^a	Estimation
Chip	1.068	1.068
Adhesive 1	1.908	2.0
Submount	2.880	3.043
Adhesive 2	0.363	0.05
Base of heat sink	0.313	0.0697
Convection	3.063	2.992
Total	9.593	9.850

^aThe FEA simulated thermal resistances are calculated by the temperatures along the central symmetry line of the LED lamp

The thermal conductivity of 6061 aluminum alloy is 167 W/m-K. Therefore, the submount has thermal resistance of the submount is 3.043 K/W. Similarly, the thermal resistance of the base of the heat sink is 0.0697 K/W. The surface area of the heat sink is about 0.0334 m². The thermal resistance of the convection is therefore 2.992 K/W.

By adding up all the thermal resistances, the total estimated thermal resistance is about 9.850 K/W which is about 2.6 % smaller than the FEA simulation result which is 9.593 K/W. The calculated and evaluated thermal resistance of each part is listed in Table 5. The calculated thermal resistance based on the FEA simulation is the temperature difference of each part along the center symmetry axis of the lamp. Please note that the estimated thermal resistances are based on several coarse assumptions. Also, thermal resistance has to be defined between isotherm surfaces which are not exactly matching the surface of the parts. In other words, the thermal resistance can't be properly defined for surfaces with non-uniform temperature distribution. Therefore, the discrepancy should be the result of the way of evaluating the thermal resistance.

This method provides a quick way to estimate the thermal resistance of each part with reasonable precision and can be applied to design the thermal management solution of LED lamps.

The contact surface resistance is not included in the simulation and estimation since the contact surface resistance is very difficult to estimate unless an experiment is performed. Since the contact thermal resistance only exists at the contacting surfaces, the area is the same as the smaller contacting surface. Therefore, contact thermal resistance can be equivalent as the adhesive material has even lower thermal conductivity. In this case, the highly conductive surfaces will be even closer to an isotherm surface.

For COB package, similar estimation can be performed. However, the submount is shared by multiple LEDs. Therefore, the effective area used by each LED should be considered when estimating the thermal resistance of the submount. Between the neighboring LEDs, the symmetry line can be considered as the boundary of the portion of the submount used for the neighboring LEDs. The thermal resistance of each small portion of the submount can be estimated. Please note that if the arrangement of the LEDs can't lead to equal area submount portions, non-uniform

temperature distribution can't be achieved. Consequently, the estimated thermal resistances will have larger error.

For a more complicated LED arrangement or lamp geometry, FEA simulation is probably a faster and the only way with good precision to estimate the thermal resistance and thermal performance of the LED lamp.

References

- Akishin GP et al (2009) Thermal conductivity of beryllium oxide ceramic. *Refract Ind Ceram* 50:465–468
- Anderson BJ (2011) Thermal stability of high temperature epoxy adhesives by thermogravimetric and adhesive strength measurements. *Polym Degrad Stab* 96:1874–1881
- Burshtein Z (2009) Radiative, nonradiative, and mixed-decay transitions of rare-earth ions in dielectric media. *Opt Eng* 49:091005
- Chung TY et al (2012) A study of large area die bonding materials and their corresponding mechanical and thermal properties. *Microelectron Rel* 52:872–877
- Chung T-Y et al. (2014) Study of the temperature distribution within the phosphor regions of white LEDs. In: Presented at the 14th international symposium on the science and technology of lighting, Spazio Como
- CREE (2014, June 5) CREE XLamp LEDs Chemical Compatibility. Support document CLD-AP63 REV
- Demtröder W (2002) *Laser spectroscopy: basic concepts and instrumentation*, 3rd edn. Springer, New York
- Faghri A (1995) *Heat pipe science and technology*. Taylor & Francis, Washington, DC
- Fujita S et al (2008) Luminescence characteristics of YAG glass-ceramic phosphor for white LED. *IEEE J Sel Top Quantum Electron* 14:1387–1391
- Huang K (1987) *Statistical mechanics*, 2nd edn. Wiley, New York
- Incropera FP et al (2007) *Introduction to heat transfer*, 5th edn. Wiley, Hoboken
- JEDEC Solid State Technology Association (2005) *Guidelines for reporting and using electronic package thermal information*
- Keppens A et al (2010) Modeling high power light-emitting diode spectra and their variation with junction temperature. *J Appl Phys* 108:043104
- Kim HH et al (2008) Thermal transient characteristics of die attach in high power LED PKG. *Microelectron Rel* 48:445–454
- Kittel C (1996) *Introduction to solid state physics*, 7th edn. Wiley, New York
- Lee YK et al (2012) Phosphor in glasses with Pb-free silicate glass powders as robust color-converting materials for white LED applications. *Opt Lett* 37:3276–3278
- Lin ME et al (1993) A comparative-study of gan epilayers grown on sapphire and sic substrates by plasma-assisted molecular-beam epitaxy. *Appl Phys Lett* 62:3479–3481
- Miyashiro F et al (1990) High thermal-conductivity aluminum nitride ceramic substrates and packages. *IEEE Trans Comp Hybrid Manufact Technol* 13:313–319
- Nakamura S et al (1998) InGaN/GaN/AlGaIn-based laser diodes with modulation-doped strained-layer superlattices grown on an epitaxially laterally overgrown GaN substrate. *Appl Phys Lett* 72:211–213
- Narendran N, Gu YM (2005) Life of LED-based white light sources. *J Disp Technol* 1:167–171
- Narendran N et al (2004) Solid-state lighting: failure analysis of white LEDs. *J Cryst Growth* 268:449–456
- Peteves SD (1996) Joining nitride ceramics. *Ceram Int* 22:527–533
- Petrie EM (2007) *Handbook of adhesives and sealants*, 2nd edn. McGraw-Hill, New York
- Reay DA et al (2006) *Heat pipes*, 5th edn. Butterworth-Heinemann, Oxford/Burlington

- Regan FJ (1993) Dynamics of atmospheric re-entry. AIAA education series, AIAA. ISBN 160086046X, 9781600860461
- Schroeder DV (2000) An introduction to thermal physics. Addison Wesley, San Francisco
- Schubert EF (2006) Light-emitting diodes, 2nd edn. Cambridge University Press, Cambridge
- Spalding DB (1963) Convective mass transfer, an introduction. McGraw-Hill, New York
- Szekely V (1997) A new evaluation method of thermal transient measurement results. *Microelectron Rel* 28:277–292
- Touloukian YS et al (1979) Master index to materials and properties. IFI/Plenum, New York
- Tran CA et al (1999) Growth of InGaN/GaN multiple-quantum-well blue light-emitting diodes on silicon by metalorganic vapor phase epitaxy. *Appl Phys Lett* 75:1494–1496
- Vurgaftman I, Meyer JR (2003) Band parameters for nitrogen-containing semiconductors. *J Appl Phys* 94:3675–3696

Optical Design: Chip and Packaging

Ching-Cherng Sun

Contents

Light Extraction Efficiency	270
Light Source Modeling	275
Optical Modeling for Phosphors	279
Chromatic Performance	281
CRI with CCT	281
Spatial CCT Uniformity	284
Design of Encapsulation Lens	285
References	287

Abstract

Light-emitting diode (LED) has been extensively applied to general lighting since the luminous efficacy exceeded 100 lm/W. In addition to high energy efficiency, the advantages of fast response, wide color range, narrow bandwidth, compact size, and environmental benefits extend the application of LED to display, communication, and others (Sun et al. 2004; Zukauskas et al. 2002; Schubert and Kim 2005). In regard to a light source with the use of an LED in lighting application, three properties related to the optical property should be addressed. The first is the optical efficiency, the second is the light pattern distribution, and the third is the color consistency. All these three properties can be determined in the chip level, the packaging level, and also the luminaire level. The chip level and the packaging level take the majority to determine the most properties of an LED luminaire. Among these properties, chip-level optics is especially important in intrinsic energy efficiency of an LED. In contrast, the packaging design is more important in color performance, especially for a phosphor-converted white LED

C.-C. Sun (✉)

Department of Optics and Photonics/Institute of Lighting and Display Sciences, National Central University, Jhongli, Taiwan

e-mail: ccsun@dop.ncu.edu.tw

(so-called pcW-LED). In addition, the optical pattern and energy efficiency are two important factors determined in this level.

In this chapter, the optics in the chip level and the packaging level for a pcW-LED will be introduced and discussed. We will start from light extraction of an LED die and then light source modeling, phosphor modeling, and packaging design to performing high-quality chromatic uniformity.

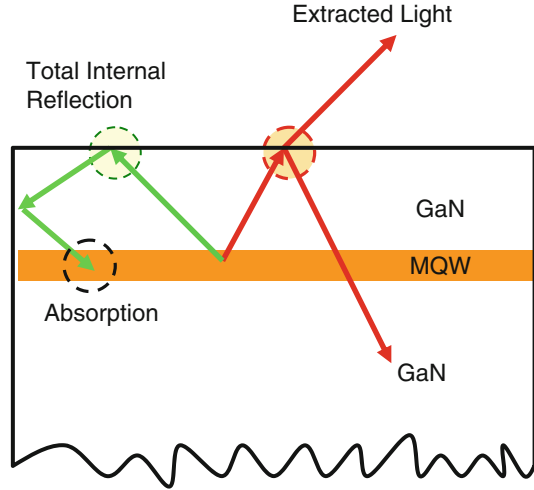
Light Extraction Efficiency

Among various types of LED, GaN-based LEDs attract the most attention for blue light emission, which is the base of a pcW-LED with the adding of yellow phosphors. The external quantum efficiency (EQE) of an LED is the best way to judge the energy efficiency of the device (Sun et al. 2011a; Krames et al. 2007). EQE is defined as the ratio between the photon number emitted by the device escaping from the die and the electron–hole pair injected into the p–n junction and is equal to a product of internal quantum efficiency (IQE) and light extraction efficiency (LEE). The IQE is defined as the ratio between the photon number emitted by the active layer and the electron–hole pair injected into the layer, and this efficiency relates to the substrate properties and epitaxy quality (Zukauskas et al. 2002; Lee et al. 2005; Steigerwald et al. 2002). The LEE is a ratio between the photon escaping from the die and the photon number emitted by the active layer, and this efficiency relates to chip processing, die geometry, and package. In the viewpoint of measurement, the EQE, rather than the IQE and the LEE, is a measurable factor, though there are several methods which are targeted to estimate the IQE and thus LEE. An alternative and effective way to estimate is to use Monte Carlo ray tracing to simulate the LEE (Sun et al. 2004; Lee et al. 2005; Ting and McGill 1995; Lee 2001; Badano and Kanicki 2001).

Generally, an LED die can be regarded as a rectangle cavity, where photons are emitted from the active layer and are incident on all exit surfaces to escape the cavity. However, if the cavity is a rectangle box, each photon could be incident on the exit surface at the same angle when it is multiply reflected in the cavity so that it has a small possibility to escape the cavity and could be absorbed finally. Therefore, an effective way for obtaining high LEE is to change the propagation direction of the photon when it hits some special layers and is incident on the exit surface at an angle within the escape cone, as shown in Fig. 1. Therefore, to change the photon angle with implanting microstructures such as pattern sapphire in a general sapphire-based GaN or surface texture in a thin GaN is an effective way. Besides, reducing the absorption coefficient of all media or reflectors is another important approach. In the packaging side, using an optical medium to encapsulate the LED die is a simple and effective way to increase light extraction efficiency.

Generally, a GaN-based LED die can be regarded as a rectangular box containing a thin layer in just few micrometers. Accordingly, one can simplify the geometry as a rectangular cavity. The most concerned part is that how to let the most photons emit

Fig. 1 Total internal reflection causes most lights trapped inside the cavity



by the active layer to go into the air. The major barrier is the total internal reflection (TIR) which traps most lights in the LED cavity. TIR is an effect for a light propagating through an interface from a medium of high refractive index (n_H) to that of low refractive index (n_L). The incident photons are completely reflected back when the incidence angle (θ_i) exceeds the critical angle (θ_c) written

$$\theta_i \geq \sin^{-1} \left(\frac{n_L}{n_H} \right) \equiv \theta_c. \quad (1)$$

The TIR effectively blocks some photons to escape from the die. A simple calculation using Eq. 1 shows that the critical angles are 27° and 36° for a medium of refractive index of 2.2 and 1.7 to the air, respectively. This means that the photons will not escape from the LED die whenever the incident angle on the surface to the air is limited to a small specific light cone, as shown in Fig. 1. The worst condition occurs when the LED die is similar to a rectangular cavity with six flat surfaces. Thus, the reflected photons have no chance to change the incident angles when they again hit the surface to the air, as shown in Fig. 2a. The photons then are reflected back again and again and become trapped photons, which finally could be absorbed by the active layer or others. A simple simulation shows that the LEE of a rectangular LED die is only around 23%. This is the major barrier in LEE for a GaN die. Figure 2b shows that the light distribution becomes different owing to implantation of microstructure in the substrate to change the light propagation direction so the LEE increases.

The geometry of GaN die could be classified into three major kinds according to specific die bonding technology, as illustrated in Fig. 3 (Wong et al. 1999; Gao et al. 2004). The first kind of geometry is called lateral wire bonding (LWB), where the two pads for n-type and p-type are located laterally and wire bonding for both n-type pad and p-type pad is required. The second type is called flip-chip bonding (FCB), where no wire bonding is needed because the two kinds of pads are faced

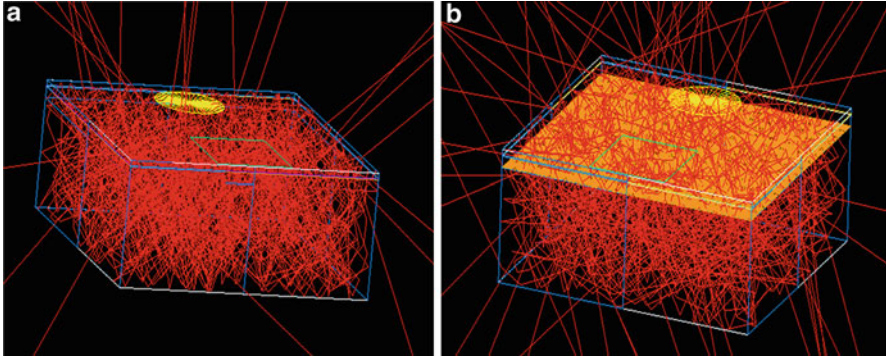


Fig. 2 An LED cavity (a) without and (b) with microstructure in the substrate

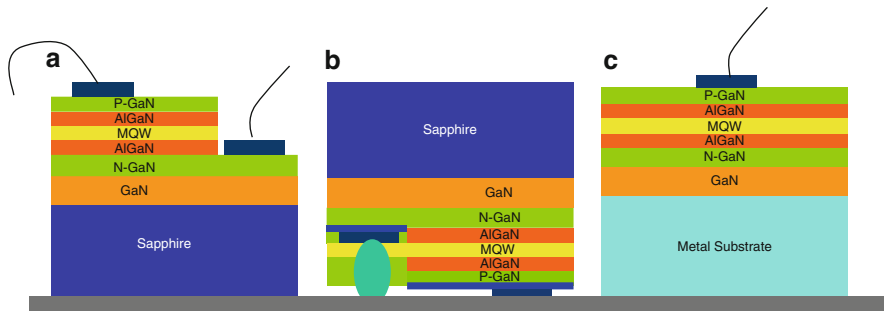


Fig. 3 The three kinds of packaging. (a) Lateral wire bonding, (b) flip-chip bonding, and (c) thin GaN

downward to connect the submount, and therefore the substrate, usually sapphire, for supporting the epitaxy is faced upward. The third is a vertical structure and is called thin GaN, where the substrate is removed through laser lift-off or other technologies. Before the lift-off process, the LED chip is bonded to a new support substrate to hold the epitaxy layer. The new support substrate should perform high electric conductivity, high thermal conductivity, high optical reflectivity, and similar thermal expansion coefficient as the epitaxy layer. Silicon is one of the medium for this purpose to meet all requirements.

To reduce the trapped photons is the main approach to increase the LEE (Sun et al. 2004; Zukauskas et al. 2002; Lee et al. 2005, 2007; Steigerwald et al. 2002; Gao et al. 2004; Kurata et al. 1981; Schnitzer et al. 1993; Lee and Song 1999; Krames et al. 1999; Baba et al. 1999; Windisch et al. 2000, 2001a, b; Linder et al. 2001; Guo et al. 2001; Wierer et al. 2001; Huh et al. 2003; Fujii et al. 2004). The trapping photons may be extracted through some useful approaches. Encapsulation with a lens on the LED die could be the simplest and the most effective way, as shown Fig. 4 (Sun et al. 2011a; Haerle et al. 2004). The encapsulation lens provides

Fig. 4 Lens encapsulation is useful to extract more lights from LED cavity

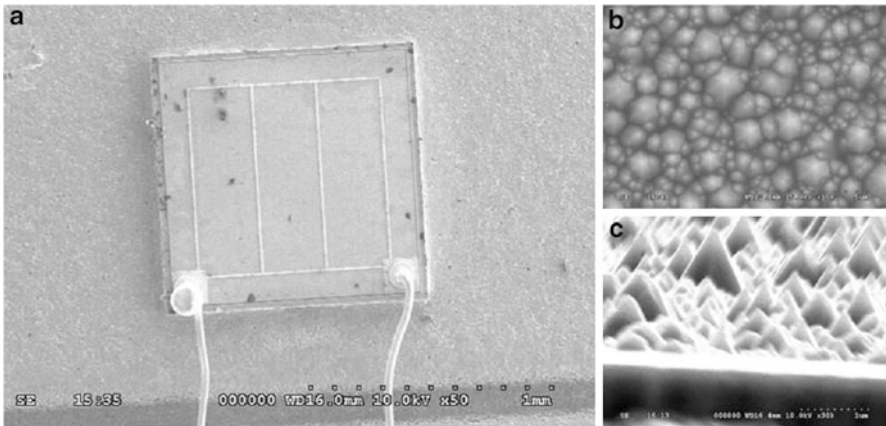
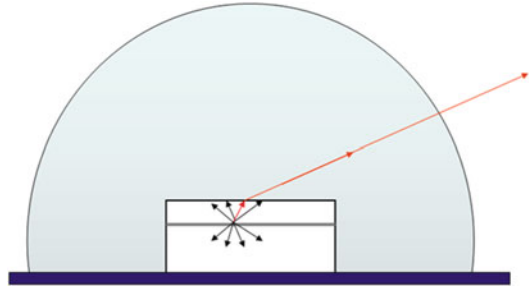


Fig. 5 (a) A photo of an LED die with surface texture. (b) The magnified *top view*. (c) The *side view* of the microstructure

more possibility for a photon to escape from the die by enlarging the critical angle because the photon goes to another medium with refractive index around 1.5 rather than 1 of the air. Usually the lens has a hemisphere shape so that the photons are easy to pass through the lens to the air. The second effective approach is to introduce the microstructure in the LED die. If the microstructure is on the top surface as shown in Fig. 5, it is called surface texture (ST) (Lee et al. 2007). If the structure is on the substrate, it is called patterned substrate (PS) or patterned sapphire substrate (PSS) (Lee et al. 2007). In this case, the microstructure is located in the interface between the substrate and the epitaxy layer. Both ST and PS structures can change the incident angle to the surface when the photons hit the surface for the second time or more. Once the incident angle falls into the light escaping cone, the photons can escape from the LED die effectively. The third is to shape the LED die to a form which is different from a rectangular box, such as a truncated (inverse) pyramid. The objective for this approach is similar to microstructure in the second approach. However, in most cases for die shaping, the die is thicker than the other types, so more lights escape from the four side faces and thus it causes wider angular distribution.

Table 1 An example of LED parameters for simulation

Layer	Bottom	Sapphire	GaN	AlGaIn	N-type	MQW	P-type	Silicone lens
Thickness (μm)	NA	330	2	0.05	2	0.1	0.05	10,0000
Reflectivity index/reflectivity	90 %	1.78	2.40	2.40	2.42	2.54	2.45	1.5

	Without Surface Texture	1D Stripe Pattern	2D Well Pattern	2D Pyramid Pattern	2D Lens Array
Without Silicone Lens	23.1%	35.9% (155%)	24.6% (116%)	45.5% (197%)	39.6% (171%)
With Silicone Lens	54.1%	64.5% (119%)	54.7% (101%)	69.6% (129%)	67.6% (125%)

Fig. 6 A comparison of the LEE enhancement for the different microstructures for surface texture, where the red number is the enhancement ratio

The simulation of the LEE can be done by the Monte Carlo ray tracing, which is one of the most suitable ways to simulate the light propagation in LED dies, but the dimensions of the microstructure should be larger than five times of the wavelength of the propagation photons (Sun et al. 2011a; Lee et al. 2005; Ting and McGill 1995; Lee 1998, 2001; Badano and Kanicki 2001; Joyce et al. 1974). The active layer in the LED die can be regarded a Lambertian surface. The emitted light is unpolarized. Table 1 shows an example for the structure parameters in the simulation of LEE for a GaN-based LED. The die size is $300 \times 300 \mu\text{m}^2$ and the absorption coefficient of active region is assumed to be 10^{-4}cm^{-1} (Muth et al. 1999; Djurić et al. 2001). For each ray, the trajectory and the energy are determined by Snell's law, Fresnel refraction, and material absorption. Figure 6 shows a comparison of the LEE enhancement for the different microstructures for surface texture. It illustrates that the pyramid array with 30° slanted surface is the best structure in increasing the LEE (Sun et al. 2011b). The effects of increasing the LEE by both encapsulating with/without a silicone lens and introducing a pattern array are summarized in Fig. 7, where the reflectivity of the bottom surface is set 90 % (Sun et al. 2011a; Guo et al. 1999; Windisch et al. 2002). Note that the absorption coefficient of the active layer in the simulation is set at 10^4cm^{-1} , but 200cm^{-1} is given more practically (Sun et al. 2011a, b). Therefore, the highest LEE could be higher than the values shown in Fig. 7. The simulation shows that both the encapsulation and the PSS with 30° pyramid array effectively increase the LEE. If both surface texture and PSS are

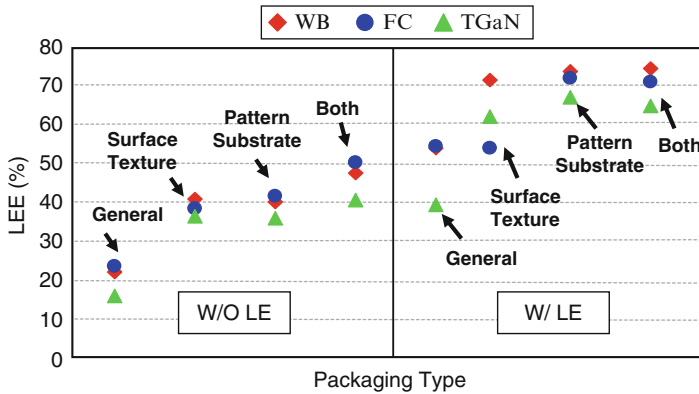


Fig. 7 Simulation of the LEE with several possible ways in to improve the LEE

applied, the LEE may be as large as to 350 %. The LEE enhancement approaches get different impacts in different packaging types. In LWB, both surface texture and PSS are very helpful in enhancing the LEE. In flip-chip packaging, surface texture is applied on the sapphire surface, and it makes less effect because the refractive index difference between the sapphire and silicone is small. In a thin GaN, there is no obvious difference between surface texture and PSS. If we look at the overall performance, PSS, surface texture, and lens encapsulation are all useful approaches in enhancing the LEE. The absorption of the active layer could be lighter in practical cases where the absorption coefficient could be as low as 200 cm^{-1} . In such a condition, surface texture or PSS with lens encapsulation could increase the LEE to as high as 90 % (Sun et al. 2011a, b).

Light Source Modeling

In contrast to other traditional light sources, LED acts as a new light source with multiple advantages. The major storage of LED light source is without standard model owing to different packaging designs, including single die, multiple dies, the cluster, and the RGB packaging. Besides, each packaging structure in an LED light source could suffer from color and spatial variation (Steigerwald et al. 2002; Sun et al. 2006; Nguyen et al. 2005). Such variations could become a major shortage in the practical application. Therefore, a useful modeling scheme to figure out the spatial distribution in optical field is important.

A useful modeling algorithm to model the optical behavior of LEDs, which is not only useful in LED lighting design but also in packaging, was proposed and demonstrated in 2006 (Sun et al. 2006). The modeling algorithm contains a simple measurement of LED geometry, Monte Carlo ray tracing, and a comparison between the simulation patterns and experimental measurements at various distances in the midfield region. The algorithm in modeling the light pattern of LEDs starts at finding

or estimating some important parameters for Monte Carlo simulation. The parameters include chip dimensions, the location of the active layer, the refractive indices, and absorption coefficients of all optical media. For simplicity, the active layer is assumed a Lambertian surface. In most conditions, large amount of rays is required to have a stable simulated light pattern since most rays cannot escape from the LED die (Zukauskas et al. 2002; Sun et al. 2006). This condition becomes worst when one needs to develop an optical system based on the LED. Therefore, in the modeling process, the distribution of the ray vectors of the six exit faces is recorded, and then the lights are reemitted from the six faces so that serious loss of rays inside the LED die is avoided.

The verification of the optical model contains two parts, where measurement is the first and a comparison between the measurement and the simulation is the second. Since most optical elements of an LED-based lamp are close to the LED, the comparison between the simulation and the real patterns of an LED should be made in this region, which is neither the near field nor the far field, and thus another optical field called midfield is defined (Sun et al. 2006). The midfield region is defined as the distance between the near-field (vector region) and the far-field (Fraunhofer region) regions, as shown in Fig. 8 (Sun et al. 2006; Goodman 1996). In the midfield region, all lights are in propagation mode and can be accurately described with the scalar theory. In contrast to the light pattern in the far field, in the midfield the angular light pattern varies from a distance to another, so the comparison should be made at various distances in the midfield to ensure that the optical model is accurate enough.

An example to show the precise modeling is as follow. Figure 9 shows a picture of a commercialized white LED, where a conformal-coated phosphor is applied on the blue die. The phosphor-coated LED die is immersed in a silicone lens. Because the phosphor may act as a scattering medium for blue light and a medium for spontaneous emission of yellow light, the top layer of the phosphor can be regarded as a Lambertian surface emitting white light in the first-step modeling (Chien et al. 2007;

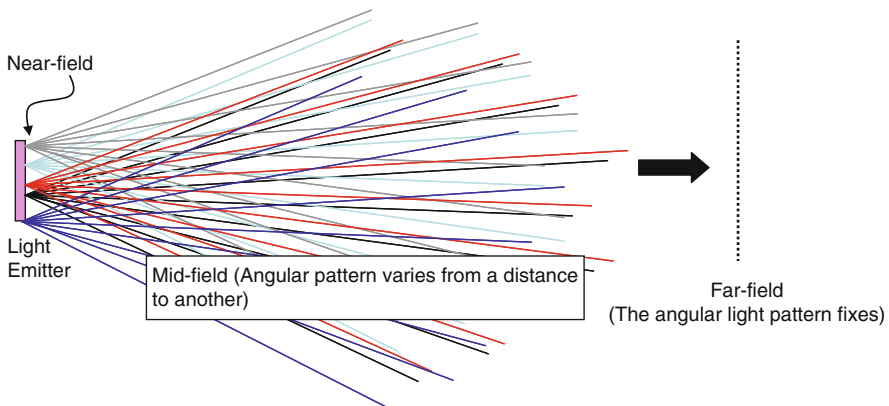


Fig. 8 Schematic diagram to illustrate the concept of midfield

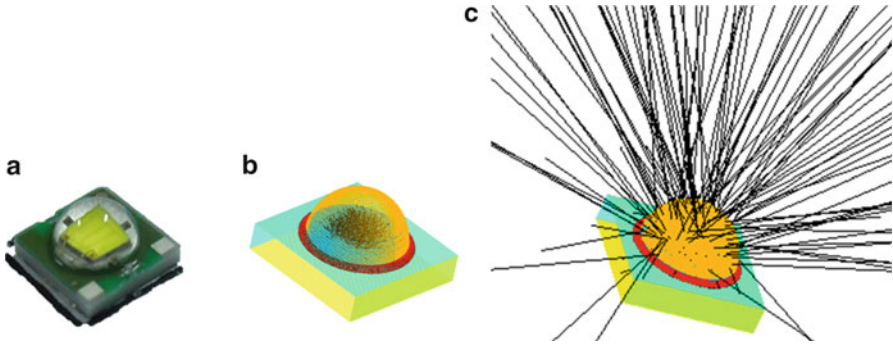


Fig. 9 (a) The photo of the modeled LED, (b) the geometrical model, and (c) lights emitted with Monte Carlo ray tracing

Borbely and Johnson 2005). Then one of the important steps is to construct a real-size white LED according to the LED. All the surfaces or media in the white LED must be equipped with suitable optical parameters, such as reflectivity, absorption coefficient, or refractive index. Through Monte Carlo ray tracing, one could simulate the light distribution in the whole space. Then the angular light distribution at several planes located at the midfield region, such as 1, 3, and 5 cm, is used to compare the measurement. In order to figure out the similarity between the simulation pattern and the experimental measurement, normalized cross-correlation (NCC) is applied (Sun et al. 2006; Lewis 1995). The NCC is written

$$\text{NCC} = \frac{\sum_x \sum_y (A_{xy} - \bar{A})(B_{xy} - \bar{B})}{\sqrt{\sum_x \sum_y (A_{xy} - \bar{A})^2 \sum_x \sum_y (B_{xy} - \bar{B})^2}} \quad (2)$$

where A_{xy} and B_{xy} are the intensity or irradiance of the simulation (A) and experimental values (B); \bar{A} (\bar{B}) is the mean value of A (B) across the x–y plane. Figures 10 shows a comparison between the simulation and the experimental measurement at several midfield distances, where the NCCs between the simulation and the corresponding measurement are all higher than 99.5 %, which is a criteria of precise model (Sun et al. 2009a).

The precise light source model is important in LED lighting. One reason is that the modeling procedure is helpful in defining an optical-qualified LED to avoid unacceptable variation from one piece to another due to low quality control in packaging process. Besides, the precise model is quite helpful in advanced optical design in LED lighting (Sun et al. 2006; Kaminski et al. 2002; Zerhau-Dreihöfer et al. 2002). Figure 11 shows a projection light pattern by the design with a TIR (total internal reflection) lens based on the precise LED model. The light pattern of the real sample is highly similar to the simulation. The optical modeling for a complicated

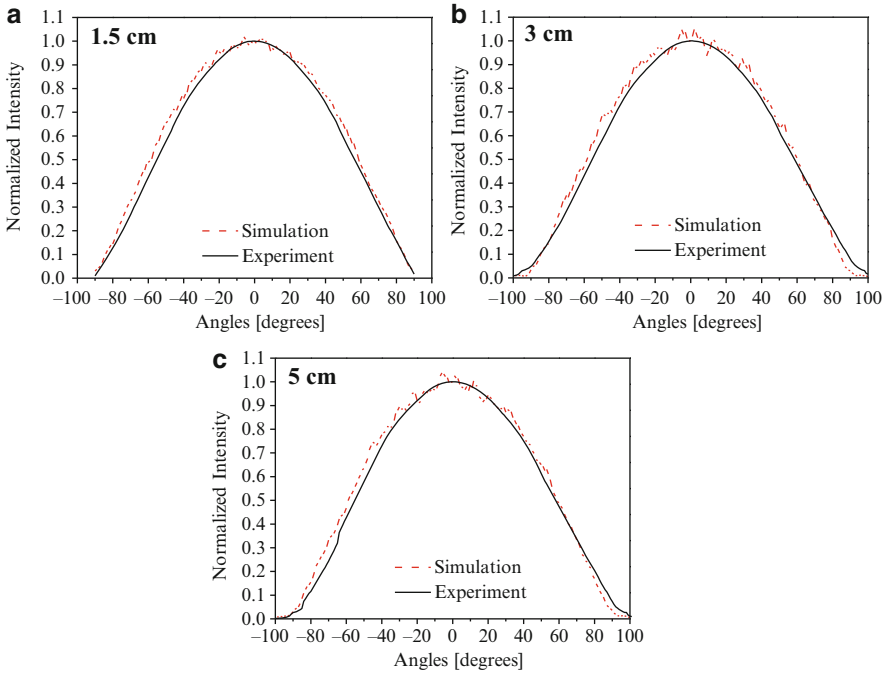


Fig. 10 One-dimensional light pattern in measurement and simulation at (a) 1.5 cm, (b) 3 cm, and (c) 5 cm

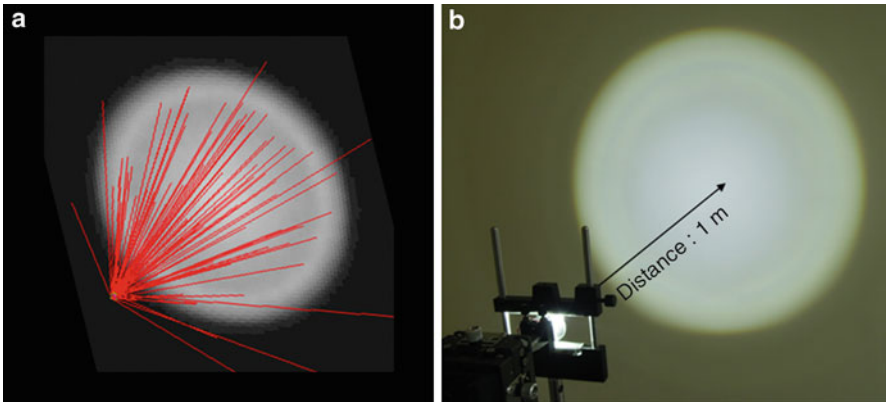


Fig. 11 (a) The projected light pattern by an optical design-based mid-field modeling and (b) the corresponding experiment

white LED could need modification in describing the surface emitting property. Figure 12 shows a case of a multi-die white LED, where eight blue dies are bonded in a power LED. To obtain a precise model, the light-emitting area needs to be adjusted. The area between dies could reflect obvious lights so that the surface in this

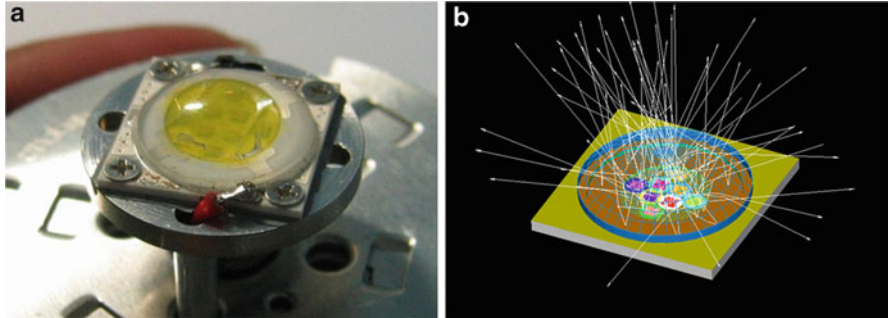


Fig. 12 (a) A photo of a multi-chip white LED and (b) the corresponding geometrical model

part could be treated as a low-brightness emitting area. A weight factor to describe emitting intensity could be introduced to have more accurate model (Chien et al. 2007; Strojnik and Paez 2001).

Optical Modeling for Phosphors

Most white LEDs in practical application contain a phosphor layer, which are always called phosphor-converted white LED, simplified as pcW-LED. In most cases, the phosphor emits lights with a peak wavelength at yellow or two peaks at green and red. The former is simple and more cost-effective, and the latter is better in color rendering index owing to broader bandwidth. In the packaging side, the phosphor associated with the packaging geometry is the main issue that decides the optical and chromatic performance. The key parameters include the particle size, concentration, absorption coefficient, quantum efficiency, thickness, and other geometrical factors (Sun et al. 2008; Kerker et al. 1979; Chang et al. 2005). A serious problem arises when the distribution of blue light and other lights, e.g., yellow light, is different. The blue light is emitted by LED die and passes through the phosphor layer with Mie scattering effect (Sun et al. 2008; Doicu and Wriedt 2001; Toublanc 1996; Boren and Huffmarn 1983). In contrast, the yellow light is emitted by the phosphor pumped by the blue light, and it is isotropic with spontaneous emission. The inherent difference in light emission causes obvious difference in spatial distribution. Therefore, optical modeling for phosphors in packaging level is strongly demanded.

To model the optical and color properties of a pcW-LED, the optical modeling algorithm is more complicated than the light source modeling. The optical property of phosphor depends on phosphor composition, particle size, absorption coefficient, quantum efficiency, and others. A reliable simulation is still based on Monte Carlo ray tracing. The simulation, however, needs to take care of light scattering, absorption ability, and photon conversion. All these simulations cannot be well done if the related parameters are not accurate enough. Thus, several corresponding experimental measurements are necessary.

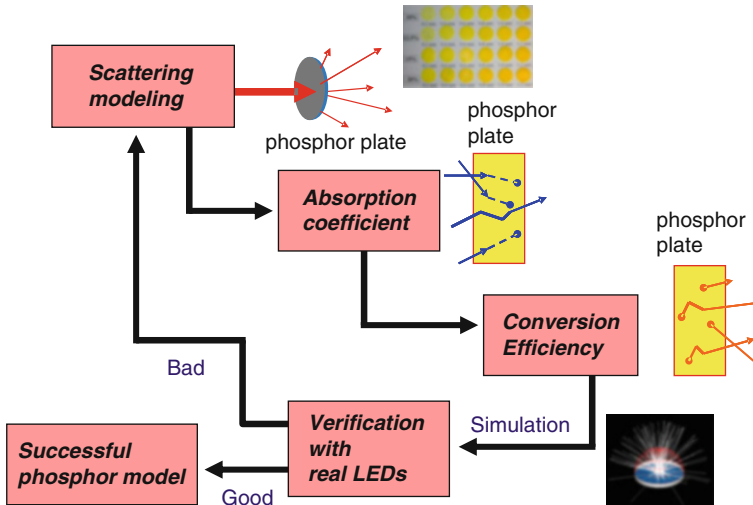


Fig. 13 The optical modeling procedure for phosphor simulation in packaging level

One effective optical modeling procedure for phosphor in packaging level is shown in Fig. 13 (Sun et al. 2008). The modeling starts from scattering simulation which relates to particle size distribution of the phosphor, thickness of the phosphor plate, and the concentration. The corresponding simulation follows Monte Carlo ray tracing incorporated with Mie scattering. Several phosphor plates of different phosphor concentrations and different thicknesses are used in the measurement. Each phosphor plate is located at the rotational center of a rotational stage. A laser with wavelength of 632.8 nm is used as the light source, and a power meter rotates around the phosphor plate and detects the scattered light. Once all angular distributions of the scattering light for different phosphor plate thicknesses and concentrations are obtained, simulation is performed with tuning the refractive index and the particle size of the phosphor until the simulated scattering distribution fits the measurement in the experiment, as shown in Fig. 14 (Sun et al. 2008). Then the next step is to figure out the absorption coefficient and conversion efficiency. To accurately obtain these two parameters, the tested phosphor plate is put over a black cavity, where a blue die is placed on the bottom side, as shown in Fig. 15 (Sun et al. 2008). The black cavity is used to absorb the reflected light from the phosphor plate. The transmitted lights in yellow and blue from the phosphor plate are measured as shown in Fig. 16 (Sun et al. 2008; Kang et al. 2006), when the tested samples are placed in an integrating sphere. The transmitted blue light is used to calculate the absorption coefficient, and the transmitted yellow light is used to calculate the conversion efficiency.

The phosphor model is finally applied to simulate the spatial distribution of the blue and the yellow lights by a real packaging pcW-LED. It is helpful to simulate the emission spectrum, the correlated color temperature (CCT), and the packaging efficiency.

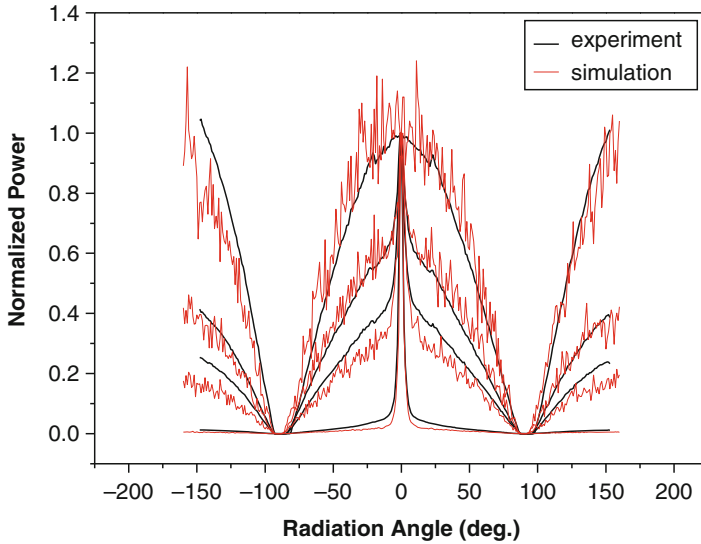


Fig. 14 The experimental measurement of scattering light distribution for different phosphor concentrations and the corresponding simulation

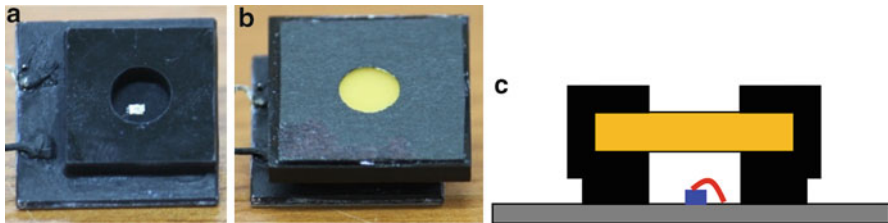


Fig. 15 (a) The LED die with a black cavity, (b) the cavity covered with a phosphor plate, and (c) the schematic diagram

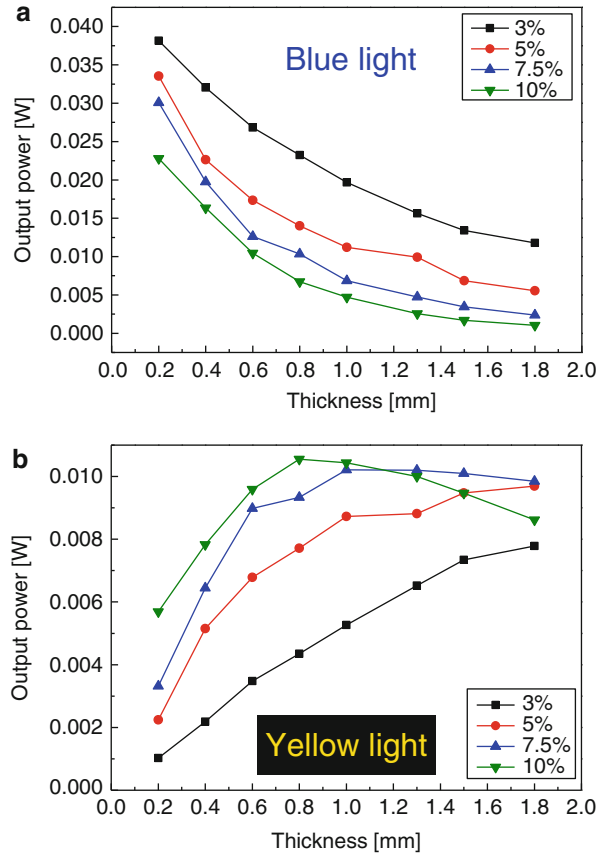
Chromatic Performance

The chromatic performance as well as the spatial distribution is another important issue in a pcW-LED. Two factors could be important in determining the chromatic performance of a pcW-LED. One is color rendering index (CRI) and the other is spatial CCT deviation. The former relates to the color appearance to the illuminated object, and the latter relates to the spatial uniformity of the color performance.

CRI with CCT

Generally, the CRI of a pcW-LED with a blue die covered by yellow phosphor is around 70 (Steigerwald et al. 2002; Sun et al. 2012a; Mueller-Mach et al. 2005;

Fig. 16 (a) The measurement of blue lights and (b) of yellow lights at different phosphor concentrations and thicknesses



Yamada et al. 2003; Xie et al. 2007; Neeraj et al. 2004; Jüstel et al. 1998). Such a pcW-LED is fine if applied to outdoor lighting, but is not good enough in indoor lighting, where the CRI is requested to reach 80 or higher. Thus, two phosphor compositions, usually emitting green and red lights, are used to extend the emission spectrum so that green and red colors can be well presented. In order to well predict the chromatic performance, a well-developed phosphor model is demanded. However, a barrier exists owing to reabsorption of the reemitted green light by red phosphor, so that well prediction of two-phosphor pcW-LED becomes difficult. If one just needs to figure out the chromatic performance of a two-phosphor pcW-LED, a simpler approach instead of the complicated two-phosphor model is applicable. Three simple approaches to predict chromatic performance are proposed in 2012 (Sun et al. 2012b). These approaches start from a linear combination of the spectra of the blue die, green, and red phosphors. The spectrum by the linear combination does not touch the reabsorption effect so the prediction of the linear model is not accurate enough. The most attractive approach is an alternative way to the linear model and is so-called linear bridge model (LBM). In the LBM, a straight line is used to connect

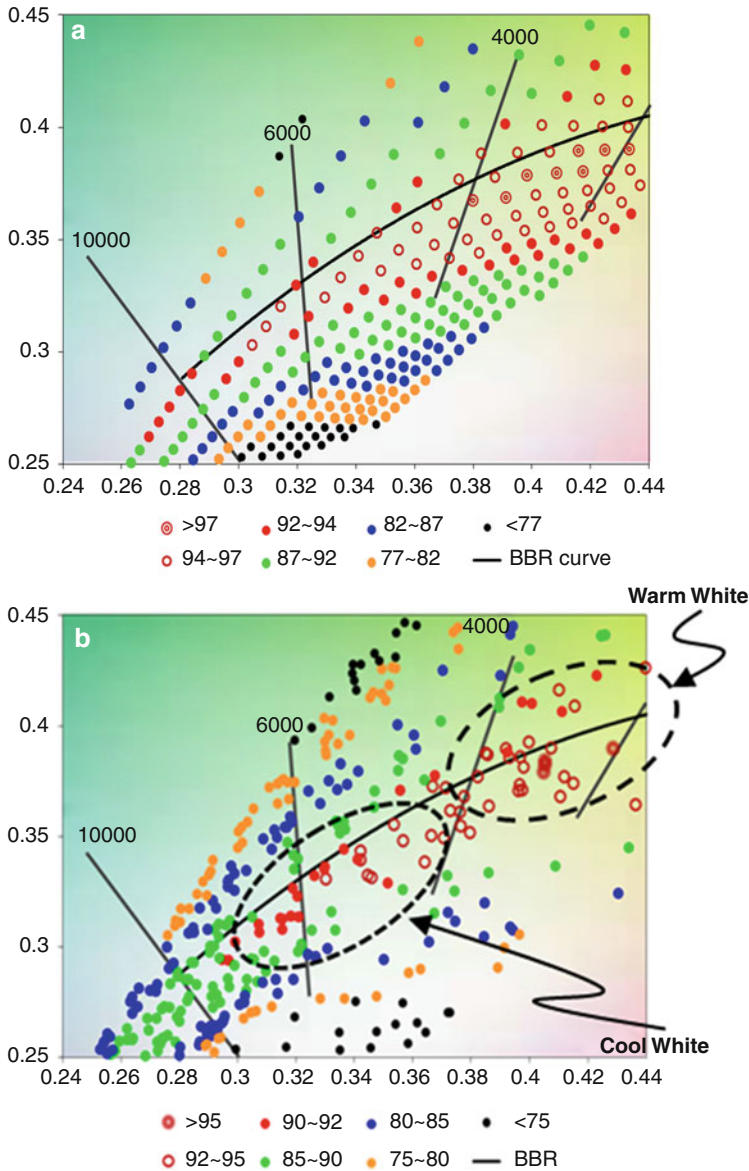


Fig. 17 (a) The simulation of the CRI with CCT locations and (b) the corresponding measurement result with phosphor plates

the wavelength peak of the two phosphors in addition to the spectrum predicted by the linear combination model. Figure 17a shows a map of chromatic performance in color coordinate to show the best performance of the CRI with the location of CCT. A corresponding experimental measurement is shown in Fig. 17b, where similar

chromatic map can be found. It shows that the LBM is a useful model to predict the chromatic performance of a two-phosphor pcW-LED. With green and red phosphors, the CRI could be higher than 90 when the CCT is larger than 5000 K, and the CRI could be as high as 95 when the CCT is located between 3000 and 5000 K.

Spatial CCT Uniformity

The spatial CCT uniformity is another important issue that must be paid attention in a pcW-LED. The spatial CCT uniformity can be evaluated through a factor called angular CCT deviation (ACCTD), which is defined by the difference between the maximum and minimum CCT in a hemisphere space centered by the normal axis of the pcW-LED. There are several ways to reduce ACCTD. A design for conformal coating of phosphor within a lens is proposed, and the simulation shows that an ACCTD around 900 K was achieved as the CCT of the white LED was around 6000 K (Sun et al. 2012b; Liu et al. 2008). A planar color conversion element is proposed to reach an ACCTD as small as 300 K in simulation when the CCT was around 5600 K (Sun et al. 2012b; Sommer et al. 2009). A design of a silicone lens is proposed to make a uniform CCT distribution, and the simulation shows that an ACCTD as small as 40 K could be achieved when the CCT was 5000 K (Sun et al. 2012b; Wang et al. 2010). A phosphor structure in a convex shape over a metal cup is proposed to have an ACCTD of 200 K when the CCT is 5000 K (Sun et al. 2012b; Shuai et al. 2011). In fact, the ACCTD depends on the CCT of a pcW-LED. If the CCT is small, such as the neutral white or warm white LED, the ACCTD could be relatively small because more scattering occurs in a phosphor volume so that the lights in different colors could be well mixed. In contrast, the ACCTD in a cool white LED is always relatively large. There is no surprise if a pcW-LED of the CCT 6500 K performs an ACCTD as large as 3000 K or more (Sun et al. 2012b; Huang et al. 2010; Borbely and Johnson 2005). A design to perform extreme small ACCTD at all CCT ranges is proposed and demonstrated based on precise phosphor model and special optical design. The design concept is to build up a phosphor dome such as a hemisphere covering a blue die. If the diameter of the phosphor dome is much larger than the blue die, the distance to pass through the phosphor volume could be equal along all directions as shown in Fig. 18 (Sun et al. 2009b, 2012b). Therefore, it is possible to have a minimum ACCTD. Such design is limited by the size of the phosphor dome because the phosphor volume is preferred to be small for compactness and cost down issue as well. A dome with a compromised diameter of 3 mm with respect to a die of 0.66 mm could perform an ACCTD as small as 300 K. Advanced design is to extend vertical length of the dome from a hemisphere to enable most directional blue light to accumulate enough optical path in the phosphor volume (Sun et al. 2012b). The optimized result of the ACCTD from warm white to cool white is shown in Fig. 19, where below 200 K in ACCTD is performed at 6500 K and below 100 K at the CCT smaller than 5000 K.

Fig. 18 The schematic diagram of a pcW-LED with a phosphor dome

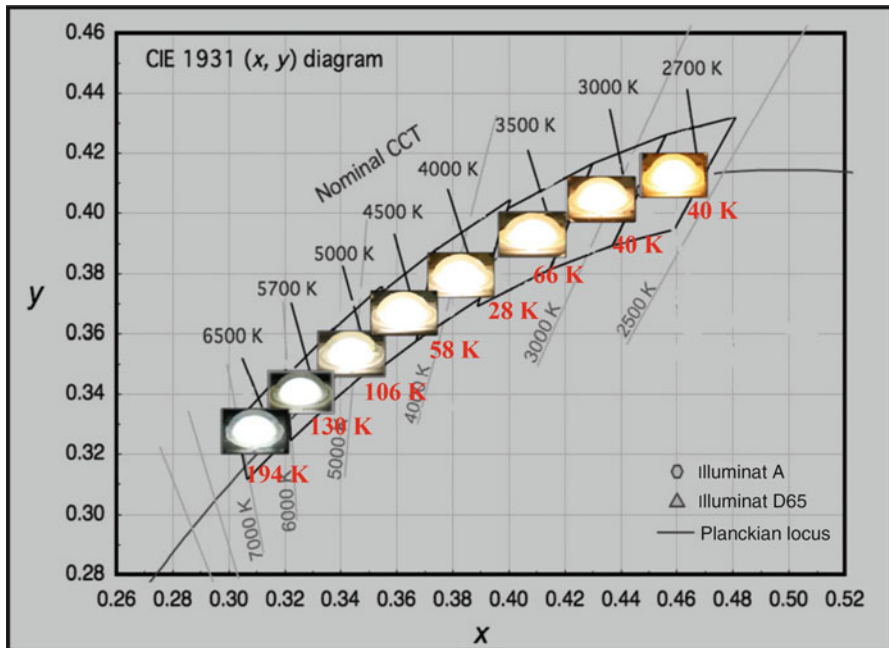
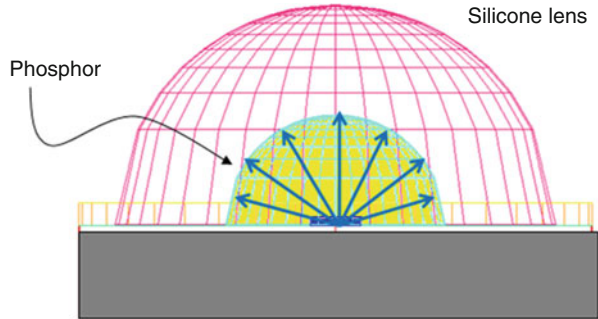


Fig. 19 The ACCTD values with the CCTs

Design of Encapsulation Lens

The encapsulation lens plays multiple roles in LED packaging, such as protection of the phosphor and die, enhancement of the light exaction efficiency, and change the light pattern. Light pattern controlling in the packaging level is important to the practical application. If an LED is applied to a spot light, the design of the encapsulation lens is to keep the etendue (Welford and Winston 1989; Fournier and Rolland 2008; Alexander 2008; Sun et al. 2013) of the LED die so that the optical

Fig. 20 Different designs in encapsulation lens. (a) A Lambertian type and (b) a side emitting

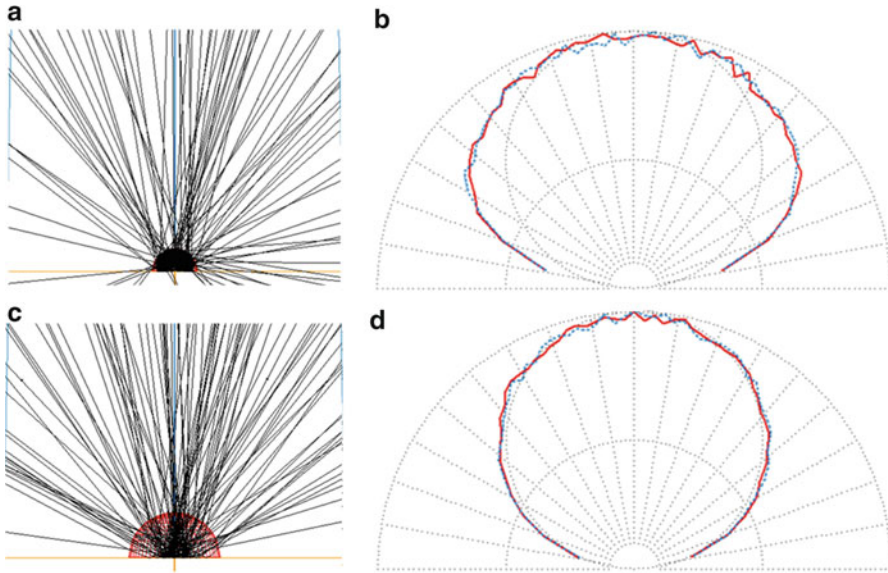
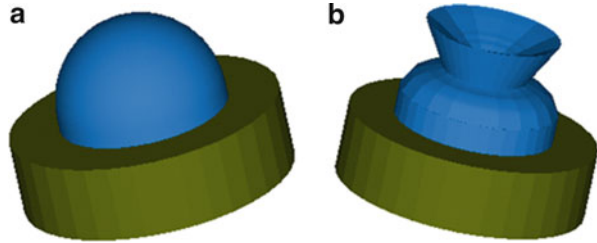


Fig. 21 Lens diameter could change the light pattern. (a) The diameter of the lens is 1.5 mm and (b) the corresponding light pattern in simulation. (c) The diameter of the lens is 3 mm and (d) the corresponding light pattern in simulation

efficiency in the luminaire can reach the theoretical maximum. If an LED is applied to a diffuse luminaire, a wide-angle light pattern could be desired. Figure 20 shows two typical designs for the encapsulation lens. The first one is a dome lens with a shape near a hemisphere, and the latter is the so-called side-emitting lens, which is to spread the lights to wider angles. In the design of the side-emitting lens, the upper boundary forms a total reflection surface so that all directional lights are redirected to lateral directions. By adjusting the slanted angle of the surface, the emitting lights can be directed to different angles centered at $60\text{--}90^\circ$.

The encapsulation lens in a shape of a hemisphere is most observed in practical application. As shown in Fig. 21, if the diameter of the hemisphere lens is large enough in comparison with the die size, the emitting lights will suffer less Fresnel loss, and the light pattern can keep the Lambertian distribution. This approach can always keep the tendue so that it is more suitable for the application of spot light or projection.

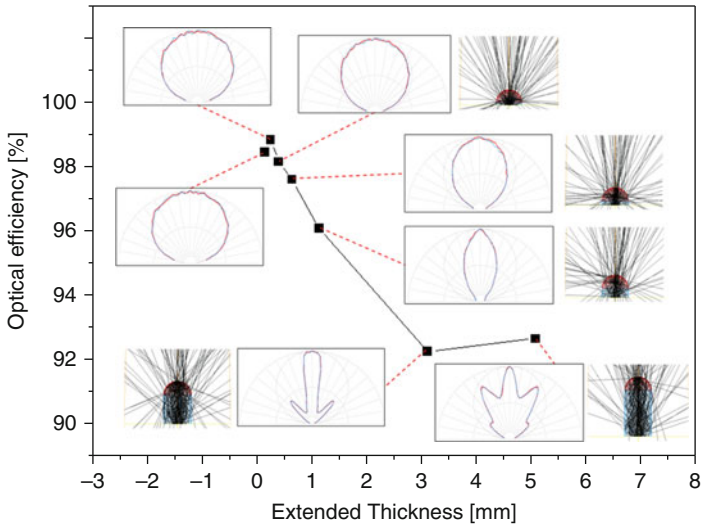


Fig. 22 The simulated optical efficiency of encapsulation lens with different extended thicknesses in the case of the lens diameter of 3 mm

In some applications, the light pattern of an LED is requested to perform more directional. Such demand can be easily met by changing the lens shape. One of the approaches is to add an extended thickness to the dome lens. A simulation of the optical efficiency and the light pattern with the different extended thicknesses for a dome lens is shown in Fig. 22. We can find that the changing of the extended thickness can alter the light pattern. One can find that in the case of 1 mm of the extended thickness, the light pattern is around 56° in FWHM in comparison with 120° without the extended thickness. However, the larger extended thickness is, the lower the optical efficiency will be, owing to larger Fresnel loss.

References

- Alexander W (2008) Requirements on LEDs in extended limited light engines. Proc SPIE 7001, 70010F:1–10
- Baba T, Inoshita K, Tanaka H, Yonekura J, Ariga M, Matsutani A, Miyamoto T, Koyama F, Iga K (1999) Strong enhancement of light extraction efficiency in GaInAsP 2-D-arranged microcolumns. J Lightwave Technol 17:2113–2120
- Badano A, Kanicki J (2001) Monte Carlo analysis of the spectral photon emission and extraction efficiency of organic light-emitting devices. J Appl Phys 90:1827–1830
- Borbely A, Johnson SG (2005) Performance of phosphor-coated light-emitting diode optics in ray-trace simulations. Opt Eng 44:111308
- Boren CF, Huffman DR (1983) Absorption and scattering of light by small particles. Wiley, New York
- Chang CC, Chern R, Chang CC, Chu C, Chi JY, Su J, Chan I-M, Wang JT (2005) Monte Carlo simulation of optical properties of phosphor-screened ultraviolet light in a white light-emitting device. Jpn J Appl Phys 44:6056–6061

- Chien WT, Sun CC, Moreno I (2007) Precise optical model of multi-chip white LEDs. *Opt Express* 15:7572–7577
- Djurić AB, Chan Y, Li BH (2001) Calculations of the refractive index of AlGaIn/GaN quantum well. *Proc SPIE* 4283:630–637
- Doicu A, Wriedt T (2001) Equivalent refractive index of a sphere with multiple spherical inclusions. *Appl Opt* 3:204–209
- Fournier F, Rolland J (2008) Design methodology for high brightness projectors. *J Disp Technol* 4:86–91
- Fujii T, Gao Y, Sharma R, Hu EL, Danbaars SP, Nakamura S (2004) Increase in the extraction efficiency of GaN-base light emitting diodes via surface roughening. *Appl Phys Lett* 84:855
- Gao Y, Fujii T, Sharma R, Fujito K, Danbaars SP, Nakamura S (2004) Roughening hexagonal surface morphology on laser lift-off (LLO) N face GaN with simple photo-enhanced chemical wet etching. *Jpn J Appl Phys* 43:637
- Goodman JW (1996) Introduction to fourier optics, 2nd edn. McGraw-Hill, San Francisco
- Guo X, Graff J, Schubert EF (1999) Photon-recycling semiconductor light-emitting diode. *IEDM Tech Dig IEDM-99:600*
- Guo X, Li Y-L, Schubert EF (2001) Efficiency of GaN/InGaIn light-emitting diodes with interdigitated mesa geometry. *Appl Phys Lett* 79:1936
- Haerle V, Hahn B, Kaiser S, Weimar A, Bader S, Eberhard F, Plössl A, EisertHigh D (2004) High brightness LEDs for general lighting applications using the new ThinGaIn™-Technology. *Phys Status Solidi* 201:2736–2739
- Huang HT, Tsai CC, Huang YP (2010) Conformal phosphor coating using pulsed spray to reduce color deviation of white LEDs. *Opt Express* 18:A201–A206
- Huh C, Lee KS, Kang EJ, Park SJ (2003) Improved light-output and electrical performance of InGaIn-based light-emitting diode by microroughening of the p-GaN surface. *Appl Phys Lett* 93:9383
- Joyce WB, Bachrach RZ, Dixon RW, Sealer DA (1974) Geometrical properties of random particles and the extraction of photons from electroluminescent diodes. *J Appl Phys* 45:2229
- Jüstel T, Nikol H, Ronda C (1998) New developments in the field of luminescent materials for lighting and displays. *Angew Chem Int Ed* 37:3084–3103
- Kaminski MS, Garcia KJ, Stevenson MA, Frate M, Koshel RJ (2002) Advanced topics in source modeling. *Proc SPIE* 4775:46
- Kang D, Wu E, Wang D (2006) Modeling white light-emitting diodes with phosphor layers. *Appl Phys Lett* 89:231102
- Kerker M, Chew H, McNulty PJ, Kratochvil JP, Cooke DD, Sculley M, Lee MP (1979) Light scattering and fluorescence by small particles having internal structure. *J Histochem Cytochem* 27:250–263
- Krames MR, Ochiai-Holcomb M, Hoffer GE, Carter-Coman C, Chen EI, Tan I-H, Grillot P, Gardner NF, Chui HC, Huang J-W, Stockman SA, Kish FA, Craford MG, Tan TS, Kocot CP, Hueschen M, Posselt J, Loh B, Sasser G, Collins D (1999) High-power truncated-inverted-pyramid $(\text{Al}_x\text{Ga}_{1-x})_{0.5}\text{In}_{0.5}\text{P}/\text{GaP}$ light-emitting diodes exhibiting 50 % external quantum efficiency. *Appl Phys Lett* 75:2365–2367
- Krames MR, Shchekin OB, Mueller-Mach R, Mueller GO, Zhou L, Harbers G, Craford MG (2007) Status and future of high-power light-emitting diodes for solid-state lighting. *J Disp Technol* 3:160–175
- Kurata K, Ono Y, Ito K, Mori M, Sano H (1981) An experimental study on improvement of performance for hemispherically shaped high-power IREDs with $\text{Ga}_{1-x}\text{Al}_x\text{As}$ grown junctions. *IEEE Trans Electron Devices* 28:374–379
- Lee SJ (1998) Analysis of InGaIn high-brightness light-emitting diodes. *Jpn J Appl Phys* 37:5990
- Lee SJ (2001) Analysis of light-emitting diode by Monte Carlo photo simulation. *Appl Opt* 40:1427
- Lee SJ, Song SW (1999) Efficiency improvement in light-emitting diodes based on geometrically deformed chips. *Proc SPIE* 3621:237–248

- Lee TX, Lin CY, Ma SH, Sun CC (2005) Analysis of position-dependent light extraction of GaN-based LEDs. *Opt Express* 13:4175–4179
- Lee TX, Gao KF, Chien WT, Sun CC (2007) Light extraction analysis for GaN-based LEDs with pyramid array. *Opt Express* 15:6670–6676
- Lewis JP (1995) Fast template matching. *Vision Interface* 95:120–123
- Linder N, Kugler S, Stauss P, Streubel KP, Wirth R, Zull H (2001) High-brightness AlGaInP light-emitting diodes using surface texturing. *Proc SPIE* 4278:19–25
- Liu Z, Liu S, Wang K, Luo X (2008) Optical analysis of color distribution in white LEDs with various packaging methods. *IEEE Photon Technol Lett* 20:2027–2029
- Mueller-Mach R, Mueller GO, Krames MR, Höpfe HA, Stadler F, Schnick W, Juestel T, Schmidt P (2005) Highly efficient all-nitride phosphor-converted white light emitting diode. *Phys Status Solidi A* 202:1727–1732
- Muth JF, Brown JD, Johnson MAL, Yu Z, Kolbas RM, Cook JW Jr, Schetzina JF (1999) Absorption coefficient and refractive index of GaN, AlN and AlGaIn alloys. *MRS Internet J Nitride Semicond Res* 4S1:G5.2
- Neeraj S, Kijima N, Cheetham AK (2004) Novel red phosphors for solid-state lighting: the system $\text{NaM}(\text{WO}_4)_{2-x}(\text{MoO}_4)_x:\text{Eu}^{3+}$ ($M = \text{Gd}, \text{Y}, \text{Bi}$). *Chem Phys Lett* 387:2–6
- Nguyen F, Terao B, Laski J (2005) Realizing LED illumination lighting applications. *Proc SPIE* 5941:31
- Schnitzer I, Yablonoivitch E, Caneau C, Gmitter TJ, Scherer A (1993) 30 % external quantum efficiency from surface textured, thin-film light-emitting diodes. *Appl Phys Lett* 63:2174–2176
- Schubert EF, Kim JK (2005) Solid-state light sources becoming smart. *Science* 308:1274–1278
- Shuai Y, He YZ, Tran NT, Shi FG (2011) Angular CCT uniformity of phosphor converted white LEDs: effects of phosphor materials and packaging structures. *IEEE Photon Technol Lett* 23:137–139
- Sommer C, Hartmann P, Pachler P, Schweighart M, Tasch S, Leising G, Wenzl FP (2009) A detailed study on the requirement for angular homogeneity of phosphor converted high power white LED light sources. *Opt Mater* 31:837–848
- Steigerwald DA, Bhat JC, Collins D, Fletcher RM, Holcomb MO, Ludowise MJ, Martin PS, Rudaz SL (2002) Illumination with solid state lighting technology. *IEEE J Sel Top Quantum Electron* 8:3:10
- Strojnik M, Paez G (2001) Radiometry. In: Malacara D, Thompson B (eds) *Handbook of optical engineering*
- Sun CC, Lin CY, Lee TX, Yang TH (2004) Enhancement of light extraction of GaN-based LED with introducing micro-structure array. *Opt Eng* 43:1700–1701
- Sun CC, Lee TX, Ma SH, Lee YL, Huang SM (2006) Precise optical modeling for LED lighting based on cross-correlation in mid-field region. *Opt Lett* 31:2193–2195
- Sun CC, Chen CY, He HY, Chen CC, Chien WT, Lee TX, Yang TH (2008) Precise optical modeling for silicate-based white LEDs. *Opt Express* 16:20060–20066
- Sun CC, Chien WT, Moreno I, Hsieh CC, Lo YC (2009a) Analysis of the far-field region of LEDs. *Opt Express* 17:13918–13927
- Sun CC, Chen CC, Chien WT, Chen CY, Lee TX, Yang TH (2009b) Precise phosphor model and the application to LED package of high uniformity in spatial CCT. In: *The second international conference on white LEDs and solid state lighting, proceedings, paper TA2–2*
- Sun CC, Lee TX, Lo YC, Chen CC, Tsai SY (2011a) Light extraction enhancement of GaN-based LEDs. *Opt Commun* 284:4862–4868
- Sun CC, Tsai SY, Lee TX (2011b) Enhancement of angular flux utilization based on implanted micro pyramid array and lens encapsulation in GaN LEDs. *J Disp Technol* 7:289–292
- Sun CC, Chen CY, Chen CC, Chiu CY, Peng YN, Wang YH, Chung CY, Yang TH, Chung TY (2012a) High uniformity in angular correlated-color-temperature distribution of white LEDs from 2800 K to 6500 K. *Opt Express* 20:6622–6630
- Sun CC, Chen CY, Chang JH, Yang TH, Ji WS, Jeng YS, Wu HM (2012b) Linear calculation model for prediction of CRI performance associated with CCT of white LEDs with two phosphors. *Opt Eng* 51:054003

- Sun CC, Chung SC, Yang SH, Yu YW, Chien WT, Chen HK, Chen SP (2013) High-directional light source using photon recycling with a retro-reflective Dome incorporated with a textured LED die surface. *Opt Express* 21:18414–18423
- Ting DZ, McGill TC (1995) Monte Carlo simulation of light-emitting diode light extraction characteristics. *Opt Eng* 34:3545–3553
- Toublanc D (1996) Henyey-Greenstein and Mie phase functions in Monte Carlo radiative transfer computations. *Appl Opt* 35:3270–3274
- Wang K, Wu D, Chen F, Liu ZY, Luo XB, Liu S (2010) Angular color uniformity enhancement of white light-emitting diodes integrated with freeform lenses. *Opt Lett* 35:1860–1862
- Welford WT, Winston R (1989) High collection nonimaging optics. Academic, San Diego
- Wierer JJ, Steigerwald DA, Krames MR, O'Shea JJ, Ludowise MJ, Gardner NF, Kern RS, Stockman SA (2001) High-power AlGaInN flip-chip light-emitting diodes. *Appl Phys Lett* 78:3379–3381
- Windisch R, Dutta B, Kuijk M, Knobloch A, Meinschmidt S, Windisch SR, Dutta B, Kuijk M, Knobloch A, Meinschmidt S, Schoberth S, Kiesel P, Borghs G, Döhler GH, Heremans P (2000) 40 % efficient thin-film surface-textured light-emitting diodes by optimization of natural lithography. *IEEE Trans Electron Devices* 47:1492–1498
- Windisch R, Meinschmidt S, Rooman C, Zimmermann L, Dutta B, Kuijk M, Kiesel P, Doehler GH, Borghs G, Heremans PL (2001a) Light extraction mechanisms in surface-textured light emitting diodes. *Proc SPIE* 4278:90–98
- Windisch R, Rooman C, Meinschmidt S, Kiesel P, Zipperer D, Döhler GH, Dutta B, Kuijk M, Borghs G, Heremans P (2001b) Impact of texture-enhanced transmission on high-efficiency surface-textured light-emitting diodes. *Appl Phys Lett* 79:2315–2317
- Windisch R, Rooman C, Dutta B, Knobloch A, Borghs G, Döhler GH, Heremans P (2002) Light-extraction mechanisms in high-efficiency surface-textured light-emitting diodes. *IEEE J Sel Top Quantum Electron* 8:248
- Wong WS, Sands T, Cheung NW, Kneissl M, Bour DP, Mei P, Romano LT, Johnson NM (1999) Fabrication of thin-film InGaN light-emitting diode membranes by laser lift-off. *Appl Phys Lett* 72:1360–1362
- Xie RJ, Hirosaki N, Kiumra N, Sakuma K, Mitomo M (2007) 2-phosphor-converted white light-emitting diodes using oxynitride/nitride phosphors. *Appl Phys Lett* 90:191101
- Yamada M, Naitou T, Izuno K, Tamaki H, Murazaki Y, Kameshima M, Mukai T (2003) Red-enhanced white-light-emitting diode using a new red phosphor. *Jpn J Appl Phys* 42: L20–L23
- Zerhau-Dreihöfer H, Haack U, Weber T, Wendt D (2002) Light source modeling for automotive lighting devices. *Proc SPIE* 4775:58
- Zukauskas A, Shur MS, Caska R (2002) Introduction to solid-state lighting. Wiley, New York

Part III

OLEDs/PLEDs

White OLED Materials

Yonghua Chen and Dongge Ma

Contents

Introduction	294
Organic Semiconductors	294
Molecular Orbitals	294
Optical Properties of Organic Molecules	295
Energy Transfer and Loss Processes	297
Electroluminescence	297
Current Status of White OLED Materials	299
Fluorescence Materials	300
Phosphorescence Materials	302
Delay Fluorescence Materials	308
Polymer Materials	312
Future Outlook	314
References	314

Abstract

White OLEDs are ultrathin, large-area light sources made from organic semiconductor materials. Over the past decades, much research has been spent on finding suitable materials to realize highly efficient monochrome and thus white OLEDs. White OLED panel efficacy has reached 90 lmW^{-1} , and a tandem white OLED panel has achieved a lifetime of over 100,000 h at $1,000 \text{ cdm}^{-2}$. LG is set to launch a 55" OLED TV in 2013, and OLEDs will be expected to make a bigger breakthrough. Although white OLED panels show superior performance, there is still much room (nearly 160 lmW^{-1}) for improvement, in view of the theoretical limit of 248 lmW^{-1} . With their high-efficiency, color tunability, and color quality, white OLEDs are emerging as one of the next-generation light sources.

Y. Chen (✉) • D. Ma

Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, China

e-mail: chenyh0302@hotmail.com; mdg1014@ciac.jl.cn

Introduction

Since the 1970s, the successful synthesis and controlled doping of conjugated polymers established the important class of organic semiconductors, which was honored with the Nobel Prize in Chemistry in the year 2000 (Chiang et al. 1977). The main impetus came from the demonstration of high-performance organic light-emitting diodes (OLEDs) incorporating an organic heterojunction of *p*- and *n*-conducting organic semiconductors from vacuum-evaporated molecular films (Tang and Vanslyke 1987) and from conjugated polymers (Burroughes et al. 1990), as well as the first successful fabrication of efficient photovoltaic cells and thin-film transistors from conjugated polymers and oligomers (Tang 1986; Koezuka et al. 1987; Burroughes et al. 1988; Horowitz et al. 1989).

The enormous progress in this field has been driven by the anticipation of novel organic light-emitting materials, including small molecules and polymers, and novel applications, such as low-cost, large-area, and flexible light sources and displays. Early efforts following these pioneering works focused on the improvement of these materials and devices with respect to their efficiency, stability, and color tunability; however, solely monochrome materials and devices have been investigated. Roughly a decade later, the first white OLEDs (Kido et al. 1994) and PLEDs (Wang et al. 1999) were reported, demonstrating that LEDs based on organic materials can become an alternative for general lighting applications.

White OLEDs are ultrathin, large-area light sources made from organic semiconductor materials. Over the past decades, much research has been spent on finding suitable materials to realize highly efficient monochrome and thus white OLEDs. White OLED panel efficacy has reached 90 lmW^{-1} , and a tandem white OLED panel has achieved a lifetime of over 100,000 h at $1,000 \text{ cdm}^{-2}$. LG is set to launch a 55" OLED TV in 2013, and OLEDs will be expected to make a bigger breakthrough. Although white OLED panels show superior performance, there is still much room (nearly 160 lmW^{-1}) for improvement, in view of the theoretical limit of 248 lmW^{-1} . With their high-efficiency, color tunability, and color quality, white OLEDs are emerging as one of the next-generation light sources.

Organic Semiconductors

Molecular Orbitals

The model of the molecular orbitals is employed in the presence of molecular systems. According to such model, when two atoms with the same energy are brought to interact, their energies are splitted, creating two different molecular energy levels: one with lower energy than the original ones and another with higher energy. Such new molecular orbitals are impossible to calculate exactly for most molecules, and so an approximation called linear combination of atomic orbitals (LCAO) is used for that purpose. According to LCAO, in proximity of an atom, the

molecular orbital can be considered as the one of the isolated atom. In the case of two hydrogen atoms, the wave function would be as follows:

$$\Psi_{\pm} = \Psi_{1s}(A) \pm \Psi_{1s}(B) \quad (1)$$

$$\Psi_{1s}(A) = \sqrt{\frac{1}{\pi a_0^3}} \cdot e^{-r_A/a_0} \quad (2)$$

where A and B are the atoms and r_A the distance between electron and atom A . The wave function for atom B is equal to Eq. 2 but with r_B as the distance between electron and atom. By linearly combining the atomic wave functions in Eq. 1, two molecular orbitals have been obtained: the *bonding orbital* (Ψ_+) and the *antibonding orbital* (Ψ_-).

In Fig. 1, the generation of molecular orbitals energetic levels is shown.

The energies of the two molecular orbitals are as follows:

$$E_+ = \frac{\beta - \gamma}{1 - \alpha} \quad (3)$$

$$E_- = \frac{\beta + \gamma}{1 + \alpha} \quad (4)$$

Where β is the Coulombic integral, which represents the energy that the electron has on the isolated atoms; γ is the exchange (or resonance) integral, which represents the interaction between the electrons; and α the amplitude of the wave function.

Optical Properties of Organic Molecules

As a first-order approximation, the optical absorption and emission spectra of organic molecules are very similar to the spectra in the gas phase or in solution (apart from a solvent shift) owing to the weak electronic delocalization. In particular,

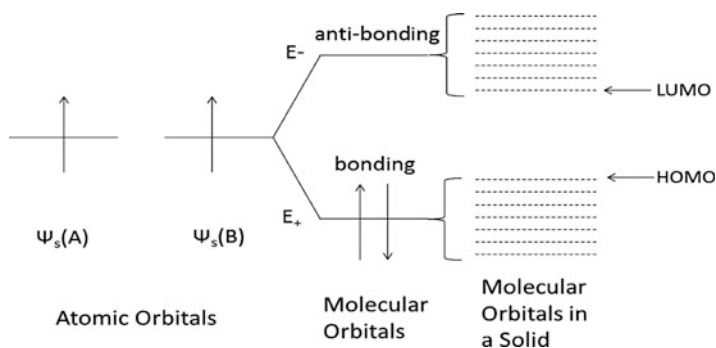


Fig. 1 Scheme of energetic levels of two isolated atoms, a bi-atomic molecule, and a solid

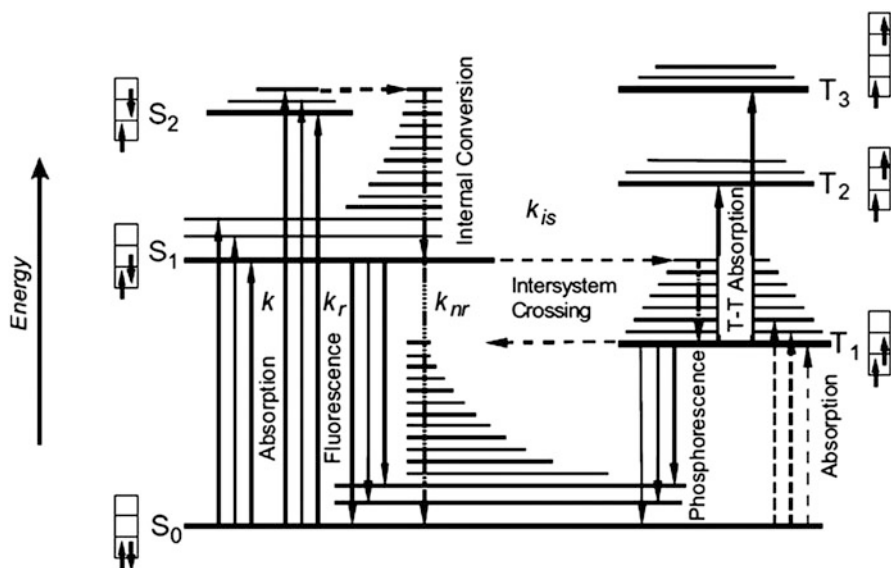


Fig. 2 Energy levels of an organic molecule (Pope and Swenberg 1982)

intramolecular vibrations play an important role in solid-state spectra, and often these vibronic modes can be resolved even at room temperature. Figure 2 shows the excitation diagram of an organic molecule. In an absorption process, an electron is excited from the electronic ground state to the electronic excited state. The process can either end in a state without a vibronic quantum or lead to the excitation of one or several of these quanta. It is shown by the Kasha's rule that the excitation in higher vibronic states will quickly relax to the lowest vibronic state of the electronically excited state. In the emission process, the electron can once more go directly to the electronic ground state without vibrations but can also end in the electronic ground states generating one or more vibrational quanta. It is thus easily recognizable that this behavior leads to a mirror symmetric optical absorption and emission spectrum. Between the two spectra, the Stokes shift is caused by the reorganization of the molecule after the electronic excitation.

As a consequence of the weak electronic delocalization, organic semiconductors have two important peculiarities as compared to their inorganic counterparts. One is the existence of well-defined spin states (singlet and triplet) as in isolated molecules which has important consequences for the photo-physics of these materials (see Fig. 2). A second important difference originates from the fact that an excitation is spatially rather localized due to the small dielectric constant of organic solids (typically $\epsilon = 3-5$); the binding energy between electron and hole is rather large ($E_X = 0.2-0.5$ eV) and definitely much larger than $k_B T$ for room temperature. This has severe consequences for the optical properties: (1) the difference between the

highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), that is, the transport gap, is significantly (by the exciton binding energy E_x) larger than the optical gap and (2) the generation of free carriers by thermal dissociation of the exciton or even electric fields is usually not possible.

Energy Transfer and Loss Processes

The energy transfer in organic molecules could take place via the interaction between an excited state and a ground state, without emitting a photon when transferring energy in two possible ways, either by Förster energy transfer or by Dexter energy transfer (Misra et al. 2006). The Förster energy transfer involves non-radiative dipole–dipole coupling between the host (donor) and guest (acceptor) molecules, where a host molecule, initially in its electronic excited state, may transfer energy to a guest molecule. The efficiency of this energy transfer is inversely proportional to the sixth power of the distance between donor and acceptor. Basically this energy transfer is diffusion of excitons from donor to acceptor having a long-range separation of about $\sim 30\text{--}100$ Å. Emission of the donor is absorbed by the acceptor, and it is favored by the spectral overlapping of donor emission and acceptor absorption spectra. Only singlet–singlet transition when undergoing the Coulombic interaction is allowed at low acceptor concentrations and at a much faster rate of $<10^{-9}$ s (Fig. 3). The Coulombic interaction will not involve the triplet–triplet energy transfer because that violates the Wigner spin conservation law. However, Dexter energy transfer is a short-range, collisional or exchange energy transfer which is a non-radiative process with electron exchange between host and guest molecules. Besides the overlap of emission spectra of host and absorption spectra of guest, the exchange energy transfer needs the overlap of wave functions. In the popular words, it needs the overlap of the electron cloud. The overlap of wave functions also implies that the excited donor and ground-state acceptor should be close enough so the exchange could happen. Therefore, Dexter energy transfer has a short-range separation of $\sim 6\text{--}20$ Å. The Dexter energy transfer allows both singlet–singlet and triplet–triplet energy transitions (Fig. 3). For Förster/Dexter energy transfer, the separation of the host–guest molecules is of prime importance, and it will correspond to an optimal dye concentration of ~ 1 % for fluorescent dyes and ~ 10 % for phosphorescent dyes.

Electroluminescence

Once the exciton is formed by the recombination of electrons and holes, it returns to its ground state via several relaxing mechanisms. The radiative decay processes from singlet (S_1) and triplet (T_1) states to the ground state give rise to electroluminescence phenomena. The former process ($S_1 \rightarrow S_0$) is known as *fluorescence* and has a decay

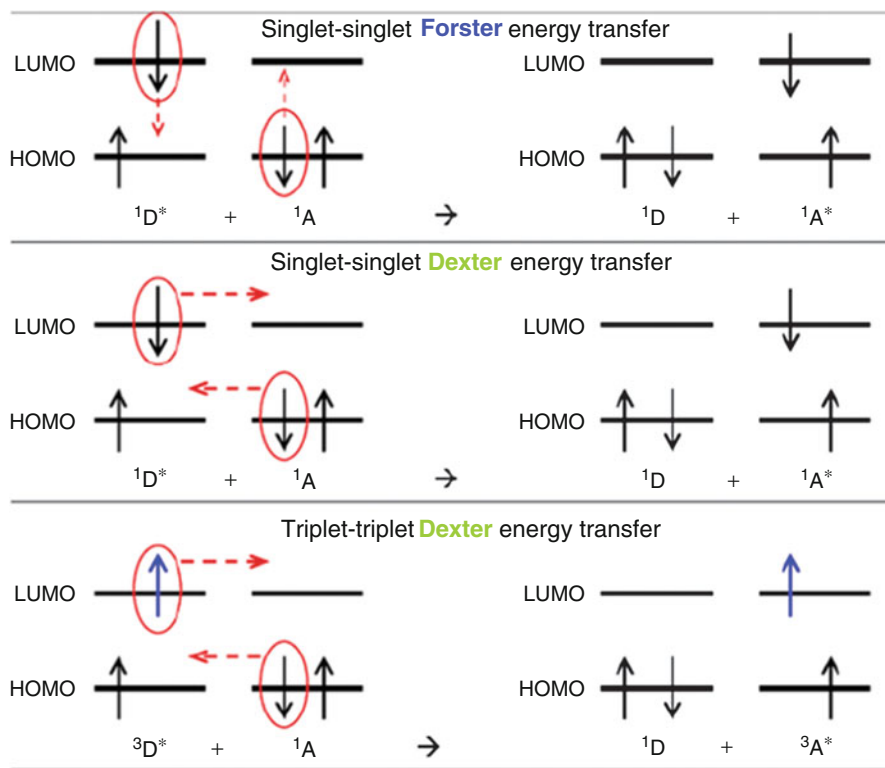


Fig. 3 Schematic diagram for Förster and Dexter energy transfers. http://chemwiki.ucdavis.edu/Theoretical_Chemistry/Fundamentals/Dexter_Energy_Transfer

time of the order of ns (10^{-9} s); the latter ($T_1 \rightarrow S_0$) is known as *phosphorescence* and has a decay time of the order of μ s-ms ($10^{-6} - 10^{-3}$ s). The reason for such higher decay times in the case of phosphorescence is that this process requires the system to change spin, as it has to relax from the triplet state T_1 to the ground singlet state S_0 . Transitions between two states with different spins are much less probable than transitions between two states with the same spin.

Since van der Waals forces determine the structure of organic semiconductors on the nanoscale due to the amorphous, disordered films (Pope and Swenberg 1999), charges are injected statistically with respect to their electron spin, finally determining the formation of singlet and triplet excited states (Reineke et al. 2013). Statistically, there is a 25 % probability of forming a singlet state and 75 % probability of forming a triplet state (Baldo et al. 1999a). The low singlet fraction causes OLEDs based on fluorescent emitter molecules to be rather inefficient with an upper limit of the internal quantum efficiency (IQE) of 25 % because emission solely occurs in its singlet manifold. However, the efficiency of OLEDs was drastically improved with the introduction of phosphorescent emitter molecules (Baldo et al. 1998; Ma et al. 1998; Reineke and Baldo 2012), which are organometallic complexes

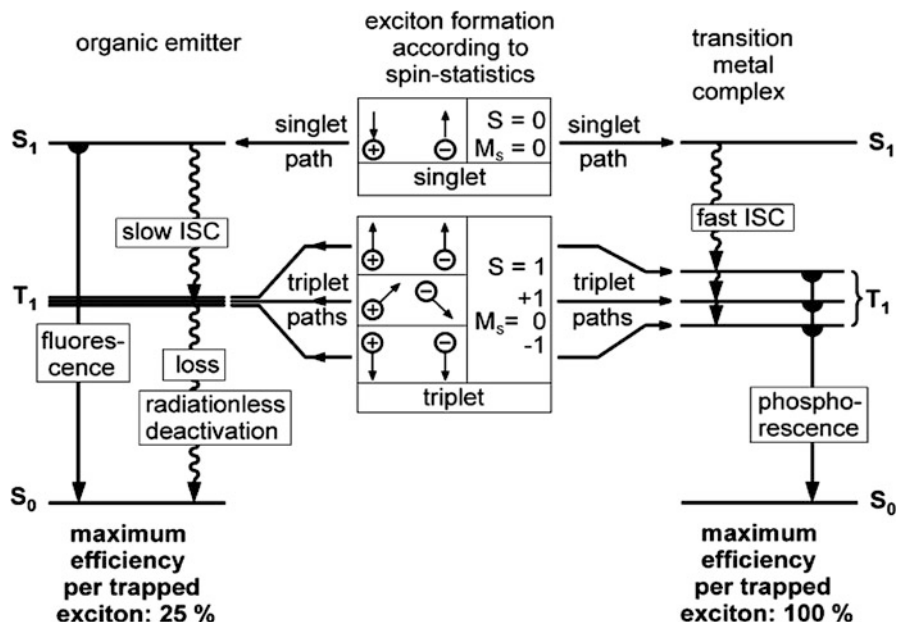


Fig. 4 Fluorescence and phosphorescence EL processes in an organic molecule under electric excitation

comprising a heavy metal atom such as iridium, platinum, palladium, etc. The heavy metal effect allows the enhanced spin-orbit interaction experienced by the molecule, weakening the selection rules for previously forbidden, radiative transitions in the triplet manifold of the molecule and facilitating intersystem crossing (ISC, 10^{-12} s) (Yersin et al. 2002; Thompson 2007), a process which mixes the singlet and triplet character of excited states. This reduces the lifetime of the triplet state (Baldo et al. 1999b); therefore, phosphorescence is readily observed in the room temperature. The fractions of singlet excitons that are created under electrical excitation are efficiently converted into triplet states before they can recombine radiatively. The ISC rate is close to unity in various phosphorescent systems (Kawamura et al. 2005, 2006). Therefore, phosphorescent materials in OLEDs can lead to IQE of 100 % (Fig. 4).

Current Status of White OLED Materials

Generally, white emission requires the mixture of complementary (e.g., blue and yellow or orange) or primary colors (red, green, and blue) or a single copolymer containing different emitting chromophores. More recent developments have been driven by the need for saturated red, green, and blue emission colors as well as higher efficiencies and longer operating lifetimes for white OLED lighting.

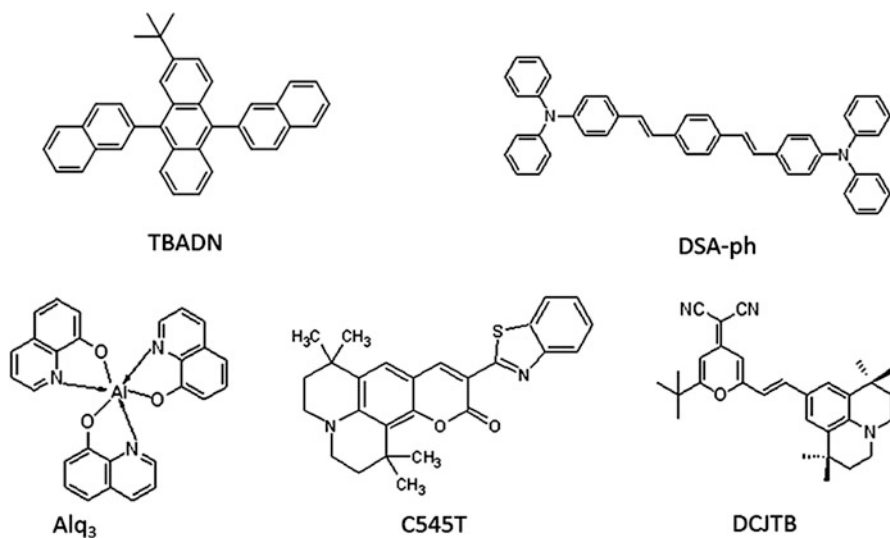


Fig. 5 Molecular structures of some classic fluorescent materials

Fluorescence Materials

Fluorescent materials that can be used to achieve white emission normally exhibit high quantum efficiency. Many fluorescent materials undergo strong concentration quenching (Swanson et al. 2003; Xie et al. 2003) so that the designated use of a material (as a pure film or dispersed in a matrix) is often determined by its photophysical properties. Figure 5 shows the molecular structures of some classic fluorescent materials.

There are a number of stable blue materials that have been disclosed in the literature which can be roughly categorized into several major classes of compounds, namely, *diarylanthracene* (Shi and Tang 2002), *di(styryl)arylene (DSA)* (Hosokawa et al. 1995), *fluorine* (Saitoh et al. 2004), and *pyrene* (Yeh et al. 2004). The 9, 10-(2-naphthyl)anthracene (ADN) and its derivatives, 2-(*n*-butyl)-9,10-(2-naphthyl)anthracene (TBADN), 2,6-(*n*-butyl)-9,10-(2-naphthyl)anthracene (DTBADN) and 2,6-(*n*-butyl)-9,10-[6-(*n*-butyl)-2-naphthyl]anthracene (TTBADN), and 2-methyl-9,10-(2-naphthyl)anthracene (MADN), can be used as deep-blue emitters or host (Wen et al. 2005). Both MADN and TBADN have better efficiency and stability than those of ADN. Low-temperature PL studies revealed also that both MADN and TBADN have different vibronic levels which shift their luminescence peaks significantly enough that their Commission Internationale De l'Eclairage (CIE) color coordinates move to deep blue. The MADN-based device showed the best performance with luminous efficiency of 1.4 cdA^{-1} and the lowest drive voltage of 6.2 V (measured at 20 mAcm^{-2}) with also the bluest color of (0.15, 0.10). Another famous deep-blue material known as *N,N*-bis-(1-naphthyl)-*N,N*-diphenyl-1,1-biphenyl-4,4-diamine (DPVBi) with a wide bandgap energy of 3.1 eV, which is

similar to that of the ADN series, in the OLED community was first published by Hosokawa and coworkers (1995). The key to its molecular design is the terminally substituted (phenyl)styryl group which is supposed to lock the styryl-conjugated chromophore in the planar conformation by twisting the extra bulky phenyl group out of the plane assumed by the (phenyl)styrene chromophore. Its derivative 9,10-di[4-(2,2-diphenylvinyl)phenyl]anthracene (DPVPA)-based device exhibited an external quantum efficiency (EQE) of 3 %, which is more efficient than MADN (EQE \sim 1.5 %), but its color appears greener with a (0.14, 0.17). Diphenyl-[4-(2-[1,1;4,1]terphenyl-4-yl-vinyl)-phenyl]-amine (BD-1) is another example for deep-blue emission. The spectra of undoped and BD-1-doped devices exhibit one dominant peak at 440 nm with a full-width at half-maximum (FWHM) of 56 and 452 nm with an FWHM of 60 nm, respectively. The coordinates of the BD-1-doped MADN devices is (0.15, 0.11). At the optimal doping concentration of 3 %, BD-1 device achieved an EL efficiency of 2.2 cdA^{-1} at 20 mAcm^{-2} with an EQE of 2.3 % (Hung and Chen 2002).

Although the above deep-blue materials show better stability and color quality, the device efficiency was too low. Therefore, tetra(*t*-butyl)perylene (TBP), which is one of the most stable sky-blue dopant materials, was developed (Hung and Chen 2002). The best device performance obtained by doping 0.5 %v/v in MADN was 3.4 cdA^{-1} with a CIE (0.13,0.20). The other stable sky-blue dopant diphenylamino-di(styryl)Arylene(DSA-Ph) has an absorption of 410 nm and a fluorescence of 458 nm and a LUMO/HOMO level of 2.7/5.4 eV with a bandgap energy of 2.7 eV (Wen et al. 2005). The device with 3 % DSA-Ph doped in MADN achieved a luminance efficiency of 9.7 cdA^{-1} and 5.5 lmW^{-1} at 20 mAcm^{-2} with a (0.16, 0.32) while in DPVPA, it yielded 10.2 cdA^{-1} and 4.8 lmW^{-1} with a (0.16, 0.35) with a lifetime of continuous DC operation of $>30,000 \text{ h}$ and the shelf storage stability is $>500 \text{ h}$ at 85°C . Moreover, the color quality was further improved in a highly efficient blue OLED based on a sterically hindered fluorescent host material of tetra (otolyl) pyrene (TOTP). Doped with DSA-Ph of matching LUMO/HOMO, the new TOTP produces blue device with luminance efficiency of 8.64 cdA^{-1} at 20 mAcm^{-2} and 7.1 V with a CIE(x,y) color coordinate of (0.15, 0.28) (Yeh et al. 2004). However, most people in the OLED materials research community should have realized by now that the best dopant/host molecules are probably not what were disclosed in the open literature (e.g., BCzVBi/DPVBi) (Hosokawa et al. 1995).

The most famous green fluorescent material is tris(8-hydroxyquinoline) aluminum (Alq_3) (Tang and Vanslyke 1987). It was widely used in OLEDs as an emitter or an electron-transporting material. However, the low efficiency limits its application. One of the best green dopants is 10-(2-benzothiazolyl)-1,1,7,7-tetramethyl-2,3,6,7-tetrahydro-1H,5H,11H-[1]benzo-pyrano[6,7,8-ij]quinolizin-11-one, known as C545T which belongs to the highly fluorescent class of coumarin laser dyes (Fox and Chen 1988). Later, the Kodak group discovered that by substituting *t*-butyl groups at the benzothiazolyl ring as in C545TB (Chen et al. 2000a), the concentration quenching problem could be further suppressed, and the thermal property was also greatly improved (T_g enhanced to 142°C from 100°C) without compromising its emissive color. In addition, the luminance efficiency could be significantly

increased from 10.5 cdA^{-1} (C545T) to 12.9 (C545TB) at a drive current density of 20 mAcm^{-2} and with a CIE of (0.30, 0.65). One of the more interesting results in subsequent research was found in C545MT, where an extra methyl group was substituted at the C-4 position of C545T (Hung and Chen 2002). When doped in Alq_3 as green emitter in OLED, C545MT has the unusual property of resistance to concentration quenching and sustaining of its EL luminance efficiency (7.8 cdA^{-1}) over a wide range of doping concentration from 2 % to 12 %, which is more than 10 times that of C545T.

Among the RGB dopants used in OLEDs, red emission, due to its low efficiency, remains to be the weakest link in realizing the full potential of an OLED display and solid-state lighting. One of the best that comes close is 4-(dicyanomethylene)-2-*t*-butyl-6-(1,1,7,7-tetramethyljulolidyl-9-enyl)-4H-pyran better known as DCJTB (Chen et al. 1999). An alternative dopant in the same class is the *i*-propyl derivative DCJTI (Chen et al. 2000b), which is easier to synthesize and more amenable to large-scale production without compromising its EL efficiency and chromaticity. Recently, it was reported that an 8-methoxy-substituted derivative (DCJMTB) doped in tris(8-hydroxyquinolinolato)gallium (Gaq_3) host matrix at 1 % achieved a luminance efficiency of 2.64 cdA^{-1} at 20 mAcm^{-2} with a power efficiency of 0.72 lmW^{-1} and CIE of (0.63, 0.36) (Chen et al. 2001). One of the more exciting developments in red emitter was recently disclosed by Sony, 1,10-dicyano-substituted bis-styrylnaphthalene derivative (BSN) (Hung and Chen 2002). The BSN has a film absorption at 507 nm and a strong PL at 630 nm with a quantum efficiency of 0.80. It has good thermal properties with $T_g \sim 115 \text{ }^\circ\text{C}$, $T_c \sim 161 \text{ }^\circ\text{C}$, and $T_m \sim 271 \text{ }^\circ\text{C}$ and was reported to form a good amorphous thin film on evaporation. Solid-state photo-ionization measurements place its LUMO at 2.93 eV and HOMO at 5.38 eV which favors the efficient recombination of electrons and holes to take place within the emitter layer. BSN as a red emitter without dopants displays an impressive luminous efficiency of 2.8 cdA^{-1} at 500 cdm^{-2} with CIE (0.63, 0.37). There were many studies on getting sharp and saturated emission from europium complexes by varying the ligand designs. These dopants, however, usually have low efficiencies and brightness. Representative examples are europium tris(dibenzoylmethide)(triphenylphosphine oxide) ($\text{Eu}(\text{DBM})_3(\text{TPPO})$) (Chen et al. 2001), europium tris(thenoyltrifluoroacetone)(phenanthroline)($\text{Eu}(\text{TFA})_3(\text{phen})$) (Capecchi et al. 2000), and europium tris(dibenzoylmethide)(1-ethyl-2-(2-pyridyl)benzimidazole)($\text{Eu}(\text{DBM})_3(\text{EPBM})$) (Huang et al. 2001).

Phosphorescence Materials

The luminous efficiency of OLEDs can be potentially improved by up to a factor of four when phosphorescent emitters are used in place of the fluorescent emitters since these compounds containing heavy metal atoms efficiently spin-orbit, coupling results in a mixing of singlet and triplet states (Reineke et al. 2013). As both singlet and triplet states are utilized, this may give rise to a theoretical maximum IQE of 100 %. Based on spin-statistics consideration, the modern approach in OLED

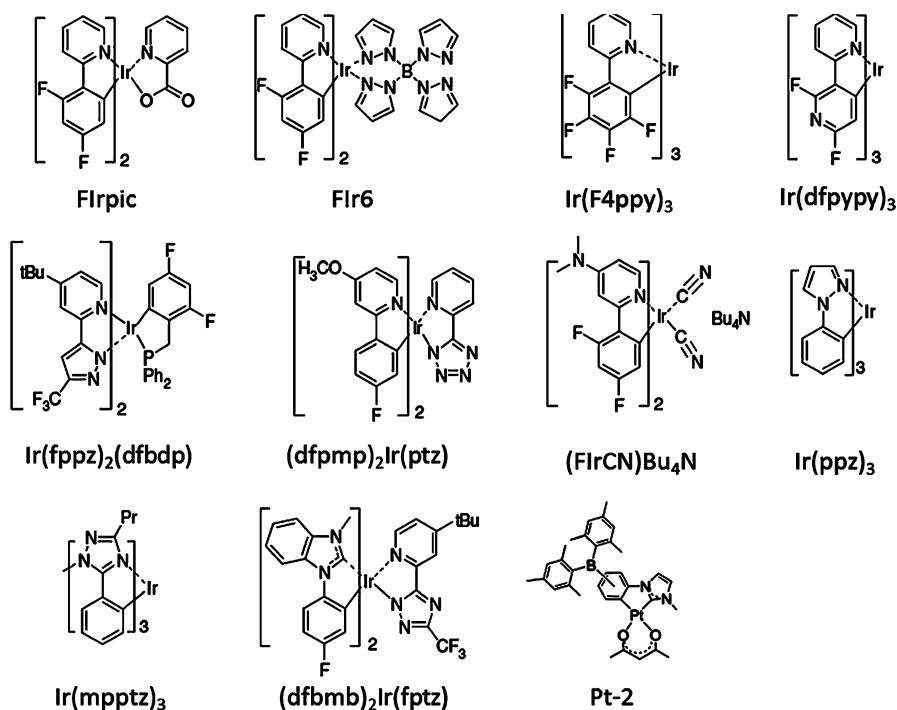


Fig. 6 Chemical structures of blue phosphorescent materials

technology has focused more on triplet-based phosphorescent complexes since the first phosphorescent material reported.

Cyclometalated iridium (III) complex is one of the best candidates of phosphorescent dyes with high quantum yields due to the short lifetime of triplet excited states. Bis[2-(4,6-difluorophenyl)pyridinato-*C*²,*N*](picolinato)iridium(III) (FIrpic) (Holmes et al. 2003) and iridium(III) bis(4,6-difluorophenylpyridinato) tetrakis(1-pyrazolyl)-borate (FIr6) (Zheng et al. 2008) have been considered as excellent dopants for greenish-blue and sky-blue emitters. However, their emissions are far from a saturated blue color, with the CIE of (0.17, 0.34) and (0.16, 0.26), respectively. Therefore, several strategies have been applied to broaden the gap of iridium complex and tune the emission into deep blue (Xiao et al. 2011), including (a) decreasing the HOMO level through the introduction of electron-withdrawing groups, (b) changing the ancillary ligand to be more electron accepting, (c) increasing the LUMO level through the addition of electron-donating groups to the pyridine ring or replacing the pyridine ring to *N*-heterocyclic ligand with a higher LUMO, and (d) using a strong σ -donating ligand. Several corresponding Ir (III) complexes are shown in Fig. 6.

Ragni et al. (2006) prepared a blue-emitting iridium (III) complex containing fluorine-substituted phenylpyridine ligands, tris[2-(3',4',5',6'-tetrafluorophenyl)pyridinato-*N*,*C*²]iridium(III) (Ir(F4ppy)₃) giving the emission maximum at

471 nm and an η_{ext} of 5.5 %. Recently, Lee et al. (2009) reported a *fac*-tris-(2',4'-difluoro-2,3'-bipyridinato-*N,C4'*)iridium(III) ($\text{Ir}(\text{dfppy})_3$) containing fluorine-substituted bipyridine ligand. The replacement of the phenyl ring in ppy with pyridine and the introduction of fluorine substituents into the pyridine ring significantly improved the thermal stability and molecular rigidity due to intermolecular interactions in the solid state. Chiu et al. (2009) designed a series of iridium (III) complexes of $\text{Ir}(\text{fppz})_2(\text{dfbdp})$ by changing ppy ligand with larger π - π^* energy gap of ppz together with weak-conjugated and strong ligand-field strength benzyl phosphine ($\text{C}^{\wedge}\text{P}$) chelate and obtained a deep-blue light with CIE coordinates of (0.15, 0.11). Wu et al. (2007) successfully expanded the E_g of Ir (III) complex by introducing a methoxy group into the ligand of 2-difluorophenylpyridine and obtained a blue emitter of bis[2-(3',4'-difluorophenyl)-4-methoxypyridinato-*N,C2'*]Ir(III) [5-(2'-pyridyl)tetrazolate] $(\text{dfpmp})_2\text{Ir}(\text{ptz})$ with the emission peak at 452 nm, corresponding to CIE coordinates of (0.18, 0.13). Censo et al. (2008) tuned the color of Ir(III) complex by changing the ligands with methylamine donor-substituted 2-phenylpyridine on the pyridine moiety and synthesized an anionic complex of $(\text{C}_4\text{H}_9)_4 \text{N}[\text{Ir}(2\text{-(2,4-difluorophenyl)-4-dimethylaminopyridine})_2(\text{CN})_2]$ (FIrCN) Bu_4N with a phosphorescence quantum yield of 0.62 and a photoluminescence shift to 451 and 471 nm. Tamayo et al. (2003) replaced the pyridine ring to *N*-pyrazole with a higher LUMO and obtained tris(phenylpyrazolyl)-iridium (III)[$\text{Ir}(\text{ppz})_3$] which exhibited photoluminescence at 414 nm for its facial isomer and 427 nm for its meridional isomer at 77 K, with no observable emission at RT. However, when the pyridine ring was replaced by triazole, which has a higher LUMO energy, to obtain tris-(1-methyl-5-phenyl-3-propyl-(Chiang et al. 1977; Tang and Vanslyke 1987; Tang 1986)triazolyl)iridium (III)[$\text{Ir}(\text{mpptz})_3$], photoluminescence at 449 and 479 nm with a high quantum yield (0.66) was achieved (Lo et al. 2006). The efficiency was further increased (0.94) when a dendron structure was used (Lo et al. 2009).

By using strong σ -donating ligands such as carbenes (Sajoto et al. 2005; Holmes et al. 2005), the meridional isomer of tris(phenyl-methylbenzimidazolyl)iridium (III) ($\text{Ir}(\text{pmb})_3$) tuned the electroluminescence emission wavelength to 395 nm corresponding to CIE coordinates of (0.17, 0.06). However, the quantum yield is quite low (0.002). One of the critical criteria in obtaining highly efficient blue phosphorescence is to decrease the radiative lifetime which may increase the emission quantum yield. Conversely, care should also be taken to keep good blue color chromaticity and avoid enhancing the radiationless decay pathways arising from enlargement of the bandgap. By decreasing the energy difference (ΔE) between $^1\text{MLCT}$ and ^3LC , Chang et al. (2008) developed a weak-conjugated benzyl carbene ligand to synthesize a complex of $(\text{dfbmb})_2\text{Ir}(\text{fptz})$ ($\text{dfbmb} = 1\text{-(2,4-difluorobenzyl)-3-methylbenzimidazolium}$) with a high quantum yield (0.73).

Although other metal complexes have been reported to give blue emission at room temperature, the color purity and efficiency are not comparable to those Ir(III) complexes (Brooks et al. 2002; D'Andrade and Forrest 2003; D'Andrade et al. 2002). However, they still provide a paradigm to develop a new area of phosphorescent blue emitters. Efficient blue OLEDs based on 1,3-difluoro-4,6-di

(2-pyridinyl)-benzeneplatinum(II) chloride (Pt-1) were reported in 2008 (Yang et al. 2008). Its solution emission wavelength is in the narrow range of 470–520 nm with η_{PL} of 0.46. Blue-emitting devices based on this complex showed peak η_{ext} of 16 %, power efficiency of 20 lm W⁻¹, and CIE coordinates of (0.15,0.26), which are comparable to some Ir complexes. Very recently, Wang et al. reported two blue-emitting Pt(II) complexes (Pt-2 and Pt-3) by combining the emissive and electron-transporting properties of the Lewis acidic triarylboron group with the strong ligand field of *N*-heterocyclic carbene (Hudson et al. 2012). The device incorporating Pt-2 as an emitter showed peak current efficiency and power efficiency values of 53.0 cd A⁻¹ and 41.6 lm W⁻¹, respectively, which remained as high as 49.6 cd A⁻¹ and 33.6 lm W⁻¹ at the display-relevant brightness of 100 cd m⁻². The device based on the blue-emitting Pt-3 also showed an attractive performance, with peak current efficiency and power efficiency of 25.8 cd A⁻¹ and 22.5 lm W⁻¹, which can be kept at 19.2 cd A⁻¹ and 13.6 lm W⁻¹ even at 100 cd m⁻².

There are a number of Cu and Zn complexes that emit blue light and are potentially useful for blue OLEDs. The advantage of these complexes is the low cost and considered to be of prime importance (Shinar and Shinar 2008). Due to the high cost and limited availability of d⁶ and d⁸ transition metal complexes, extensive research efforts have been directed toward the discovery of less expensive alternatives (Barbieri et al. 2008). However, the device performance-based thesis complexes show much poorer compared to Ir- and Pt-based complexes (Son et al. 2008; Liu et al. 2011; Deaton et al. 2010; Si et al. 2009; Zhang et al. 2009; Manbeck et al. 2011; Czerwieńiec et al. 2011).

As shown in Fig. 7, the Ir (III) complexes have also been widely used as the highly efficient green phosphorescent dyes since the first most famous report of Ir (ppy)₃-based green phosphorescent OLED due to its high spin-orbit coupling in the presence of Ir (III) (Baldo et al. 1999b). Though Ir(ppy)₃-based devices can achieve high efficiencies, a high doping concentration of Ir(ppy)₃ causes self-quenching as well as triplet-triplet annihilation. To avoid the self-quenching, a modification to ppy has been performed to obtain Ir(x-ppy)₃. In 2001, Xie and his coworkers reported a phosphorescent tris(2-(4-tolyl)phenylpyridine)iridium (Ir(mppy)₃) by introducing a pinene group as the sterically hindered spacer in the backbone of 2-phenylpyridine to reduce bimolecular interactions and suppress self-quenching (Xie et al. 2001). Jung et al. (2004) synthesized dimethyl-substituted tris(pyridylphenyl)iridium(III) derivatives Ir(dmppy)₃ and found that the substitution position of methyl groups caused a variation in the electroluminescence. Kang et al. (2008) used rigid and bulky cyclometalation ligands of [2-(1-cyclohexenyl)pyridine (chpy) and 2-(3-methyl-1-cyclohexenyl)pyridine (mchpy)] to suppress triplet-triplet annihilation and achieved an efficiency of 18.7 %, 69.0 cdA⁻¹ and 62.0 lmW⁻¹. In addition, the triplet-triplet annihilation and roll-off can also be reduced by the adoption of energy well-matched host and dopant as well as a buffer layer as reported by Mi and his coworkers, who used tris[3,6-bis(phenyl)-pyridazinato-*N1,C2'*]iridium Ir(BPPya)₃ as the phosphorescent dye (Gao et al. 2008). To tune the wavelength of emission, Zhou et al. (2008a) synthesized tris-cyclometalated homoleptic iridium (III) complexes

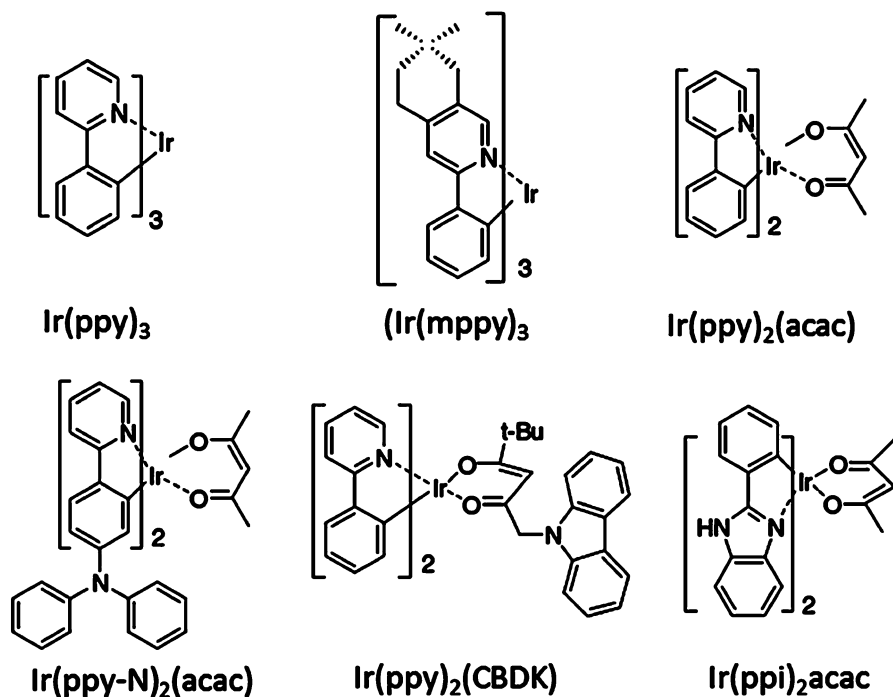


Fig. 7 Chemical structures of green phosphorescent materials

[Ir(ppy-X)₃] (X = SiPh₃, GePh₃, NPh₂, POPh₂, OPh, SPh, and SO₂Ph) by introducing functional groups onto the phenyl ring of ppy and obtained a high efficiency of 13.9 %, 60.8 cdA⁻¹, and 49.1 lmW⁻¹ with bluish-green to yellow-green emission. Although triarylsilyl moieties are more chemical stable and have higher efficiency for emission than trialkylsilyl, the later one shows high steric bulk and thermal stability. Jung et al. (2009) reported a homoleptic Ir (III) complex of *fac*-tris[2-(3'-trimethylsilylphenyl)-5-trimethylsilylpyridinato]iridium [Ir(dsippy)₃] with narrow green emission (FWHM = 50 nm) and higher efficiency than Ir(ppy)₃. The bulky silyl group on the ppy ring seems to play a key role to suppress various intermolecular excited-state interactions, which illustrates a way to solve the problem of color purity for green electrophosphorescence.

Owing to the strong spin-orbit coupling of Ir (III) complex, a mixture of ³MLCT with ³(π-π*) of ligand states leads to allow phosphorescent emission. Forrest and Thompson et al. (Lamansky et al. 2001) reported a series of cyclometalated iridium (III) complexes, (C^N)₂Ir(acac) [(acac) = acetylacetonate], one of which is [Ir(ppy)₂(acac)] with an emission peak at 525 nm. To investigate the substitution effect on emission color, Wong and his coworkers (Zhou et al. 2008b) introduced some electron-withdrawing groups (to improve EI and ET capability) onto the phenyl ring of ppy to obtain [Ir(ppy-X)₂(acac), X = SiPh₃, GePh₃, NPh₂, POPh₂, OPh, SPh, and

SO₂Ph]. PhOLED based on Ir-OPh emitted bluish-green light at 505 nm, and devices based on Ir-NPh₂ emitted bright green light at 528 nm. The color tuning method can be applied to multicolor and white light.

To suppress self-quenching, Liu et al. (2008) introduced carbazole-functionalized β -diketonate to improve carrier transport and suppress triplet-triplet annihilation and obtained bis(2-phenylpyridinato-*N,C*₂)iridium(1-(carbazol-9-yl)-5,5-dimethylhexane-2,4-diketonate) [Ir(ppy)₂(CBDK)] and bis(2-(2,4-difluorophenyl)pyridinato-*N,C*₂)iridium(1-(carbazol-9-yl)-5,5-dimethylhexane-2,4-diketonate) [Ir(dfppy)₂(CBDK)] for non-doped devices. In addition, the triplet-triplet annihilation can be reduced by active hydrogen to shorten the excited state lifetime of the complex as reported in bis(2-phenyl-benzoimidazole)iridium (III) acetylacetonate [Ir(ppi)₂acac] (Han et al. 2008). Although many efforts have been made on the color tuning of green phosphors, the purity of highly efficient green phosphors is still in need of further improvement.

The first phosphorescent material used in the red phosphorescent OLED to improve the IQE was a red Pt (II) organic chelate dye. In 1998, Forrest and Thompson as well as their coworkers reported that 2,3,7,8,12,13,17,18-octaethyl-21*H*,23*H*-porphine platinum (II) (PtOEP) could be used as an energy acceptor (guest) doped in the donor of Alq₃ (host) to fabricate red phosphorescent OLED (Baldo et al. 1998). The device generated saturated red emission at 650 nm from the triplet excited state, while emission at ~580 nm from the singlet state was not observed. The peak EQE and IQE reached 4 % and 23 %, respectively. Successively, another two Pt (II) porphyrins of PtOX and PtDPP were synthesized (Kwong et al. 1999). Their devices exhibited strong red electrophosphorescence with narrow line widths. However, the quantum efficiency and the EL spectra changed with current density. The bright saturated red emission was blue-shifted to orange at low current density, and the quantum efficiency decreased at high current density.

In order to achieve a stable red phosphorescent OLED, a phosphor of bis(2-(2'-benzo [4,5-*a*] thienyl)pyridinato-*N, C*₃') metal (acetylacetonate)[btp₂M(acac), M = Ir, Pt] with short lifetime of triplet state was synthesized (Adachi et al. 2001). The lifetime of btp₂M(acac) was short with value less than 10 μ s, minimizing the saturation of triplet emissive states and triplet-triplet annihilation. Therefore, significant improvement to the quantum efficiency was achieved at high current density in comparison to PtOEP, PtOX, and PtDPP.

Ligands, therefore, play an important role in color emission and quantum efficiency. Forrest and Thompson et al. synthesized a series of Ir-organometallic phosphors consisting of two cyclometalated (C[^]N) ligands and a single monoanionic, bidentate ancillary ligands (LX), i.e.,(C[^]N)₂Ir(LX) (Lamansky et al. 2001). Studies of photophysical characterization indicated that the lowest energy (emissive) excited state in these organometallic phosphors was a mixture of ³MLCT and ³(π - π^*) states. The emission color can be tuned from green to red through selection of an appropriate C[^]N ligand. In addition to the larger π -conjugation space, the intramolecular donor-acceptor (D-A) systems can also shift the absorption and emission spectra to the red region to obtain pure red

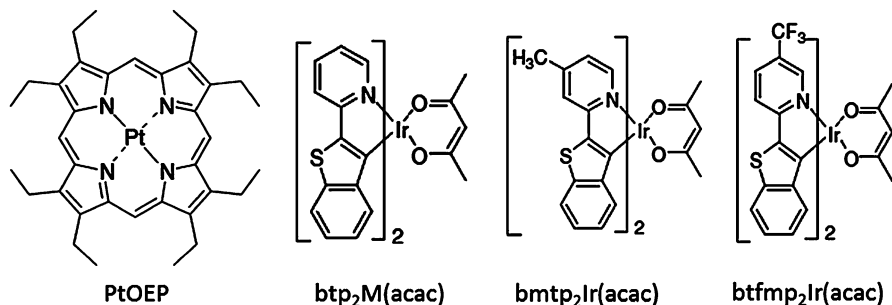


Fig. 8 Chemical structures of red phosphorescent materials

phosphorescence (Tsuboyama et al. 2003). Xu and his coworkers introduced the substituents of CH₃ and CF₃ into the pyridyl ring of btp ligand to tune Ir(III) complexes into the red region (Xu et al. 2007).

To improve charge injection and transport ability of phosphorescent materials, charge transporting groups have been chosen to be incorporated onto key ligands. Zhou et al. reported pure red phosphorescent Ir(III) complexes by using triphenylamine dendrons to chemically modify the key ligands, where the triphenylamine group lifted the HOMO level of the Ir(III) complexes (Zhou et al. 2007). Compared with the tri(phenylisoquinoline)Ir(III), the HOMO of the Ir(III) complexes tailoring phenylisoquinoline with triphenylamine groups was increased from -5.11 eV to -4.96 eV.

In addition to Ir (III) and Pt (II) complexes, there are many organometallic phosphors based on other heavy metal ions, e.g., Os (II), Ru (II), and Re (I), we just discussed here organometallic phosphors based on Ir(III) and Pt (II) as the core ion for red electrophosphorescences. Some representative red Ir complexes are shown in Fig. 8.

Delay Fluorescence Materials

A conventional organic molecule, depicting singlet (S_1) and triplet (T_1) excited states and a ground state (S_0), which the S_1 level, was considerably higher in energy than the T_1 level by 0.5–1.0 eV, because of the electron exchange energy between these levels. However, organic molecules with a small energy gap (ΔE_{ST}) between S_1 and T_1 levels can be achieved by careful design. Thus, the T_1 to S_1 reverse intersystem crossing (ISC) becomes feasible and the light emission can be totally decayed from its singlet. This process is thermally activated delayed fluorescence (TADF) (Uoyama et al. 2012), as shown in Fig. 9. A molecule with efficient TADF requires a very small ΔE_{ST} between its S_1 and T_1 excited states, which can harness both singlet and triplet excitons for light emission through fluorescence decay channels, leading to an intrinsic fluorescence efficiency in excess of 90 % and a very high

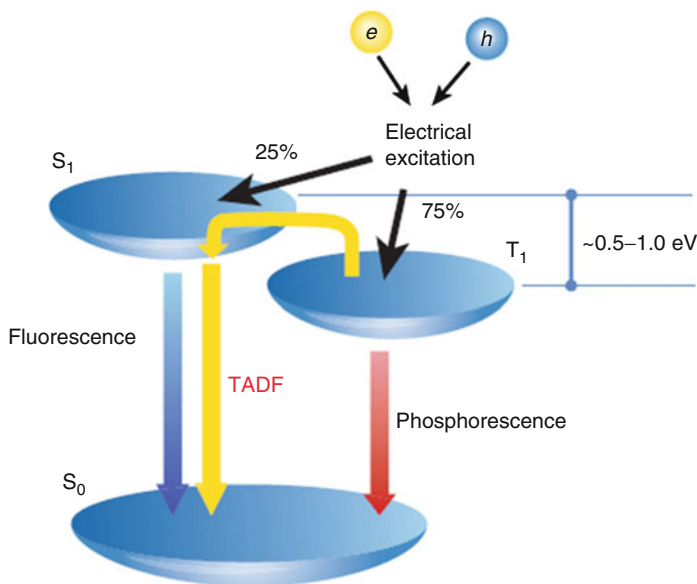


Fig. 9 Energy diagram of a conventional TADF organic molecule

external electroluminescence efficiency, of more than 19 %, which is comparable to that achieved in high-efficiency phosphorescence-based OLEDs.

The first TADF organic materials reported by Adachi et al. were some tin (IV) complexes (Endo et al. 2009). The successful achievement of TADF orange–red emission was assisted by heat, and the emission was increased with increasing the temperature, demonstrating the thermally activated process from T_1 to S_1 . After that, a series of RGB TADF materials at room temperature were designed and synthesized, as shown in Fig. 10.

For blue emission, a series of efficient TADF deep-blue organic materials have been characterized for carbazole/sulfone derivatives in both solutions and doped films by Zhang et al. (2012). At room temperature, all compounds exhibit broad and structureless emission bands with a maximum at 402–419 nm. A pure blue OLED based on this compound demonstrates a very high EQE of nearly 10 % at low current density. Furthermore, a series of highly efficient TADF emitters based on carbazolyl dicyanobenzene, with carbazole as a donor and dicyanobenzene as an electron acceptor, were designed (Uoyama et al. 2012). Because the carbazolyl unit is markedly distorted from the dicyanobenzene plane by steric hindrance, the HOMO and LUMO of these emitters are localized on the donor and acceptor moieties, respectively, leading to a small ΔE_{ST} . Different emissions were achieved. For green emission, a very high external electroluminescence quantum efficiency of 19.3 % was achieved, which is equivalent to an internal electroluminescence quantum efficiency of 64.3–96.5 % assuming a light outcoupling efficiency of 20–30 % (Smith et al. 2004; Tanaka et al. 2007). The orange and sky-blue OLEDs had

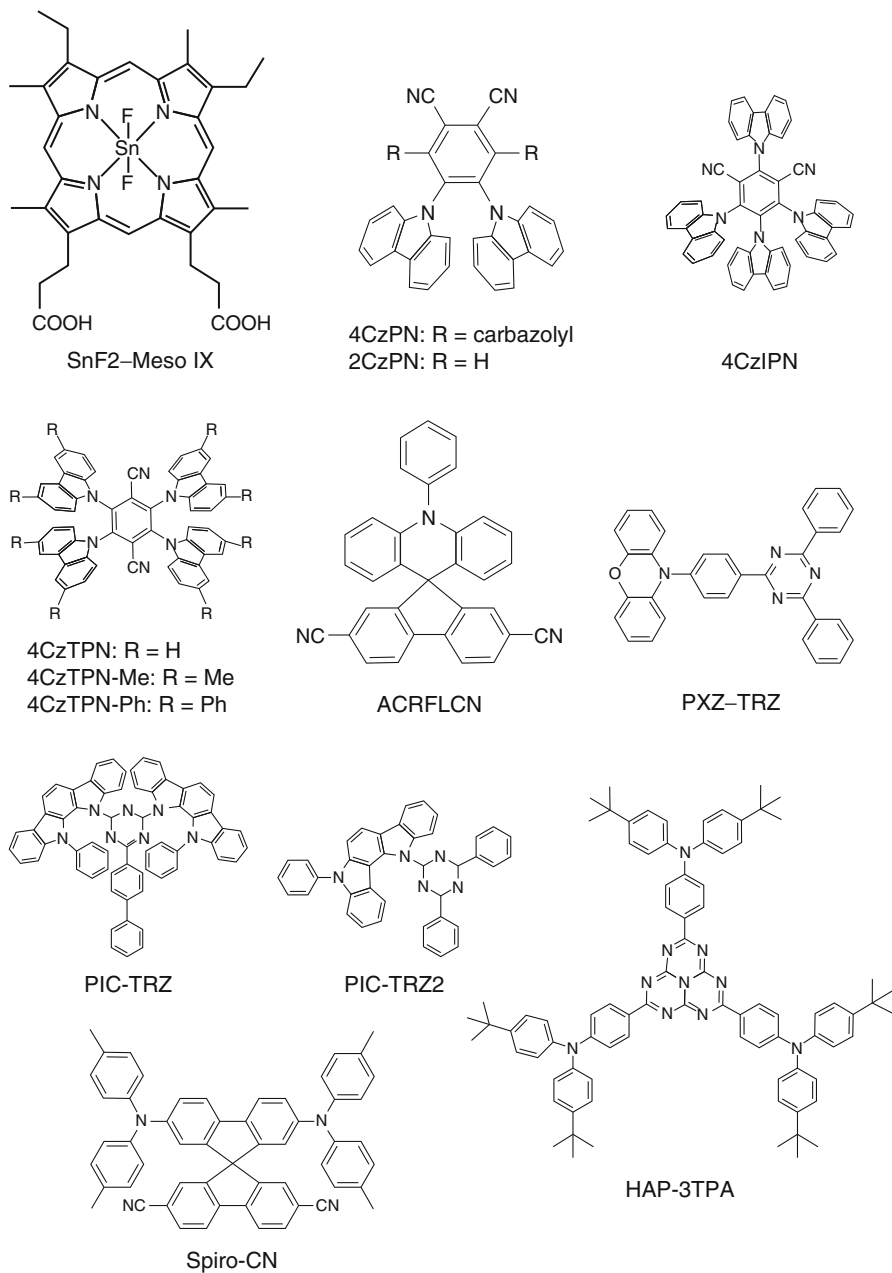


Fig. 10 Some typical RGB TADF organic materials

external electroluminescence quantum efficiencies of 11.2 % and 8.0 %, respectively, which are also higher than those of conventional fluorescence-based OLEDs. Very recently, highly efficient TADF emitters containing 2,5-diphenyl-1,3,4-oxadiazole (OXD) or 3,4,5-triphenyl-4H-1,2,4-triazole (TAZ) electron acceptor and phenoxazine (PXZ) electron donor moieties were developed by Lee et al. (2013). Oxadiazole-based compounds PXZ-OXD and 2PXZ-OXD showed green emission, while the triazole-based ones PXZ-TAZ and 2PXZ-TAZ exhibited sky-blue emission. An OLED using 2PXZ-OXD as an emitter exhibited an EQE of 14.9 %. The peak position of the EL spectra of 2PXZ-TAZ was observed at 456 nm with CIE color coordinates of (0.16, 0.15) and EQE of 6.4 %. TADF properties of a dicarbazole–triazine compound, 9-(4,6-diphenyl-1,3,5-triazin-2-yl)-9*H*-phenyl-3,3*b*-bicarbazole (CzT), and its OLED characteristics with sky-blue emission were also investigated (Serevicius et al. 2013). An estimated small energy gap of about 90 meV between the singlet and triplet energy states of CzT made the up-conversion of triplet excitons back to a singlet state possible. The origin of the observed delayed fluorescence has been shown to be TADF. An OLED with CzT as an emitter showed the maximum EQE of 6 %.

For green emission, they demonstrated an efficient green TADF material activated by a spiro-acridine derivative at room temperature with high η_{PL} values (more than 67 %) for the first time and a large TADF component enables excitons to be generated with high efficiency in OLEDs (Mehes et al. 2012). The OLED achieved a high external quantum efficiency of 10.1 % at a low current density of $3.3 \times 10^{-4} \text{ mA cm}^{-2}$. This value approaches the theoretical maximum, which was calculated to be 12.7 % by a method that was reported (Nakagawa et al. 2012). A high-efficiency purely organic luminescent material, 2,4-bis{3-(9*H*-carbazol-9-yl)-9*H*-carbazol-9-yl}-6-phenyl-1,3,5-triazine (CC2TA) comprising the bicarbazole donor and phenyltriazine acceptor units, was designed and synthesized (Lee et al. 2012). The molecular design of CC2TA allows spatial separation of HOMO and LUMO on the donor and acceptor fragments, respectively, leading to an exceptionally small singlet–triplet exchange energy ($\Delta E_{\text{ST}} = 0.06 \text{ eV}$) together with a high triplet energy. Furthermore, a high external quantum efficiency as high as 11 % has been achieved in the sky-blue OLED employing CC2TA as an emitter. They also found that a spirobifluorene derivative (Spiro-CN) having the donor–acceptor moieties as an emitter gave a strong TADF emission (Nakagawa et al. 2012). This device gives yellow electroluminescence with maximum current and power efficiencies of 13.5 cd A^{-1} and 13.0 lm W^{-1} and a maximum external quantum efficiency of 4.4 %, which are the highest values reported to date for a device containing spirobifluorene as an emitter. Furthermore, a high luminance of $12,000 \text{ cd m}^{-2}$ was observed at 15 V. Very recently, Sato et al. demonstrate an organic molecule with an energy gap between its singlet and triplet excited states of almost zero ($\Delta E_{\text{ST}} = 0 \text{ eV}$). Such separation was realized through proper combination of an electron donating in dolocarbazole group and a diphenyltriazine electron-accepting moiety (PIC-TRZ and PIC-TRZ2). Calculated and measured ΔE_{ST} were 0.003 and 0.02 eV, respectively. A total photoluminescence efficiency of 59 % with 45 % from a delayed component and 14 % from a prompt component was obtained for a doped

film. OLEDs containing this molecule as an emitting dopant exhibited an unexpectedly high external electroluminescence efficiency of $\eta_{\text{EQE}} \sim 14\%$ (Sato et al. 2013).

For red–orange emission, emission wavelength tuning of TADF from green to orange in solid-state films is demonstrated. Emission tuning occurs by stabilization of the intramolecular charge transfer state between a phenoxazine (PXZ) donor unit and 2,4,6-triphenyl-1,3,5-triazine (TRZ) acceptor unit separated by a large twist angle. The emission wavelengths of mono-, bis-, and tri-PXZ-substituted TRZ exhibit a gradual red shift while maintaining a small energy gap between the singlet and triplet excited states. An OLED containing a tri-PXZ-TRZ emitter exhibited a maximum external quantum efficiency of 13.3 % for yellow–orange emission (Lee et al. 2013). Li et al. reported an efficient orange–red TADF emitter, 4,4',4''-(1,3,3a,1,4,6,7,9-heptaazaphenalene-2,5,8-triyl) tris(*N,N*-bis(4-(*tert*-butyl)phenyl)aniline) (HAP-3TPA) (Li et al. 2013). An orange–red OLED incorporating HAP-3TPA exhibits high EL performance with a maximum η_{ext} of 17.5 %, a maximum power efficiency of 22.1 lm W⁻¹, a maximum current efficiency of 25.9 cd A⁻¹, a turn-on voltage of 4.4 V, and a peak luminance of 17,000 cd m⁻² without any light outcoupling enhancement.

Polymer Materials

The key idea of white polymers for white OLEDs is to realize a single copolymer that contains all the different emitting chromophores needed to cover the visible spectrum (Reineke et al. 2013). The advantages of this approach are the simple fabrication, the isotropic yet statistical distribution of the chromophores within the film (Gather et al. 2007), the control of the interspecies energy transfer by the molecular design, and the low probability of phase separation within the film (Berggren et al. 1994; Forrest 2004; Gather et al. 2011). Figure 11 shows some typical white polymer structures.

Tu et al. (2004) reported on an efficient white light-emitting polymer by mixing moieties of an orange fluorophore (1,8-naphthalimide) into the blue PFO main polymer (WP-1). Used as a single emissive layer device, a chromophore concentration of 0.05 % in the PFO main chain yields a device efficiency of 5.3 cdA⁻¹ and 2.8 lmW⁻¹ with CIE of (0.26, 0.36). By changing the orange chromophore to TPABT (WP-2), Wang and coworkers improved the device efficiency to 7.3 cdA⁻¹, 3.34 lmW⁻¹, and 3.8 % EQE even with improved color quality CIE (0.35, 0.34) (Liu et al. 2007a). Lee et al. (2005) were the first to report on a main chain copolymer containing emitting units for the three basic colors of blue (PDHF), green (DTPA), and red (TPDCM) (WP-3). The overall content of green and red chromophores makes up less than 3 % in total. Despite the broad spectrum realized with CIE coordinates (0.34, 0.35), the device efficiency was very low with a maximum current efficiency of 0.04 cdA⁻¹. Chuang et al. (2007) and Luo et al. (2007) used highly efficient fluorescent benzothiadiazole derivatives for green and red chromophores and reached maximum efficiencies of EQE $\sim 2.22\%$ with CIE coordinates of (0.37, 0.36) and a maximum EQE of 3.84 %, respectively.

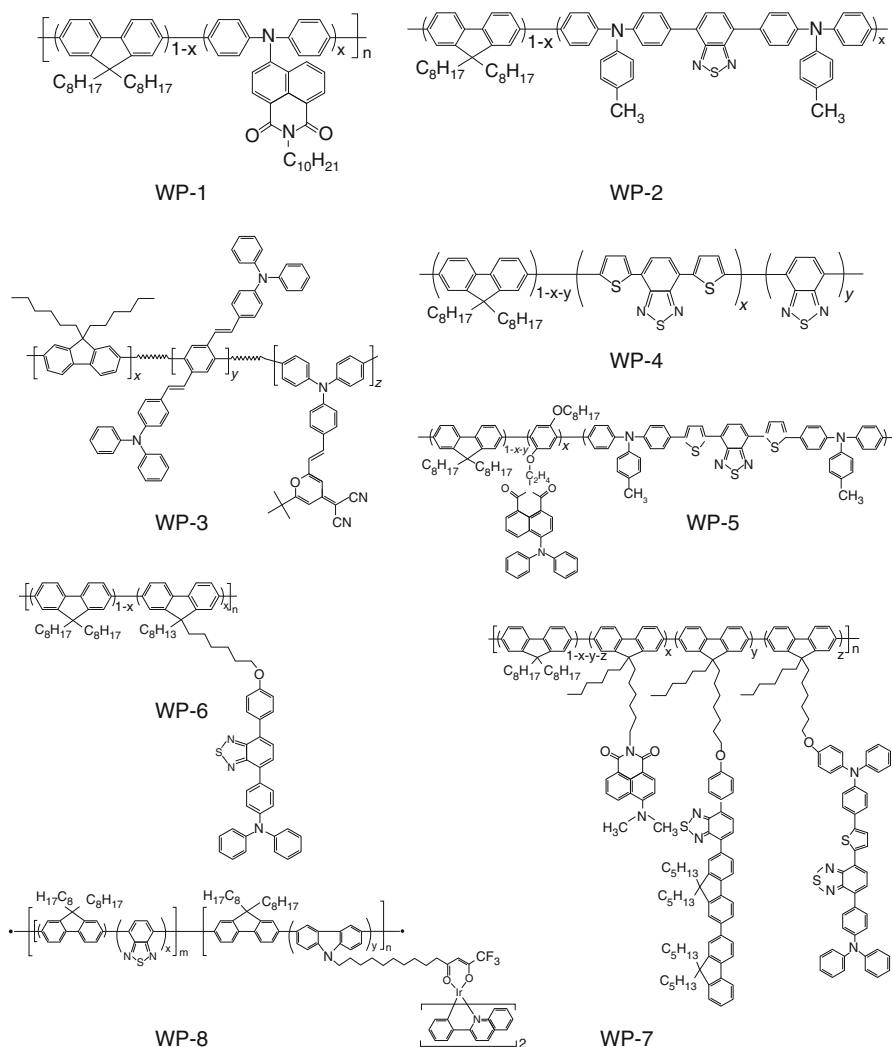


Fig. 11 Chemical structures of white polymers

corresponding to a current efficiency of 6.20 cdA^{-1} with CIE (0.35, 0.34) (WP-4), respectively.

By attaching the emitting units to the main chain via alkyl chains (Liu et al. 2005), Liu et al. synthesized a benzothiadiazole derivative (TPATBT) in the polyfluorene main chain for red emission and additionally a naphthalimide derivative (DPAN) as a pendant chain (WP-5). This configuration reached maximum values of 0.83 lmW^{-1} and 1.59 cdA^{-1} with CIE (0.31, 0.34). Using a more efficient red chromophore (MB-BT-ThTPA), they compared the influence of the position of the red emitter in the copolymer, i.e., either in the main chain or as a side chain attached by an alkyl

bridge (Liu et al. 2007b). By repositioning the MB-BT-ThTPA from the main to the side chain, the device efficiency is more than doubled [from 1.99 lmW^{-1} and 3.80 cdA^{-1} to 4.17 lmW^{-1} and 7.30 cdA^{-1}]. When the orange chromophore is attached as a side chain rather than incorporated into the polymer backbone (Liu et al. 2007a), they reported on an improvement in device efficiency by factor of 1.5–1.8 for a two-color single-component copolymer (WP-6). Recently, Zhang et al. (2010) reported on a highly efficient single-component polymer system containing three chromophores that are covalently attached to the polymer backbone (WP-7). With a correlated color temperature of approximately 4,500 K with CIE coordinates (0.37, 0.42), the best device reaches 6.2 % EQE and a luminous efficacy of 10.4 lmW^{-1} measured at 500 cdm^{-2} .

All the concepts from above were solely based on fluorescence-emitting materials. However, similar to the general consideration that phosphorescence should enhance the device efficiency, the incorporation of phosphors into a single-component copolymer seems promising. Jiang et al. (2006) discussed an approach for a hybrid fluorescent/phosphorescent copolymer. Based on a polyfluorene backbone, they added a benzothiadiazole chromophore for green emission to the polymer backbone and attached a phosphorescent emitter (2-phenylquinoline iridium complex) via an alkyl bridge. Despite the fact that the emission is close to warm white color point A, with CIE coordinates of (0.44,0.38), the device efficiencies of 5.6 cdA^{-1} is too low.

Future Outlook

The enormous interest in the development of new materials with improved efficiency and lifetime as well as optimized light quality for low-cost, low-energy consumption solid-state lighting has directed considerable scientific efforts and economical investment. It appears clear that research efforts focusing on new emitting molecular and polymeric compounds are at the heart of the progress of white OLED technology, and advancements will largely rely on the ability of chemists to design and synthesize new efficient organic materials and properly combine them. It may reasonably be expected that TADF materials exhibit great potential for the high efficiency and long lifetime application in the future white OLED technology. However, the report on white OLEDs using TADF materials is still on the way. No matter how to develop, we trust that white OLED-based lighting sources will enter the market as innovative systems which will produce deep changes in the illumination architectures combined with low-energy consumption and long lifetime.

References

- Adachi C, Baldo MA, Forrest SR, Lamansky S, Thompson ME, Kwong RC (2001) High-efficiency red electrophosphorescence devices. *Appl Phys Lett* 78:1622–1624

- Baldo MA, O'Brien DF, You Y, Shoustikov A, Sibley S, Thompson ME, Forrest SR (1998) Highly efficient phosphorescent emission from organic electroluminescent devices. *Nature* 395:151–154
- Baldo MA, O'Brien DF, Thompson ME, Forrest SR (1999a) Excitonic singlet-triplet ratio in a semiconducting organic thin film. *Phys Rev B* 60:14422–14428
- Baldo MA, Lamansky S, Burrows PE, Thompson ME, Forrest SR (1999b) Very high-efficiency green organic light-emitting devices based on electrophosphorescence. *Appl Phys Lett* 75:4–6
- Barbieri A, Accorsi G, Armaroli N (2008) Luminescent complexes beyond the platinum group: the d(10) avenue. *Chem Commun* 19:2185–2193
- Berggren M, Inganäs O, Gustafsson G, Rasmussen J, Andersson MR, Hjertberg T, Wennerström O (1994) Light-emitting-diodes with variable colors from polymer blends. *Nature* 372:444–446
- Brooks J, Babayan Y, Lamansky S, Djurovich PI, Tsyba I, Bau R, Thompson ME (2002) Synthesis and characterization of phosphorescent cyclometalated platinum complexes. *Inorg Chem* 41:3055–3066
- Burroughes JH, Jones CA, Friend RH (1988) New semiconductor-device physics in polymer diodes and transistors. *Nature* 335:137–141
- Burroughes JH, Bradley DDC, Brown AR, Marks RN, Mackay K, Friend RH, Burns PL, Holmes AB (1990) Light-emitting-diodes based on conjugated polymers. *Nature* 347:539–541
- Capecchi S, Renault O, Moon DG, Halim M, Etchells M, Dobson PJ, Salata OV, Chrisou V (2000) High-efficiency organic electroluminescent devices using an organoterbium emitter. *Adv Mater* 12:1591–1594
- Chang CF, Cheng YM, Chi Y, Chiu YC, Lin CC, Lee GH, Chou PT, Chen CC, Chang CH, Wu CC (2008) Highly efficient blue-emitting iridium(III) carbene complexes and phosphorescent OLEDs. *Angew Chem Int Ed* 47:4542–4545
- Chen CH, Shi J, Klubek KP (1999) US Patent 5, 908,581
- Chen CH, Tang CW, Shi J, Klubek KP (2000a) US Patent 6, 020,078
- Chen CH, Tang CW, Shi J, Klubek KP (2000b) Recent developments in the synthesis of red dopants for Alq(3) hosted electroluminescence. *Thin Solid Films* 363:327–331
- Chen BJ, Lin XQ, Cheng LF, Lee CS, Gambling WA, Lee ST (2001) Improvement of efficiency and colour purity of red-dopant organic light-emitting diodes by energy levels matching with the host materials. *J Phys D Appl Phys* 34:30–35
- Chiang CK, Fincher CR, Park YW, Heeger AJ, Shirakawa H, Louis EJ, Gau SC, Macdiarmid AG (1977) Electrical-conductivity in doped polyacetylene. *Phys Rev Lett* 39:1098–1101
- Chiu YC, Hung JY, Chi Y, Chen CC, Chang CH, Wu CC, Cheng YM, Yu YC, Lee GH, Chou PT (2009) En route to high external quantum efficiency (similar to 12 %), organic true-blue-light-emitting diodes employing novel design of iridium (III) phosphors. *Adv Mater* 21:2221–2225
- Chuang CY, Shih PI, Chien CH, Wu FI, Shu CF (2007) Bright-white light-emitting devices based on a single polymer exhibiting simultaneous blue, green, and red emissions. *Macromolecules* 40:247–252
- Czerwieńiec R, Yu JB, Yersin H (2011) Blue-light emission of Cu(I) complexes and singlet harvesting. *Inorg Chem* 50:8293–8301
- D'Andrade BW, Forrest SR (2003) Effects of exciton and charge confinement on the performance of white organic p-i-n electrophosphorescent emissive excimer devices. *J Appl Phys* 94:3101–3109
- D'Andrade BW, Brooks J, Adamovich V, Thompson ME, Forrest SR (2002) White light emission using triplet excimers in electrophosphorescent organic light-emitting devices. *Adv Mater* 14:1032–1036
- Deaton JC, Switalski SC, Kondakov DY, Young RH, Pawlik TD, Giesen DJ, Harkins SB, Miller AJM, Mickenberg SF, Peters JC (2010) E-type delayed fluorescence of a phosphine-supported Cu-2(μ -NAr₂)₂ diamond core: harvesting singlet and triplet excitons in OLEDs. *J Am Chem Soc* 132:9499–9508
- Di Censo D, Fantacci S, De Angelis F, Klein C, Evans N, Kalyanasundaram K, Bolink HJ, Grätzel M, Nazeeruddin MK (2008) Synthesis, characterization, and DFT/TD-DFT calculations

- of highly phosphorescent blue light-emitting anionic iridium complexes. *Inorg Chem* 47:980–989
- Endo A, Ogasawara M, Takahashi A, Yokoyama D, Kato Y, Adachi C (2009) Thermally activated delayed fluorescence from Sn^{4+} -porphyrin complexes and their application to organic light-emitting diodes – a novel mechanism for electroluminescence. *Adv Mater* 21:4802–4806
- Forrest SR (2004) The path to ubiquitous and low-cost organic electronic appliances on plastic. *Nature* 428:911–918
- Fox JL, Chen CH (1988) US Patent 4, 736,032
- Gao ZQ, Mi BX, Tam HL, Cheah KW, Chen CH, Wong MS, Lee ST, Lee CS (2008) High efficiency and small roll-off electrophosphorescence from a new iridium complex with well-matched energy levels. *Adv Mater* 20:774–778
- Gather MC, Alle R, Becker H, Meerholz K (2007) On the origin of the color shift in white-emitting OLEDs. *Adv Mater* 19:4460–4465
- Gather MC, Kohnen A, Meerholz K (2011) White organic light-emitting diodes. *Adv Mater* 23:233–248
- Han LL, Yang DF, Li WL, Chu B, Chen Y, Su ZS, Zhang DY, Yan F, Hu ZZ, Zhang ZQ (2008) The reduced triplet-triplet annihilation of electrophosphorescent device doped by an iridium complex with active hydrogen. *Appl Phys Lett* 93:153303
- Holmes RJ, Forrest SR, Tung YJ, Kwong RC, Brown JJ, Garon S, Thompson ME (2003) Blue organic electrophosphorescence using exothermic host-guest energy transfer. *Appl Phys Lett* 82:2422–2424
- Holmes RJ, Forrest SR, Sajoto T, Tamayo A, Djurovich PI, Thompson ME, Brooks J, Tung YJ, D'Andrade BW, Weaver MS, Kwong RC, Brown JJ (2005) Saturated deep blue organic electrophosphorescence using a fluorine-free emitter. *Appl Phys Lett* 87:243507
- Horowitz G, Fichou D, Peng XZ, Xu ZG, Garnier F (1989) A field-effect transistor based on conjugated alpha-sexithienyl. *Solid State Commun* 72:381–384
- Hosokawa C, Higashi H, Nakamura H, Kusumoto T (1995) Highly efficient blue electroluminescence from a distyrylarylene emitting layer with a new dopant. *Appl Phys Lett* 67:3853–3855
- Huang L, Wang KZ, Huang CH, Li FY, Huang YY (2001) Bright red electroluminescent devices using novel second-ligand-contained europium complexes as emitting layers. *J Mater Chem* 11:790–793
- Hudson ZM, Sun C, Helander MG, Chang YL, Lu ZH, Wang SN (2012) Highly efficient blue phosphorescence from triarylboron-functionalized platinum(II) complexes of *N*-heterocyclic carbenes. *J Am Chem Soc* 134:13930–13933
- Hung LS, Chen CH (2002) Recent progress of molecular organic electroluminescent materials and devices. *Mater Sci Eng R* 39:143–222
- Jiang JX, Xu YH, Yang W, Guan R, Liu ZQ, Zhen HY, Cao Y (2006) High-efficiency white-light-emitting devices from a single polymer by mixing singlet and triplet emission. *Adv Mater* 18:1769–1773
- Jung SG, Kang YJ, Kim HS, Kim YH, Lee CL, Kim JJ, Lee SK, Kwon SK (2004) Effect of substitution of methyl groups on the luminescence performance of Ir-III complexes: preparation, structures, electrochemistry, photophysical properties and their applications in organic light-emitting diodes (OLEDs). *Eur J Inorg Chem* 2004:3415–3423
- Jung SO, Zhao Q, Park JW, Kim SO, Kim YH, Oh HY, Kim J, Kwon SK, Kang Y (2009) A green emitting iridium(III) complex with narrow emission band and its application to phosphorescence organic light-emitting diodes (OLEDs). *Org Electron* 10:1066–1073
- Kang DM, Kang JW, Park JW, Jung SO, Lee SH, Park HD, Kim YH, Shin SC, Kim JJ, Kwon SK (2008) Iridium complexes with cyclometalated 2-cycloalkenylpyridine ligands as highly efficient emitters for organic light-emitting diodes. *Adv Mater* 20:2003–2007
- Kawamura Y, Goushi K, Brooks J, Brown JJ, Sasabe H, Adachi C (2005) 100 % phosphorescence quantum efficiency of Ir(III) complexes in organic semiconductor films. *Appl Phys Lett* 86:071104

- Kawamura Y, Brooks J, Brown JJ, Sasabe H, Adachi C (2006) Intermolecular interaction and a concentration-quenching mechanism of phosphorescent Ir(III) complexes in a solid film. *Phys Rev Lett* 96:017404
- Kido J, Hongawa K, Okuyama K, Nagai K (1994) White light-emitting organic electroluminescent devices using the poly(*N*-vinylcarbazole) emitter layer doped with 3 fluorescent dyes. *Appl Phys Lett* 64:815–817
- Koezuka H, Tsumura A, Ando T (1987) Field-effect transistor with polythiophene thin-film. *Synth Met* 18:699–704
- Kwong RC, Sibley S, Dubovoy T, Baldo M, Forrest SR, Thompson ME (1999) Efficient, saturated red organic light emitting devices based on phosphorescent platinum(II) porphyrins. *Chem Mater* 11:3709–3713
- Lamansky S, Djurovich P, Murphy D, Abdel-Razzaq F, Lee HE, Adachi C, Burrows PE, Forrest SR, Thompson ME (2001) Highly phosphorescent bis-cyclometalated iridium complexes: synthesis, photophysical characterization, and use in organic light emitting diodes. *J Am Chem Soc* 123:4304–4312
- Lee SK, Hwang DH, Jung BJ, Cho NS, Lee J, Lee JD, Shim HK (2005) The fabrication and characterization of single-component polymeric white-light-emitting diodes. *Adv Funct Mater* 15:1647–1655
- Lee SJ, Park KM, Yang K, Kang Y (2009) Blue phosphorescent Ir(III) complex with high color purity: fac-tris(2',6'-difluoro-2,3'-bipyridinato-N, C-4')iridium(III). *Inorg Chem* 48:1030–1037
- Lee SY, Yasuda T, Nomura H, Adachi C (2012) High-efficiency organic light-emitting diodes utilizing thermally activated delayed fluorescence from triazine-based donor-acceptor hybrid molecules. *Appl Phys Lett* 101:093306
- Lee J, Shizu K, Tanaka H, Nomura H, Yasuda T, Adachi C (2013) Oxadiazole- and triazole-based highly-efficient thermally activated delayed fluorescence emitters for organic light-emitting diodes. *J Mater Chem C* 1:4599–4604
- Li J, Nakagawa T, MacDonald J, Zhang QS, Nomura H, Miyazaki H, Adachi C (2013) Highly efficient organic light-emitting diode based on a hidden thermally activated delayed fluorescence channel in a heptazine derivative. *Adv Mater* 25:3319–3323
- Liu J, Zhou QG, Cheng YX, Geng YH, Wang LX, Ma DG, Jing XB, Wang FS (2005) The first single polymer with simultaneous blue, green, and red emission for white electroluminescence. *Adv Mater* 17:2974–2978
- Liu J, Guo X, Bu LJ, Xie ZY, Cheng YX, Geng YH, Wang LX, Jing XB, Wang FS (2007a) White electroluminescence from a single-polymer system with simultaneous two-color emission: polyfluorene as blue host and 2,1,3-benzothiadiazole derivatives as orange dopants on the side chain. *Adv Funct Mater* 17:1917–1925
- Liu J, Xie ZY, Cheng YX, Geng YH, Wang LX, Jing XB, Wang FS (2007b) Molecular design on highly efficient white electroluminescence from a single-polymer system with simultaneous blue, green, and red emission. *Adv Mater* 19:531–535
- Liu ZW, Bian ZQ, Ming L, Ding F, Shen HY, Nie DB, Huang CH (2008) Green and blue-green phosphorescent heteroleptic iridium complexes containing carbazole-functionalized beta-diketonate for non-doped organic light-emitting diodes. *Org Electron* 9:171–182
- Liu ZW, Qayyum MF, Wu C, Whited MT, Djurovich PI, Hodgson KO, Hedman B, Solomon EI, Thompson ME (2011) A codeposition route to CuI-pyridine coordination complexes for organic light-emitting diodes. *J Am Chem Soc* 133:3700–3703
- Lo SC, Shipley CP, Bera RN, Harding RE, Cowley AR, Burn PL, Samuel IDW (2006) Blue phosphorescence from iridium(III) complexes at room temperature. *Chem Mater* 18:5119–5129
- Lo SC, Harding RE, Shipley CP, Stevenson SG, Burn PL, Samuel IDW (2009) High-triplet-energy dendrons: enhancing the luminescence of deep blue phosphorescent iridium(III) complexes. *J Am Chem Soc* 131:16681–16688
- Luo J, Li XZ, Hou Q, Peng JB, Yang W, Cao Y (2007) High-efficiency white-light emission from a single copolymer: fluorescent blue, green, and red chromophores on a conjugated polymer backbone. *Adv Mater* 19:1113–1117

- Ma YG, Zhang HY, Shen JC, Che CM (1998) Electroluminescence from triplet metal-ligand charge-transfer excited state of transition metal complexes. *Synth Met* 94:245–248
- Manbeck GF, Brennessel WW, Eisenberg R (2011) Photoluminescent copper(I) complexes with amido-triazolato ligands. *Inorg Chem* 50:3431–3441
- Mehes G, Nomura H, Zhang QS, Nakagawa T, Adachi C (2012) Enhanced electroluminescence efficiency in a spiro-acridine derivative through thermally activated delayed fluorescence. *Angew Chem Int Ed* 51:11311–11315
- Misra A, Kumar P, Kamalasanan MN, Chandra S (2006) White organic LEDs and their recent advancements. *Semicond Sci Technol* 21:R35–R47
- Nakagawa T, Ku SY, Wong KT, Adachi C (2012) Electroluminescence based on thermally activated delayed fluorescence generated by a spirobifluorene donor-acceptor structure. *Chem Commun* 48:9580–9582
- Pope M, Swenberg CE (1982) *Electronic processes in organic crystals*. Clarendon, Oxford
- Pope M, Swenberg CE (1999) *Electronic processes in organic-crystals*. Oxford University Press, New York
- Ragni R, Plummer EA, Brunner K, Hofstraat JW, Babudri F, Farinola GM, Naso F, De Cola L (2006) Blue emitting iridium complexes: synthesis, photophysics and phosphorescent devices. *J Mater Chem* 16:1161–1170
- Reineke S, Baldo MA (2012) Recent progress in the understanding of exciton dynamics within phosphorescent OLEDs. *Physica Status Solidi A* 209:2341–2353
- Reineke S, Thomschke M, Lusse B, Leo K (2013) White organic light-emitting diodes: status and perspective. *Rev Mod Phys* 85:1245–1293
- Saitoh A, Yamada N, Yashima M, Okinaka K, Senoo A, Ueno K, Tanaka D, Yashiro R (2004) Novel fluorene-based blue emitters for high performance OLEDs. *Proc Soc Inf Disp* 150–153
- Sajoto T, Djurovich PI, Tamayo A, Yousufuddin M, Bau R, Thompson ME, Holmes RJ, Forrest SR (2005) Blue and near-UV phosphorescence from iridium complexes with cyclometalated pyrazolyl or *N*-heterocyclic carbene ligands. *Inorg Chem* 44:7992–8003
- Sato K, Shizu K, Yoshimura K, Kawada A, Miyazaki H, Adachi C (2013) Organic luminescent molecule with energetically equivalent singlet and triplet excited states for organic light-emitting diodes. *Phys Rev Lett* 110:247401
- Serevicius T, Nakagawa T, Kuo MC, Cheng SH, Wong KT, Chang CH, Kwong RC, Xia S, Adachi C (2013) Enhanced electroluminescence based on thermally activated delayed fluorescence from a carbazole-triazine derivative. *Phys Chem Chem Phys* 15:15850–15855
- Shi JM, Tang CW (2002) Anthracene derivatives for stable blue-emitting organic electroluminescence devices. *Appl Phys Lett* 80:3201–3203
- Shinar J, Shinar R (2008) Organic light-emitting devices (OLEDs) and OLED-based chemical and biological sensors: an overview. *J Phys D Appl Phys* 41:133001
- Si ZJ, Li J, Li B, Liu SY, Li WL (2009) High light electroluminescence of novel Cu(I) complexes. *J Lumin* 129:181–186
- Smith LH, Wasey JAE, Barnes WL (2004) Light outcoupling efficiency of top-emitting organic light-emitting diodes. *Appl Phys Lett* 84:2986–2988
- Son HJ, Han WS, Chun JY, Kang BK, Kwon SN, Ko J, Han SJ, Lee C, Kim SJ, Kang SO (2008) Generation of blue light-emitting zinc complexes by band-gap control of the oxazolyl phenolate ligand system: syntheses, characterizations, and organic light emitting device applications of 4-coordinated bis(2-oxazolylphenolate) zinc(II) complexes. *Inorg Chem* 47:5666–5676
- Swanson SA, Wallraff GM, Chen JP, Zhang WJ, Bozano LD, Carter KR, Salem JR, Villa R, Scott JC (2003) Stable and efficient fluorescent red and green dyes for external and internal conversion of blue OLED emission. *Chem Mater* 15:2305–2312
- Tamayo AB, Alleyne BD, Djurovich PI, Lamansky S, Tsyba I, Ho NN, Bau R, Thompson ME (2003) Synthesis and characterization of facial and meridional tris-cyclometalated iridium(III) complexes. *J Am Chem Soc* 125:7377–7387
- Tanaka D, Sasabe H, Li YJ, Su SJ, Takeda T, Kido J (2007) Ultra high efficiency green organic light-emitting devices. *Jpn J Appl Phys* 2(46):L10–L12

- Tang CW (1986) 2-layer organic photovoltaic cell. *Appl Phys Lett* 48:183–185
- Tang CW, Vanslyke SA (1987) Organic electroluminescent diodes. *Appl Phys Lett* 51:913–915
- Thompson M (2007) The evolution of organometallic complexes in organic light-emitting devices. *MRS Bull* 32:694–701
- Tsuboyama A, Iwawaki H, Furugori M, Mukaide T, Kamatani J, Igawa S, Moriyama T, Miura S, Takiguchi T, Okada S, Hoshino M, Ueno K (2003) Homoleptic cyclometalated iridium complexes with highly efficient red phosphorescence and application to organic light-emitting diode. *J Am Chem Soc* 125:12971–12979
- Tu GL, Zhou QG, Cheng YX, Wang LX, Ma DG, Jing XB, Wang FS (2004) White electroluminescence from polyfluorene chemically doped with 1,8-naphthalimide moieties. *Appl Phys Lett* 85:2172–2174
- Uoyama H, Goushi K, Shizu K, Nomura H, Adachi C (2012) Highly efficient organic light-emitting diodes from delayed fluorescence. *Nature* 492:234–238
- Wang YZ, Sun RG, Meghdadi F, Leising G, Epstein AJ (1999) Multicolor multilayer light-emitting devices based on pyridine-containing conjugated polymers and *para*-sexiphenyl oligomer. *Appl Phys Lett* 74:3613–3615
- Wen SW, Lee MT, Chen CH (2005) Recent development of blue fluorescent OLED materials and devices. *J Disp Technol* 1:90–99
- Wu LL, Yang CH, Sun IW, Chu SY, Kao PC, Huang HH (2007) Photophysical and electrochemical properties of blue phosphorescent iridium(III) complexes. *Organometallics* 26:2017–2023
- Xiao LX, Chen ZJ, Qu B, Luo JX, Kong S, Gong QH, Kido JJ (2011) Recent progresses on materials for electrophosphorescent organic light-emitting devices. *Adv Mater* 23:926–952
- Xie HZ, Liu MW, Wang OY, Zhang XH, Lee CS, Hung LS, Lee ST, Teng PF, Kwong HL, Zheng H, Che CM (2001) Reduction of self-quenching effect in organic electrophosphorescence emitting devices via the use of sterically hindered spacers in phosphorescence molecules. *Adv Mater* 13:1245–1248
- Xie WF, Liu SY, Zhao Y (2003) A nondoped-type small molecule white organic light-emitting device. *J Phys D Appl Phys* 36:1246–1248
- Xu ML, Wang GY, Zhou R, An ZW, Zhou Q, Li W (2007) Tuning iridium(III) complexes containing 2-benzo[b]thiophen-2-yl-pyridine based ligands in the red region. *Inorg Chim Acta* 360:3149–3154
- Yang XH, Wang ZX, Madakuni S, Li J, Jabbour GE (2008) Efficient blue- and white-emitting electrophosphorescent devices based on platinum(II) [1,3-difluoro-4,6-di(2-pyridinyl)benzene] chloride. *Adv Mater* 20:2405–2409
- Yeh CC, Lee MT, Chen HH, Chen CH (2004) High-performance blue OLEDs based on sterically hindered pyrene ost material. *Proc Soc Inf Disp* 788–792
- Yersin H, Donges D, Humbs W, Strasser J, Sitters R, Glasbeek M (2002) Organometallic Pt(II) compounds. A complementary study of a triplet emitter based on optical high-resolution and optically detected magnetic resonance spectroscopy. *Inorg Chem* 41:4915–4922
- Zhang LM, Li B, Su ZM (2009) Realization of high-energy emission from [Cu(N-N)(P-P)](+) complexes for organic light-emitting diode applications. *J Phys Chem C* 113:13968–13973
- Zhang BH, Qin CJ, Ding JQ, Chen L, Xie ZY, Cheng YX, Wang LX (2010) High-performance all-polymer white-light-emitting diodes using polyfluorene containing phosphonate groups as an efficient electron-injection layer. *Adv Funct Mater* 20:2951–2957
- Zhang QS, Li J, Shizu K, Huang SP, Hirata S, Miyazaki H, Adachi C (2012) Design of efficient thermally activated delayed fluorescence materials for pure blue organic light emitting diodes. *J Am Chem Soc* 134:14706–14709
- Zheng Y, Eom SH, Chopra N, Lee JW, So F, Xue JG (2008) Efficient deep-blue phosphorescent organic light-emitting device with improved electron and exciton confinement. *Appl Phys Lett* 92:223301
- Zhou GJ, Wong WY, Yao B, Xie ZY, Wang LX (2007) Triphenylamine-dendronized pure red iridium phosphors with superior OLED efficiency/color purity trade-offs. *Angew Chem Int Ed* 46:1149–1151

- Zhou GJ, Wang Q, Ho CL, Wong WY, Ma DG, Wang LX, Lin ZY (2008a) Robust tris-cyclometalated iridium(III) phosphors with ligands for effective charge carrier injection/transport: synthesis, redox, photophysical, and electrophosphorescent behavior. *Chem Asian J* 3:1830–1841
- Zhou GJ, Ho CL, Wong WY, Wang Q, Ma DG, Wang LX, Lin ZY, Marder TB, Beeby A (2008b) Manipulating charge-transfer character with electron-withdrawing main-group moieties for the color tuning of iridium electrophosphors. *Adv Funct Mater* 18:499–511

White OLED Devices

Dongge Ma

Contents

Introduction	322
Basics of White OLED Devices	324
Operational Principle of White OLEDs	324
Advantages of White OLEDs	326
Efficiency Characterization of White OLEDs	327
Challenges of White OLEDs	329
Current Status of White OLED Devices	330
Fluorescence White OLEDs	331
Phosphorescence White OLEDs	337
Fluorescence/Phosphorescence Hybrid White OLEDs	346
Tandem White OLEDs	352
Future Outlook	357
References	357

Abstract

In this chapter, we will review the progress of white OLEDs based on organic small molecules in view of device architectures. The basics of white OLED devices are, firstly, demonstrated, and then the advanced architectures and current status of white OLEDs based on fluorescent, phosphorescent, and hybrid emitters are discussed. Because tandem structures, where similar or different emitting units are connected through a charge generation layer (CGL), provide further improvement in the efficiency and stability of white OLEDs, the advances of tandem white OLEDs are also discussed. Finally, the future outlook of white OLEDs is given.

D. Ma (✉)

Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, China
e-mail: mdg1014@ciac.jl.cn

Introduction

Organic light-emitting devices (OLEDs) are electroluminescence devices based on organic molecules with simple sandwiching structures (Tang and Vanslyke 1987; Sun et al. 2006; Schwartz et al. 2006). After the discovery of an efficient bilayer OLED technology in 1987, OLEDs have attracted much attention because of their potential applications in full-color flat-panel displays and high-efficiency lightings. Device architecture was studied intensively (Pfeiffer et al. 2002; Baldo et al. 1998a; Adachi et al. 2001). Its short lifetime that has been a roadblock to development is now tackled. Some OLEDs show a lifetime of more than 200 Khrs (D'Andrade et al. 2008). Material systems commercially available at present showed that OLED technology was considered as a possible next lighting technology (Wang et al. 2009a, b; Wang and Ma 2010; Reineke et al. 2013; Sasabe and Kido 2011). The power efficiency of OLEDs developed by Novaled has reached 124 lm/W with a 3D light extraction system (Reineke et al. 2009), and the power efficiency of commercially available large-area white OLEDs by LG Chem is over 80 lm/W by using internal and external out-coupling technology (Moon et al. 2013), which are high enough to replace a bulb and even a fluorescent lamp. These improvements in technology and product development have led to an explosive growth in the number of OLED lighting patents being filed as well as scientific publications.

As promising candidates for lighting applications, OLEDs have several advantages compared with common light sources, such as incandescent bulbs, fluorescent tubes, or inorganic LEDs: they are flat-area emitters offering a pleasant diffuse light perception, wide viewing angles, vivid colors, light and thin luminaire, and possibility to be made transparent or to be processed on flexible substrates by low-cost processing based on various thin-film deposition techniques such as vacuum thermal evaporation, spin coating, or casting from solution. Many examples of white OLED samples are seen, and white OLED tiles and lamps (OSRAM, Philips, LG, Panasonic, Lumiotec, and others) are commercially available now too.

As we know, white light has three characteristics: (i) the Commission Internationale de L'Eclairage (CIE) coordinates, (ii) the color rendering index (CRI), and (iii) the correlated color temperature (CCT). For lighting purposes, the light source should give CIE coordinates similar to that of a blackbody radiator, and, at the same time, the closer the light source to the ideal white point (0.33, 0.33) is, the better the color purity. However, there is a quite broad region of the diagram around this point that can be considered white light. CRI is a number ranging from 0 to 100 that measures the ability of a source lighting an object to reproduce the true color of the object. It is important to note that CRI is defined only in the proximity of the Planckian locus. A lower limit for a good light source is a CRI of 80 for indoor lighting applications. Besides CIE and CRI, the high quality of white light illumination sources requires a CCT between 2,500 and 6,500 K, depending on the different markets; for example, incandescent lamps have a CCT = 2,700 K (warm white) and fluorescent lamps can range from about 3,000 K to more than 4,000 K (cool white) (D'Andrade and Forrest 2004). Cool white light is less desirable for consumers, particularly for in-house lighting. To meet the requirement of lighting

applications in color characteristics, it is very necessary to achieve a satisfactory white light by combining three primary colors: red, green, and blue. However, in the case of OLEDs, besides the fact that white light emission can be achieved by mixing either of the three primary colors (red, green, and blue), the combination of two complementary colors is also a kind of simple and effective way to produce high-quality white light due to the wide emission property of organic molecules (Wang et al. 2009b; Wang and Ma 2010; Su et al. 2008a; Chang et al. 2010; Wang et al. 2010; Zhao et al. 2012a).

Because of the sandwiched structure of OLEDs, several possibilities to generate white light with OLEDs have been proposed. An alternative way is to use paralleled red, green, and blue pixels together, which are well used in OLED display. The advantage of this approach is that every color can be addressed individually, which can be used to compensate differential aging or to easily change the color of the light source. However, it is also technically the most complicated structuring process, making production expensive. As lighting sources, its higher current density for each color also accelerates the degradation. Therefore, it is not a suitable way in practical lighting application. The other, also a very simple, approach to achieve white light emission in OLEDs is based on color conversion, which has been well established in fluorescent tubes and white LEDs (Schwab et al. 2011). This approach uses only one electrically excited blue emitter to optically excite the color down-conversion layers (CCLs). It can be seen that its key is that high-efficiency blue emitter and conversion materials have to be used. At present, only a few publications dealing with the light conversion in OLEDs are reported due to the problem of low-efficiency blue emitter and the absence of efficient conversion materials (Schwab et al. 2011; Li et al. 2007; Cho et al. 2010; Chen and Kwok 2011). Most of the white OLEDs to realize good white light reported in the literatures are the structures of emission layers sandwiched between electrodes and charge carrier transport layers, including single emission layer doped with two- or three-color molecules (Wang et al. 2009b; D'Andrade et al. 2004a; Lee et al. 2009; Eom et al. 2009), separated emission layers containing different color sub-layers (Sun et al. 2006; Wang et al. 2009a; Reineke et al. 2009; Kido et al. 1995; D'Andrade et al. 2002, 2004b; Schwartz et al. 2007; Su et al. 2008b; Schwartz et al. 2009; Sasabe et al. 2010; Chang et al. 2013), and stacked emission units with different colors connected by charge generation layer (CGL) (Matsumoto et al. 2003; Liao et al. 2004; Guo and Ma 2005; Wang et al. 2009c; Chen et al. 2011, 2012a). This is not only the easiest way to fabricate a white OLED but also the most interesting one in terms of device physics. Among the challenges to achieve high-efficiency white light are the optimum choice of emitter materials, tuning of charge carrier balance, adjusting the recombination zone, and mutual energy transfer.

Besides the satisfaction of the color characteristics, power efficiency and lifetime are also two important performance parameters for OLEDs as lighting source in applications. To be competitive with classic lighting sources such as fluorescent tubes or LEDs, OLED lighting still needs to have high power efficiency and long stability in production. Aiming at improving the efficiency and stability, various key technologies on device architectures and organic material design have been created (Wang and Ma 2010; Reineke et al. 2013; Sasabe and Kido 2011). It can be seen that

not only the device structures but also the performance of used materials and the matching of its energy level are very important. It is well known that organic electroluminescent materials classify either as fluorescence or phosphorescence. When white OLEDs are based on fluorescent emitters, we call them fluorescence white OLEDs, whereas white OLEDs based on phosphorescent emitters are called phosphorescence white OLEDs. As known, fluorescence OLEDs have the advantage of a long lifetime, but its efficiency is low due to an upper limit of the internal quantum efficiency of 25 % originating from singlet exciton emission (Segal et al. 2003). Phosphorescence OLEDs can emit at very high efficiency because the introduction of phosphorescent emitter molecules theoretically allows 100 % excitons composed of 25 % singlets and 75 % triplets used in electroluminescence (EL) processes (Baldo et al. 1998b; Ma et al. 1998). However, the poor stability of blue phosphorescence emitter molecules greatly limits the application of phosphorescence OLEDs. Actually, the above problems can be well resolved by the combination of fluorescent blue emitters with phosphorescent red and green emitters in white OLEDs by distributing the singlet and triplet excitons in a suitable way (Sun et al. 2006; Schwartz et al. 2007, 2009). It is found that this approach achieved 100 % internal quantum efficiency in white OLEDs and also avoided the use of unstable phosphorescent blue emitters. Obviously, the difficulties in finding a stable blue phosphorescent emitter system make the combination of a stable blue fluorophor with red and green phosphors in so-called hybrid white OLEDs an attractive alternative to fully phosphorescent or fully fluorescent devices.

In fact, regardless of whatever device architectures and endeavors we create, the core is always involved with fine-tuning the charges and excitons (singlet and triplet) that are associated with the internal physical processes within the device, causing them to ultimately depreciate by giving off efficient white light. That is, the careful management of the charges and excitons in the emission region by advanced device architectures is one key factor in realizing high-performance white OLEDs. In this chapter, we will review the progress of white OLEDs based on organic small molecules in view of device architectures. The basics of white OLED devices are, firstly, demonstrated, and then the advanced architectures and current status of white OLEDs based on fluorescent, phosphorescent, and hybrid emitters are discussed. Because tandem structures, where similar or different emitting units are connected through a charge generation layer (CGL), provide further improvement in the efficiency and stability of white OLEDs, the advances of tandem white OLEDs are also discussed. Finally, the future outlook of white OLEDs is given.

Basics of White OLED Devices

Operational Principle of White OLEDs

Organic light-emitting diodes (OLEDs) are ultrathin, flat-area emitting devices made of thin-film organic semiconductors sandwiched between two electrodes. State-of-the-art small molecule-based OLEDs consist of various layers – each layer having a

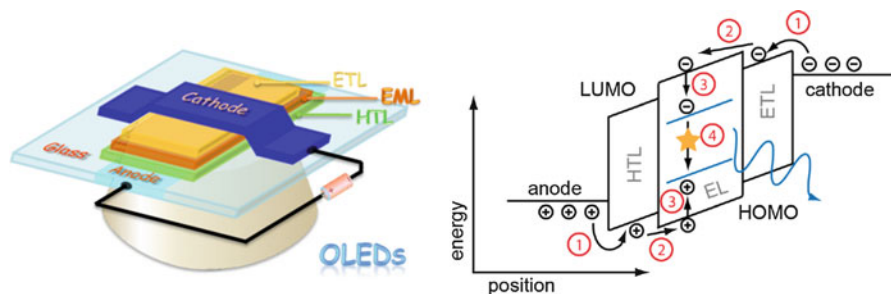


Fig. 1 Architecture (*left*) and principles of light emission (*right*) in a multilayer OLED

distinct functionality. These films are prepared by thermal evaporation in high vacuum or organic vapor-phase deposition. Depending on the required properties, different OLED architectures exist. The general architecture of an OLED is shown in Fig. 1 (left). The conventional bottom-emitting OLEDs comprise a transparent electrode on top of a glass substrate, followed by one or more layers of organic material and capped with a highly reflective metal electrode. Depending on the choice of organic materials for the emissive layer, OLEDs can be designed to emit any color, including white light with various color temperatures. This, in combination with high luminous efficacy and long lifetimes, reputedly makes OLEDs ideal candidates for lighting applications.

The right of Fig. 1 shows the principles of operation of multilayer OLEDs. When a voltage is applied to the transparent electrodes, the current flow through the organic layers generates light as the electrons and holes recombine in the emissive layer. Actually, the EL includes four basic processes: (1) the injection of charge carriers (holes and electrons) from the respective electrodes into the organic layers, (2) the charge carrier transport in the organic layers, (3) the charge carrier recombination on emission molecules with the creation of an excited state (excitons) and possible diffusion of these excitons, and (4) the decay of excitons accompanied by the emission of photons. For organic molecule-based OLEDs, the excitons are generally formed by means of *Langevin*-type recombination (Albrecht and Bassler 1995).

Because the formed excitons on organic molecules have a relatively strong binding energy (up to 1.5 eV) and are localized on one molecule, the excited molecule state is generally called a Frenkel exciton. The formation, transfer, and decay of the excitons are of essential importance for the luminescent efficiency of OLEDs. In OLEDs, the charges are injected statistically with respect to their electron spin, finally determining the formation of singlet and triplet excited states. Due to spin-statistics, the exciton-formed electron-hole recombination leads to 25 % singlet and 75 % triplet state population. In fluorescent organic molecules, as shown in Fig. 2, only the singlets emit light (fluorescence) while the triplet excitation energy is transferred into heat (left-hand side). On the other hand, phosphorescent organic molecules, such as organometallic compounds with transition metal centers, do not exhibit fluorescence but show a fast intersystem crossing (ISC) to the lowest triplet state. Thus, the emitter harvests singlet and triplet excitation energy and can

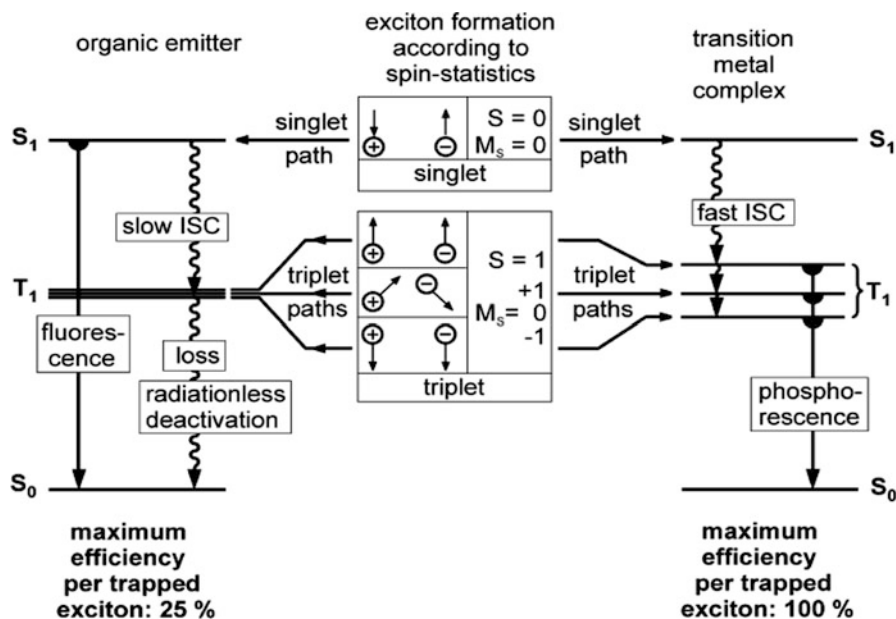


Fig. 2 Exciton-formation and light-emission processes in an organic molecule-based OLED (Yersin 2004)

efficiently emit light (phosphorescence). Obviously, a triplet emitter can exhibit a four times higher EL efficiency than a singlet emitter. This means that the best way to obtain high-efficiency OLEDs is by using phosphorescent molecules as the emitter. Recently, a possible way to enhance fluorescence efficiency to 100 % like phosphorescence was discovered by the up-conversion of a triplet excited state (T_1) into a singlet excited state (S_1). The possible up-conversion mechanism includes triplet–triplet annihilation (TTA, i.e., *p-type* delayed fluorescence) (Birks 1970) and thermally activated delayed fluorescence (TADF, i.e., *e-type* delayed fluorescence) (Valeur 2002). For this, it is necessary for fluorescent emitters to have a small energy difference between triplet and singlet excited states so that the triplet energies can be effectively transferred to singlet states. The recent achievement of high-efficiency red, green, and blue OLEDs based on fluorescent emitters with the up-conversion property has shown the possible promise of using them to fabricate high-efficiency and long-lifetime white OLEDs (Fukagawa et al. 2012; Lee et al. 2012; Zhang et al. 2012; Uoyama et al. 2012).

Advantages of White OLEDs

In fact, white OLEDs were targeted toward display applications for use primarily as liquid-crystal display backlights and OLED itself. As their power efficiencies have surpassed those of incandescent lamps and even arrived at the level of fluorescent

tubes by the improvements in device architectures and synthesis of novel materials, the interest in the application of white OLED technology for solid-state lighting has been steadily increasing. In comparison to common light sources, such as incandescent bulbs, fluorescent tubes, or inorganic LEDs, white OLEDs have several advantages. White OLEDs offer distinct properties to replace classical light sources or even inorganic LEDs. In particular, white OLEDs can be used as an ultrathin area light source, emitting diffuse light for a potentially large area. This means that they do not need light distribution elements themselves, almost like a lamp with lampshade, and its thickness could be made very thin that could allow the lighting to be placed directly on ceilings as a planar sheet of light. As we know, organic materials emit generally very wide spectrum; this makes the CRI of white OLEDs very high, which is especially suitable for indoor general lighting and even for professional photography application. Moreover, through the adjustment of each color of light-emitting material proportion, one can produce any color of light, adapting to different applications. And white OLEDs are also fully dimmable and can be switched on and off without any time delay. It is worth noting that being an area light source, the heat generation inside OLEDs is small during standard operation, not like inorganic LEDs, which is focused on a small volume. Therefore, OLEDs do not need a heat management to specially do in power source and out-shelf, thus saving power consumption, space, and cost. In addition, it is also possible for white OLEDs to have transparent panels, which could be used as lighting elements in windows, screens, or room dividers. Most interestingly, white OLEDs can be fabricated onto flexible substrates to provide new architectural design opportunities that cannot be realized with any other technologies. Therefore, white OLEDs are the most promising candidates for lighting applications in the near future.

Efficiency Characterization of White OLEDs

Power efficiency (PE, η_p), current efficiency (CE, η_c), and external quantum efficiency (EQE, η_{ext}) are the common parameters used to describe white OLED performance. CE in candelas of light per ampere current (cd/A) is the ratio of the luminous intensity in forward direction and the current through the device. This measure weighs photons of light based on the photonic response curve of the human eye, so photons of light near 550 nm give the highest cd/A efficiency. PE in lumen per watt (lm/W) represents the output light power from a device (measured in lumens, i.e., the unit of light intensity perceived by the human eye) per electrical power input (measured in watts). Therefore, PE can be taken as the standard to characterize whether white OLEDs save electricity or not. Accordingly, high PE values are always the key target in the pursuit of energy-saving white OLEDs. Generally, EQE is the total number of photons emitted by the device per electron-hole pair injected into the device and can be described as

$$\eta_{\text{ext}} = \eta_{\text{int}}\eta_{\text{ph}} = \gamma_{\text{e-h}}\eta_{\text{s-p}}\phi_{\text{p}}\eta_{\text{ph}} \quad (1)$$

where η_{int} is the internal quantum efficiency defined as the total number of photons generated inside the device per electron–hole pair injected into the device; η_{ph} is the out-coupling efficiency; $\gamma_{\text{e-h}}$ is the ratio of electrons to holes (or vice versa, to maintain $\gamma_{\text{e-h}} \leq 1$) injected from opposite contacts; $\eta_{\text{s-p}}$ is the fraction of the overall excitons formed which result in radiative transitions ($\eta_{\text{s-p}} \sim 0.25$ for fluorescent species and 1 for phosphorescent dyes); and ϕ_{p} is the intrinsic quantum efficiency for radiative decay (both phosphorescence and fluorescence are included).

When the emitting materials combination is defined, $\eta_{\text{s-p}}$ and ϕ_{p} will be decided. Furthermore, η_{ph} is usually not subjected to the electronic process occurring within the device. Hence, from the device engineering point of view, EQE value is mostly sensitive to $\gamma_{\text{e-h}}$. In amorphous organic layers, the recombination of free electrons and holes is generally by means of *Langevin* type, where the bimolecular recombination efficiency should be associated with $\gamma_{\text{e-h}}$, which follows (Reineke et al. 2013)

$$\gamma_{\text{e-h}} = j_{\text{low}}/j_{\text{high}} \leq 1 \quad (2)$$

where j_{low} and j_{high} are the lower and higher of the two charge carriers, respectively. Apparently, this charge balance is a key factor in deciding device efficiency. An ideal condition is that electrons and holes are injected and transported equally well and both charges confined inside the emission layer, which is sandwiched between the effective charge-blocking layers; $\gamma_{\text{e-h}}$ can then be close to unity. However, in general white OLEDs, the large difference in electron and hole charge mobilities within emitting layer (s) (EML) or between different functional layers and the formed energy barriers associated with the mismatch of energy levels between layers will make it hard to realize. Therefore, balancing the charge injection and transport is a prerequisite to the realization of high efficiency. It has been shown that introducing bipolar hosts, electron-transporting layer (ETL) with high mobility, and effective charge-blocking layers is beneficial to the improvement of this issue (Wang and Ma 2010; Su et al. 2008a).

For a common white OLED, if the emission pattern is assumed to be the *Lambertian* type, then the correlation between PE and EQE can be described as:

$$\eta_{\text{p}} \propto \eta_{\text{ext}}/U \quad (3)$$

where U is the operational voltage of the device. Clearly, in addition to all the factors that influence device EQE, PE is also decided by the driven voltage. However, compared to inorganic semiconductors, the intrinsic mobility of organic materials is rather low. Much higher voltage is required to achieve the correspondingly high brightness, thus significantly decreasing PE. Furthermore, the unfavorable barriers associated with organic–organic or metal–organic contacts also contribute to this effect. Hence the operational voltage must be lowered in order to achieve higher PE. One effective method is by adding charge transfer dopants into organic layers or introducing materials with high mobility to increase the density of charge carriers and hence the conductivity of organic layers, ultimately decreasing the driven voltage (Schwartz et al. 2007). Other methods such as employing thin emissive layers (Su et al. 2008a; D’Andrade et al. 2004a),

modifying electrodes by the addition of buffer layers (You et al. 2007), and designing a stepped progression of the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) energy levels between layers that can facilitate charge injection and transportation (Sun and Forrest 2007) are all proposed to increase PE.

As white OLEDs need to work at higher brightness to satisfy the requirement of practical lighting applications, as a result, the efficiency is greatly reduced. Specially, it is more severe for phosphorescence-based OLEDs due to the existence of exciton density, which will increase the probability of excited annihilation processes. Experimental and theoretical evidences have demonstrated several exciton-quenching processes responsible for the observed efficiency roll-off in phosphorescent devices: triplet–triplet and triplet–charge carrier (polaron) annihilation and field-assisted dissociation of Coulombically correlated electron–hole pairs preceding the exciton formation. As known, the triplet concentration is proportional to its intrinsically long lifetime and increases linearly with current density j . The mutual triplet–triplet or triplet–polaron annihilation, a process that becomes very efficient at high triplet concentration, is a general explanation for the efficiency roll-off under high current density, which has been well discussed by Baldo et al. (2000) and Reineke et al. (2007). Furthermore, Kalinowski et al. (Kalinowski et al. 2002) have claimed that the field-assisted dissociation of charge pairs is another principal cause of this efficiency drop at high electric fields. On the basis of their experimental and analytical results, they successfully solved the problem that the current density threshold for the efficiency reduction by triplet–triplet interaction quenching was shown to be in the order of A/cm^2 , while, experimentally, the efficiency tended to decrease rapidly from its peak in the mA/cm^2 region in some phosphorescent devices. Moreover, they also demonstrated the coexistence of annihilation and dissociation of excitons on charge carriers to be common phenomena in phosphorescent devices (Kalinowski et al. 2006). To select molecular systems with strongly bound excitons, low exciton diffusivity and a wide recombination zone in addition to the short lifetime of triplet excitons are thus expected to minimize the exciton-quenching effect, which has been strongly supported by some experimental evidences (Su et al. 2008a; Giebink and Forrest 2008).

Challenges of White OLEDs

Because of the existence of competition for white OLEDs in lighting fields with fluorescent tubes and inorganic LEDs, the application of white OLEDs faces enormous challenge in efficiency, lifetime, and cost. For white OLEDs, it is very necessary to have high power efficiency at high brightness (>80 lm/W at $3,000$ cd/m^2) and large emission area without sacrificing operational lifetime ($>20,000$ h). This is yet very difficult for white OLEDs at present. For example, as described, the efficiency of white OLEDs is sharply reduced at such high brightness due to the degradation of used organic materials. And working at a large area, a short circuit, nonuniform light emission, hot spot, efficiency reduction (power loss), and heat

generation also significantly exist in devices, which are also directly related to the lifetime. Therefore, the further development of highly efficient and stable white OLED materials and efficient device structures based on these materials is greatly important. Moreover, from the point of view of scientific research, because an accurate theory of the operation in white OLEDs is still in its infancy, therefore, a quantitative and universal model is ardently anticipated, and once it is established, methods for molecular scale and device structure manipulation for the prediction of device performance may become possible. Clearly, it is only through the joint endeavors of both chemists and physicists that revolutionary advances in this technique can be attained. For the challenge of lifetime, another problem is that the color is not also expected to change as OLED device ages or operates at different voltages. In other words, the relative intensity of the respective color sources should remain constant over its lifetime and voltage. This not only requires the red, green, and blue emitters to keep good stability but also needs the exciton recombination region not to change with the operational voltage. Scientists are always resolving these problems, especially the problems on efficiency reduction and lifetime degradation in the case of large-area emission, which become a large hindrance for OLED commercialization.

It is still a challenge to realize large flexible white OLED lighting, which is the unique feature of OLED lighting in future applications. However, as we know, OLEDs are very sensitive to moisture and oxygen in the air, so it is very necessary to encapsulate OLEDs to keep its stability. Generally, plastic substrates do not prevent from moisture and oxygen. Therefore, besides developing new flexible substrates, thin-film encapsulation technology has become an important research topic in this field.

Actually, if white OLED technology obtains wide application for general illumination purposes, reducing the cost per lumen will become a major challenge in the future. As lighting is not a new market, white OLEDs must be cheaper than the current technology in order to be adopted. The decorative lighting market has the least stringent requirements and requires a brightness of 100–500 cd/m^2 at an efficiency of >15 lm/W for an operational lifetime of at least 10,000 h. The market for general illumination requires much larger panels at high brightness (3,000–5,000 cd/m^2) and very high efficiencies of >60 lm/W . Whatever develops, it is believed that white OLEDs must be widely used in lighting fields if its cost comes down even to US\$30.0/ m^2 . It can be seen that the reduction of cost needs not only to develop cheap EL material systems and effective processing technology of devices but also to greatly enhance the productivity of large-area OLED panels.

Current Status of White OLED Devices

Over the following years significant progress has been made both in the architecture and the performance of white OLEDs. For lighting, high-efficiency and stable white OLEDs have been reported based upon fluorescence, phosphorescence, and combinations of fluorescence and phosphorescence (hybrid) emitters. As demonstrated

above, fluorescence devices are typically more stable than phosphorescence devices. However, phosphorescence structures have been demonstrated to be more efficient. In this section, we will review the current status of white OLEDs based on different organic systems for lighting applications.

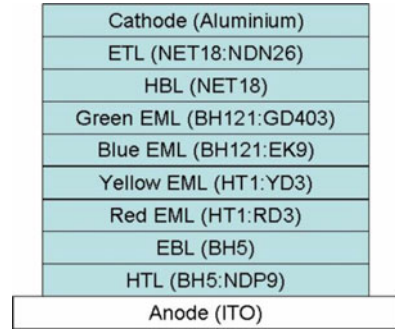
Fluorescence White OLEDs

Due to the advantages of long lifetime and low cost of fluorescence materials used in the fabrication of white OLEDs, people are always making an effort to further enhance the efficiency of fluorescence white OLEDs. For lighting, white OLEDs must have high power efficacy of >30 lm/W at 1,000 nits) besides a long lifetime (50,000 h @ 1,000 nits) and CRI > 80 (Tyan et al. 2008). Based on the limitation of intrinsic quantum efficiency of fluorescence materials, in order to arrive at the requirement of above performance, one of the key opportunities for improving fluorescence white OLED device efficiency is to improve the light extraction efficiency, including the external extraction structure (EES), where the extraction enhancement structure is coated on the free surface of the glass substrate, and the internal extraction structure (IES), where an extraction enhancement structure is inserted between the substrate and the anode. It is experimentally demonstrated that the efficiency can be improved by over three times when using external and internal extraction together (Reineke et al. 2009; Tyan et al. 2008). This means that the fabricated white OLEDs based on fluorescence have to emit an efficiency of over 10 lm/W.

High-efficiency fluorescence white OLEDs were first obtained by using two-emission-layer structure composed of blue and yellow. Here the white emission is obtained by adjusting the ratio of emissions from the adjacent blue- and yellow-emitting layers by varying layer thicknesses or by using multiple cohosts and/or co-dopants in both the yellow and blue emissive layers. The yellow dopants are often rubrene derivatives while efficient blue dopants are used in the blue EML. The best device showed a low voltage (<4 V), high efficiency (14 cd/A), and long lifetime ($\sim 50,000$ h) (Hatwar et al. 2004). However, the two-color white OLED is deficient in green emission, which cannot satisfy the need of display and even lighting; thus three- and four-color white OLEDs were configured well (Nishimura et al. 2008). In the four-color configuration, the efficiency was 9.3 cd/A, and the emissive spectrum spans the whole visible wavelength. It was found that the thickness of the yellow layer is crucial for achieving high efficiency and maintaining a high color gamut. Unfortunately, the efficiency of all devices is low.

Idemitsu Kosan Co., Ltd. improved the efficiency of fluorescence white OLEDs by developing both carrier transport materials and emitting materials (He et al. 2004). The fundamental device structure of the fabricated white OLED was ITO/HT/HT/RH:RD/CBL/BH:BD/BH:GD/ET/LiF and Al, R, G, and B emission layers and a carrier-blocking layer (CBL) were stacked. The carrier balance was optimized by using the new HIL and HTL materials. Good white emission with CIE (0.369, 0.425) was obtained. The white device showed a long lifetime of over

Fig. 3 Structure of the fluorescent white PIN OLED (Reproduced from Murano et al. (2009))



70,000 h at 1,000 cd/m² maintaining a high efficiency of 20 cd/A with the optimization of a carrier balance by using new carrier transport materials. The lifetime improvement has been attributed to this delocalization of recombination zone.

It is well known that the enhancement of the power efficiency of white OLEDs is reducing the operational voltage of devices. The PIN structure is one of the best methods. Since its development it could be demonstrated that PIN technology is both capable of achieving very high efficiencies (He et al. 2004) and very long lifetimes (Birstock et al. 2005), and it can be considered now as a very well-established technology for OLED display and lighting applications. Highly efficient and stable PIN white OLED structures based on Kodak's proprietary emitters and Novaled's proprietary *p*- and *n*-type dopants have been developed (Murano et al. 2009). The basic device structure of fluorescent white PIN OLEDs is shown in Fig. 3. In this device, both the red and yellow emission layers (EMLs) have predominantly hole-transporting properties, whereas the blue and green EMLs are predominantly electron transporting. The recombination of charges will therefore mainly take place in the yellow and blue EMLs. One can assume that the red and green emitters are mostly excited through an exciton diffusion process from the inner emission layers to the outer sides of the emission zone. This layout is advantageous to control the color balance over the application luminance range and also to limit color shift with device aging. It is clearly seen that the device allows for low operational voltage and a good charge carrier balance in the EML through the use of *p*- and *n*-type-doped transport layers. 15.1 cd/A current efficiency and 3.03 V driving voltage were achieved at a brightness of 1,000 cd/m². This device also showed a long lifetime of 27,000 h by extrapolating at a starting brightness of 1,000 cd/m². Furthermore, by using the HBL materials EK-ET44 from Kodak or NET8 from Novaled instead of NET18, the efficiencies of the devices could be improved to 15.8 cd/A and 16.8 cd/A, and the lifetime reaches 33,000 and 22,000 h, respectively, showing the important role of PIN structure.

As we know, most of the internally generated light is trapped within the device and only a small fraction (~20–30 %) exits due to the high refractive index (~1.8) of the emitting organic layers and the glass substrate (~1.5). The internally generated light is distributed between several different modes, and details of the emission loss and light-trapping mechanism have been described elsewhere (Tyan et al. 2008;

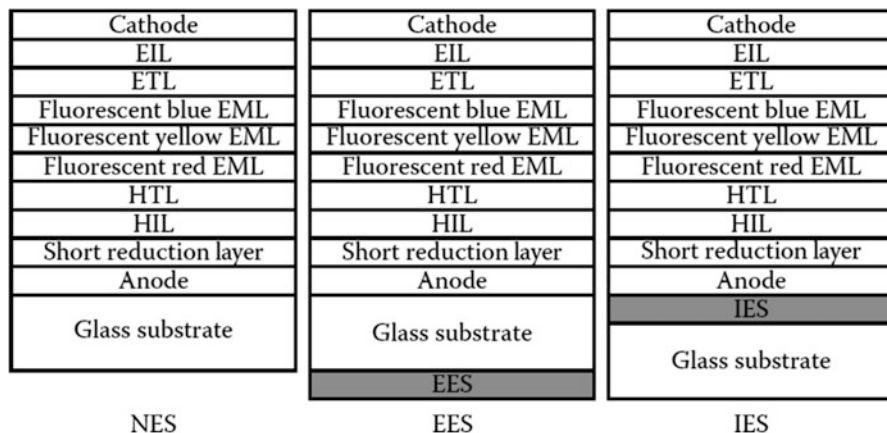


Fig. 4 Schematics of the device architectures, comparing devices with no extraction structure (NES), an EES, and an IES (Reproduced from Tyan et al. (2008))

Brutting et al. 2013). Therefore, the efficiency of the fabricated OLEDs can be greatly improved if the trapped light can be altered in direction to allow it to exit the structure. The out-coupling methods include 2 types of external extraction structure (EES) and internal extraction structure (IES). The EES is relatively easy to implement and has been shown to improve light output by as much as 60 %. It does not, however, access the organic-ITO light, and hence the potential of extraction efficiency enhancement is limited. The IES is expected to be more effective because it can access the organic-ITO mode in addition to the air mode and the substrate mode. However, respectively speaking, the IES is much more difficult to fabricate. Yuan-Sheng Tyan et al. in Eastman Kodak Company have successfully developed the IES and applied it in all-fluorescence white OLEDs (Tyan et al. 2008). Figure 4 shows the schematics of the device architectures, comparing devices with no extraction structure (NES), an EES, and an IES. In comparison with the devices without extraction structure and with EES, it is experimentally found that the efficiency of the fabricated white OLEDs is significantly enhanced due to the utilization of IES. A 14.5 % external quantum efficiency and 31.2 lm/W efficacy have been demonstrated. The color coordinate at (0.387, 0.381) falls well within the DOE Energy Star tolerance quadrangle for 4,000 K CCT. The T_{50} lifetime is projected to be over 10,000 h at an initial brightness of 1,000 cd/m^2 .

Actually, the fabrication of high-efficiency all-fluorescent white OLEDs is still difficult due to the waste of the triplet excitons formed from the majority of recombination events. To fully achieve optimum efficiency and color rendition from different fluorescent emission molecules in white OLEDs, besides developing more efficient electro-fluorescent host and guest materials, some effective new architectural approaches seem to be rather important. Recently, D.G. Ma's group designed a high-efficiency fluorescence white OLEDs (Zhang et al. 2009). The white OLEDs included three sequent red, green, and blue separately

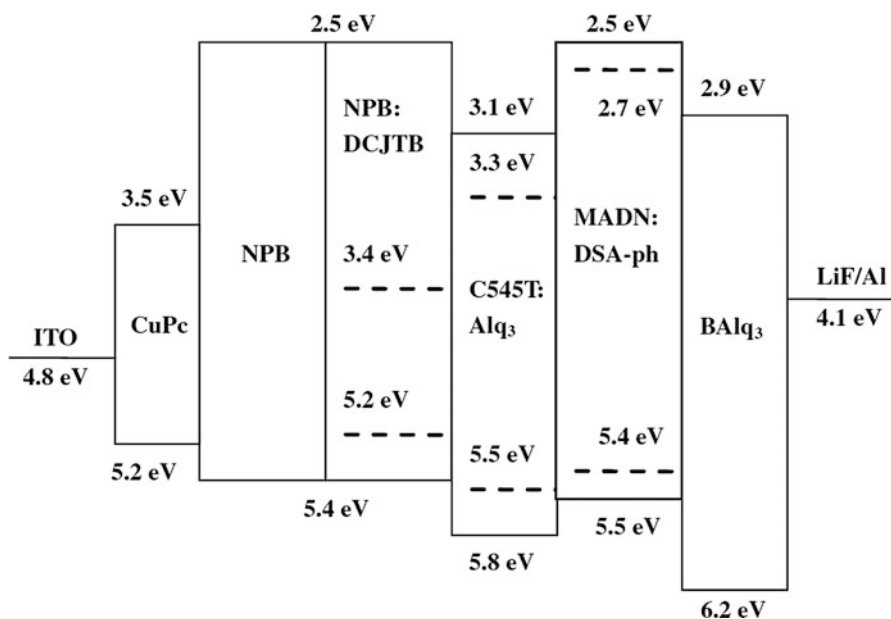
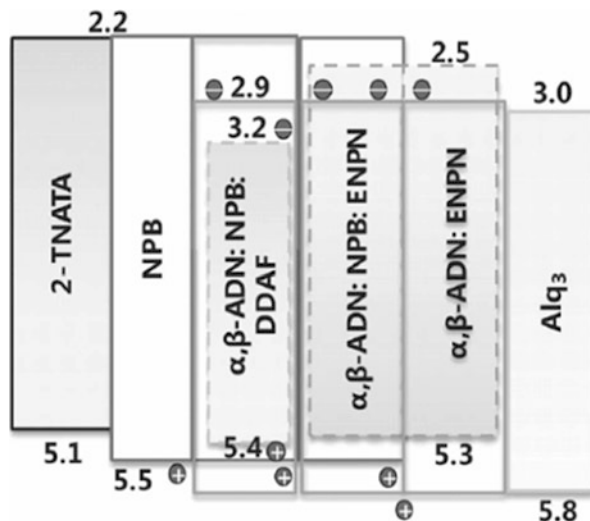


Fig. 5 Structure and energy level diagram of the studied devices (Reproduced from Zhang et al. (2009))

monochromatic emission layers. The red and blue emissive layers were based on 4-(dicyanomethylene)-2-*tert*-butyl-6-(1,1,7,7-tetramethyljulolidin-4-yl-vinyl)-4*H*-pyran (DCJTB) doped *N,N*-di(naphthalene-1-yl)-*N,N*-diphenyl-benzidine (NPB) and *p*-bis(*p*-*N,N*-diphenyl-amino-styryl)benzene (DSA-ph) doped 2-methyl-9,10-di(2-naphthyl) anthracene (MADN), respectively, and the green emissive layer was based on tris(8-hydroxyquinoline)aluminum (Alq₃) doped with 10-(2-benzothiazolyl)-2,3,6,7-tetrahydro-1,1,7,7-tetramethyl-1*H*,5*H*,1 [*H*-(1)-benzopyrroprano(6,7-8-*i,j*)quinolizin-1]-one (C545T), which is sandwiched between the red and the blue emissive layers. Figure 5 gives the structure and the energy level diagram of the studied devices. It is experimentally found that the introduction of a thin C545T-doped Alq₃ green layer between red and blue emission layers plays an important role in the performances of the fabricated white OLEDs. It controls the location of the exciton recombination region well, leading to the high performance of the devices. The devices emitted stable white light with Commission Internationale de L'Eclairage coordinates of (0.41, 0.41) and color rendering index (CRI) of 84 in a wide range of bias voltages. The maximum power efficiency, current efficiency, and quantum efficiency reached 15.9 lm/W, 20.8 cd/A, and 8.4 %, respectively. The power efficiency at brightness of 500 cd/m² still arrived at 7.9 lm/W, and the half-lifetime under the initial luminance of 500 cd/m² is over 3,500 h.

More recently, Y. Qiu's group presented a new strategy to achieve white OLEDs with an extremely long lifetime by wisely controlling the recombination zone (Duan et al. 2011). As the structure shown in Fig. 6, a blue emitting layer of

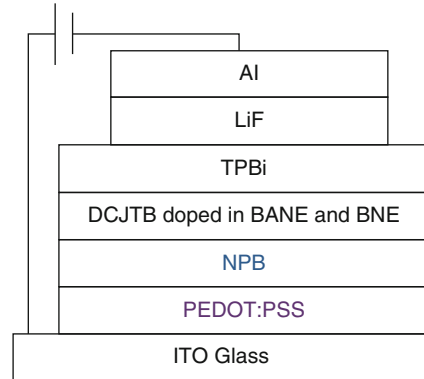
Fig. 6 Device structure of white OLEDs with MHYEL/MHBEL/BEL (Reproduced from Duan et al. 2011)



6,6'-(1,2-ethenediyl)bis(*N*-2-naphthalenyl-*N*-phenyl-2-naphthalenamine doped 9-(1-naphthyl)-10-(2-naphthyl)-anthracene was deposited on top of the mixed host blue-emitting layer to prevent hole penetration into the electron-transporting layer and to attain better confinement of carrier recombination. In this way, a white OLED with a record high lifetime of over 150,000 h at an initial brightness of 1,000 cd/m² was obtained. And its peak current efficiency reached 14.7 cd/A at a high current density of 2,624 A/m². The electroluminescent spectra showed almost no color shifting after accelerated aging. This indicates that these results might be a starting point for further research towards ultra-stable OLED for displays and lightings.

Among the construction of high-efficiency fluorescence white OLEDs, using a single emissive layer composed of a yellow or red fluorescence dye doped in a blue single host or a blue cohost is also an effective method to fabricate high efficient white OLEDs. In this white emission, generally the red light is from the dye dopant, whereas the blue light is from the host (Jou et al. 2008; Yang et al. 2011). As a main result based on single-EML structure, a fluorescence white OLED by using double hole-transporting layers (HTLs) composed of poly(3,4-ethylene-ioxothiophene)-poly-(styrenesulfonate) (PEDOT) and *N,N'*-bis-(1-naphthyl)-*N,N'*-diphenyl-1,10-biphenyl-4-4'-diamine (NPB) was realized (Jou et al. 2008). Figure 7 shows the structure of the developed fluorescence white OLEDs. This device utilized the structure of a cohost single emissive layer with 0.5 wt% red 4-(dicyanomethylene)-2-tbutyl-6-(1,1,7,7-tetramethyljulolidyl-9-enyl)-4*H*-pyran (DCJTb) doped in a mixed host of 25 % *trans*-1,2-bis(6-(*N,N*-di-*p*-tolylamino)-naphthalene-2-yl)ethane (BNE) and 75 % 1-butyl-9,10-naphthaleneanthracene (BANE). It was clearly seen from the experimental results that the thickness of NPB HTL possesses significant effect on device efficiency. When using only PEDOT:PSS as the HTL, excessive holes may have been injected into the EML,

Fig. 7 Device structure of fluorescence white OLEDs with double HTLs (Reproduced from Jou et al. (2008))



causing an electron-insufficient unbalanced carrier injection. As the applied voltage was increased, more excessive holes may have been injected, causing worse unbalanced carrier injection. This may, partly, explain why the corresponding power efficiency observed at $1,000 \text{ cd/m}^2$ was decreased from 11.9 to 9.1 lm/W , a deterioration of 24 %. However, when an NPB layer was added between the PEDOT:PSS and the EML, the formation of a step hole injection barrier enables the blocking of excessive holes and leads to a more balanced carrier injection. As a result, the efficiency was increased. It can be seen that the hole-blocking function of the second HTL NPB was seemingly more marked at high voltage. The power efficiency reached 16.5 lm/W at $1,000 \text{ cd/m}^2$, which is much higher 9.1 lm/W for that without NPB. For this 7.5 nm NPB-composing device, its efficiency deterioration was only 13 % as the observed brightness was increased from 100 to $1,000 \text{ cd/m}^2$. This result fully exhibits the important role of the device structure design in the efficiency improvement for white OLEDs.

With respect to single-unit OLEDs, a tandem OLED structure showing improved current efficiency and lifetime was developed. Tandem OLEDs can be accomplished by vertically stacking several individual electroluminescent (EL) units and driving the entire device with a single power source. In a tandem OLED having N EL units ($N > 1$), the current efficiency (cd/A) can be about N times as high as that of a single-unit OLED. Therefore, the stacked OLED needs only about $1/N$ the current density used in the single-unit OLED to obtain the same luminance, resulting in an operational lifetime of at least N times that of the conventional OLED. In a tandem OLED, all of the EL units are electrically connected in series by a called charge generation layer (CGL) unit between adjacent EL units. The CGL plays an important role in the performance of the fabricated tandem OLEDs. The detailed analysis on the mechanism and construction of CGLs and the main advances on tandem white OLEDs will be given in section “[Tandem White OLEDs](#).” Here, we only give a representative result of all-fluorescence tandem white OLEDs.

Figure 8 shows the structure of a two-stack tandem all-fluorescence white OLED (Spindler and Hatwar 2009). Low-voltage and high-efficiency electron-transporting materials and architecture were used in this device, with the addition of two emission

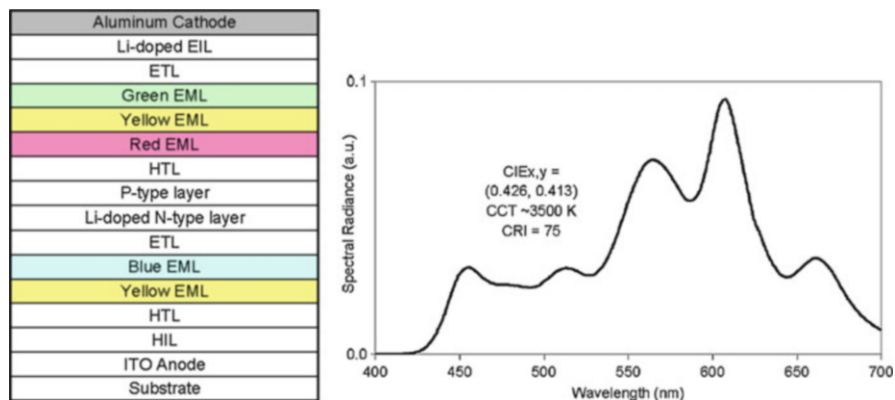


Fig. 8 Device structure (*left*) and EL spectrum (*right*) of the designed tandem fluorescence white OLEDs (Reproduced from Spindler and Hatwar (2009))

layers to provide a broader spectrum that can meet the color requirements for lighting. The EL spectrum is also shown in Fig. 8. This device had a CCT of 3,500 K and a color within the Energy Star specifications for SSL with a CRI of 75. This device also showed high efficiency and long lifetime; a current efficiency of 38 cd/A at a driven voltage of 6.0 V was achieved, resulting in a power efficiency of nearly 20 lm/W. The external quantum efficiency was 13.9 %. The lifetime was estimated to be 140,000 h at an initial luminance of 1,000 cd/m². If considering the light extraction technique, this same tandem white OLED is estimated to achieve over 40 lm/W. The exceptional efficiency and lifetime achieved demonstrates that tandem white OLEDs based on fluorescent emitters are a compelling choice for first-generation OLED solid-state lighting products having a long lifetime as the most critical requirement.

Phosphorescence White OLEDs

The high internal efficiency is important for white OLEDs to be competitive with existing lighting technologies. Among the various concepts for white OLEDs, by far the most effort has been spent on research dealing with devices based solely on phosphorescence-emitting materials. This is probably due to the fact that phosphors inherently offer internal efficiencies of unity (Baldo et al. 1998a), so that in general the only remaining task in device engineering is the distribution of excitons to different emitters for white light emission. For the design of high-performance phosphorescence white OLEDs, not only the used phosphors but also the used hosts and the electron- and hole-transport layers near the emitter must be carefully chosen and effectively matched in energy levels. A key requirement for the host materials of phosphorescence emitters is to have a high triplet level than the phosphors, and besides the high triplet level, the electron-blocking layer and

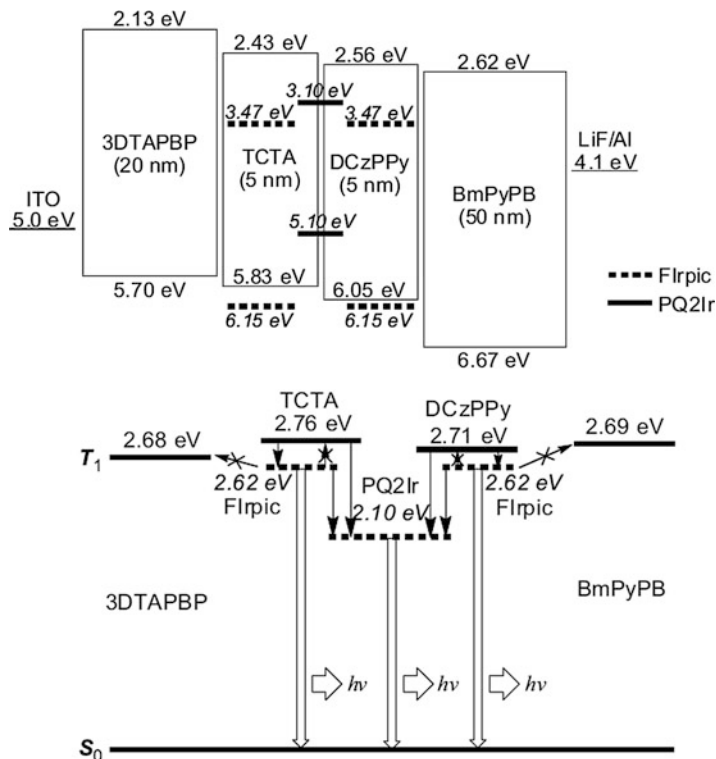


Fig. 9 *Top*: proposed energy level diagrams of the white OLEDs showing the highest occupied and lowest unoccupied molecular orbital energies relative to vacuum. The thicknesses of the constituents used are also shown. *Bottom*: alignment of triplet energy levels for the constituents used in the blue and white OLEDs. Effective triplet energy transfer from host materials to Flrpic and PQ2Ir and good triplet energy confinement are proposed (Reproduced from Su et al. (2008a))

electron-transporting/hole-blocking layer should have small LUMO and large HOMO, respectively. By that, the triplet excitons can be efficiently confined to the emissive states. Comparatively, the excitonic confinement is especially a challenge for blue emitters as they require host materials with widest band gap.

J. Kido's group even reported a high-efficiency two-color all-phosphorescence white OLED (Su et al. 2008a). In their devices, a red dopant iridium(III) bis-(2-phenylquinoly-N, $C^{2'}$)dipivaloylmethane (PQ2Ir) was used as a complementary emitter, and the EML is designed to form a strong carrier- and exciton-confinement structure. Figure 9 shows the device structure and its triplet energy diagram. In order to confine triplet excitons of the blue emitter Flrpic ($T_1 = 2.62$ eV), they composed the device structure solely with materials having higher triplet levels. It can be seen that the LUMO level of hole-transport layer bis(*m*-di-*p*-tolylaminophenyl)-1,1'-biphenyl (3DTAPBP) is 0.30 and 1.34 eV higher than those of 4,4',4''-tri(*N*-carbazolyl)triphenylamine (TCTA) and iridium(III) bis(4,6-(difluorophenyl) pyridinato-N, $C^{2'}$)picolinate (Flrpic), respectively,

confining electrons within the emissive layers. In addition, the HOMO level of electron-transport layer 1,3-bis(3,5-dipyrid-3-yl-phenyl)benzene (BmPyPB) is 0.62 and 0.52 eV deeper than those of 2,6-bis(3-(carbazol-9-yl)phenyl)pyridine (DCzPPy) and FIrpic, respectively, confining holes within the emissive layers. Moreover, the introduction of a double emission layer may lead to a distributed exciton-formation zone across the full emissive layers, reducing the possibility of triplet-triplet annihilation and rendering improved efficiency and reduced efficiency roll-off. Indeed, the white OLEDs showed a high efficiency, even though at higher luminance. The 25 % high external quantum efficiencies along with low driving voltages gave an external power efficiency of 53 lm/W at 100 cd/m², which rolled off slightly to 44 lm/W at 1,000 cd/m² luminance. At higher luminance values of 5,000 and 10,000 cd/m², the external power efficiency of this device yet remained 31 and 24 lm/W, respectively, indicating the effectivity of the designed device structure in confining carriers and excitons in the emission region. The device showed a Commission Internationale de L'Eclairage (CIE) coordinates of (0.341, 0.396) at 100 cd/m² and shifted slightly to (0.335, 0.396) at 1,000 cd/m². This indicates that the insertion of ultrathin orange-light-emitting layers at the DCzPPy/TCTA interface induced sufficient orange light emission as to achieve stable white emission, although the thickness of the orange-light-emitting layers is only a 20th to that of the total emissive layers. The color rendering index (CRI) of the current white OLED is 68. Of course, this low CRI can be resolved by using a deep-blue emitter instead of FIrpic in the current architecture.

Although two-color phosphorescence white OLEDs can show high efficiency, the low CRI is its main problem. In order to increase the color quality of phosphorescent white OLEDs, three primary colors need to be employed in device design. The representative structure is a different EML design with multiple exciton-formation zones while using exactly the same emitter molecules reported by S. R. Forrest's group (Sun et al. 2006; Sun and Forrest 2008). In the 3-EML white OLED, the three hosts were arranged in the order of TCTA/*N,N'*-dicarbazolyl-3,5-benzene (mCP)/*p*-bis-(triphenylsilyl)benzene (UGH2) to form a stepped progression of HOMO levels from TCTA (5.7 eV) to mCP (5.9 eV) to UGH2 (7.2 eV). The energy barrier for holes is larger at the mCP/UGH2 interface compared to the TCTA/mCP interface, resulting in a larger exciton density at the former interface. To reduce the accumulation of the holes, a blue dopant FIr6 at 20 wt% was doped in UGH2 to promote charge injection directly onto the dopant. The adjusting of doping concentrations and thicknesses of the R and G EMLs was used to optimize the white color balance and the efficiency. The detailed structure is shown in Fig. 10. Based on this design, the peak forward-viewing EQE of this 3-EML white OLED is (15.3 ± 0.8)% at $J = 12 \mu\text{A}/\text{cm}^2$ and the peak PE is (38 ± 2)lm/W at $J = 1.3 \mu\text{A}/\text{cm}^2$. The total PE peaks at 64 ± 3 lm/W and rolls off to 34 ± 2 lm/W at 1,000 cd/m². As shown in Fig. 10, three emission peaks are clearly seen. The Commission Internationale de L'Eclairage (CIE) coordinates and the CRI values are (0.37, 0.41) and 81 at $J = 1 \text{ mA}/\text{cm}^2$ and (0.35, 0.38) and 79 at $J = 100 \text{ mA}/\text{cm}^2$, respectively (Fig. 10).

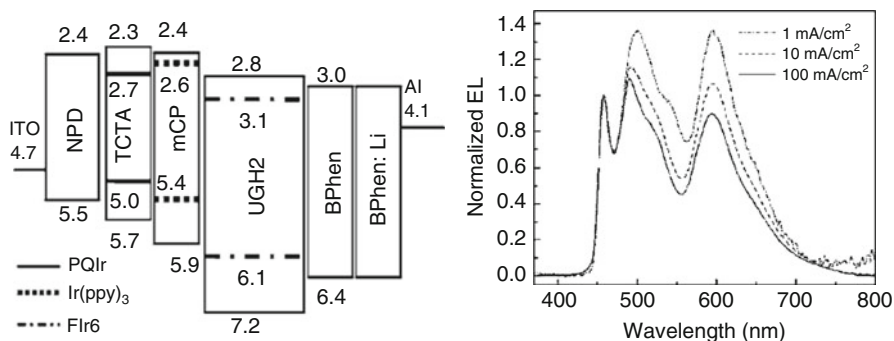


Fig. 10 Proposed energy level diagram (*left*) and normalized EL spectra at various current densities (*right*) of the three-emission-layer white OLED (Reproduced from Sun et al. (2006))

Later, J. Kido's group realized a much more high-efficiency white OLED with three phosphorescent emitters by introducing a newly developed *mer*-tris(*N*-dibenzofuranyl-*N'*-methylimidazole)iridium (III) [Ir(dbfmi)] having blue emission at 445 nm (Sasabe et al. 2010). In this device, a new host material 3,6-bis(diphenylphosphoryl)-9-phenylcarbazole (PO9) with a triplet level of 2.95 eV was used. It was clearly shown that the Ir(dbfmi)-doped PO9 film showed a high PL quantum efficiency value of $70 \pm 1\%$. As expected from the phosphorescent spectrum of PO9, the transient photoluminescence (PL) decay curve of the Ir(dbfmi):PO9 film exhibited an almost single-exponential decay (94 %) with a phosphorescence lifetime of 19.6 μ s at room temperature, indicating the effective suppression of the Ir (dbfmi) exciton quenching in the host. The final structure of the optimized phosphorescence white OLEDs was ITO (130 nm)/TAPC (40 nm)/TCTA (5 nm)/PQ2 Ir (dpm) 2 wt%-doped CBP (1 nm)/Ir(ppy)₃ 6 wt%-doped CBP (1 nm)/Ir(dbfmi) 10 wt %-doped PO9 (10 nm)/B3PyPB (50 nm)/LiF (0.5 nm)/Al (100 nm). The white OLEDs based on this EML sequence reached very high efficiencies of 21.5 % EQE and 43.3 lm/W at 1,000 cd/m². Even at 5,000 cd/m², efficiencies of 16.7 % and 28.7 lm/W still remained. The color quality was improved to a CRI = 80.2 with CIE coordinates of (0.43, 0.43), which is acceptable for the illumination light source.

As shown in many experiments, the use of host materials with extremely wide band gap to match blue emitter in turn increases the operational voltage of the resulting white OLEDs, ultimately leading to the reduction of power efficiency. K. Leo's group presented an all-phosphorescence white OLED structure that combines a novel concept for energy-efficient photon generation (Reineke et al. 2009). The key feature of the OLED layer structure is the positioning of the blue phosphor within the emission layer and its combination with a carefully chosen host material. The blue host-guest system is surrounded by red and green sub-layers of the emission layer. Energetically, the triplet energy of the blue emitter is resonant to its host so that the blue phosphorescence is not accompanied by internal triplet energy relaxation before emission, rendering the red or green species to harvest unused excitons without energy loss. The energy level diagram of their EML

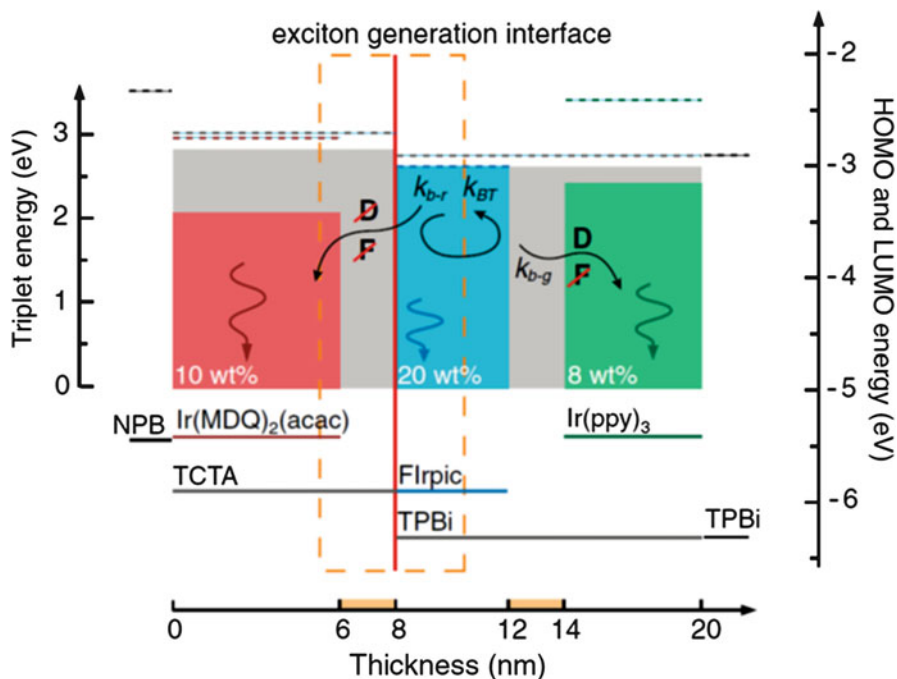


Fig. 11 Energy level diagram of the phosphorescent white emission layer concept. *Dashed lines* are LUMO levels and *solid lines* HOMO levels. The *filled boxes* indicate the respective triplet levels of host (gray) and emitter (colored) materials. The *dashed box* indicates the exciton-formation zone. *D* and *F* denote Dexter and Förster energy transfers, respectively. Furthermore, the rates refer to *blue-to-red* transfer k_{b-r} , back transfer k_{BT} , and *blue-to-green* transfer k_{b-g} (Reproduced from Reineke et al. (2009))

architecture is depicted in Fig. 11. It is clear that the exciton-formation region is at the interface of a double-emission-layer structure. Obviously, for holes and electrons, the emission layer is nearly barrier-free until they reach the region of exciton formation, which keeps the operating voltage low. And placing the blue sub-EML at the position of exciton generation, the total layer thickness can be reduced because the exciton density is accordingly higher. Thus, the TPBi:FIrpic layer is only 4 nm thick. This is also favored in reducing the operational voltage. Actually, the use of reduced-band-gap materials will also further reduce the operational voltage of the device. It can be seen that with the EML structure, very low voltages of 3.22 and 3.95 V are obtained for 1,000 and 10,000 cd/m^2 , respectively, operating close to the thermodynamic limit. The corresponding device efficiencies are 13.1 % EQE and 30 lm/W at 1,000 cd/m^2 with CIE color coordinates of (0.45, 0.47) and CRI of 80. By combining a carefully chosen emitter layer with high-refractive-index substrates and using a periodic out-coupling structure, the device can emit a power efficiency of 90 lm/W at 1,000 cd/m^2 and has the potential to be raised to 124 lm/W if the light out-coupling can be further improved.

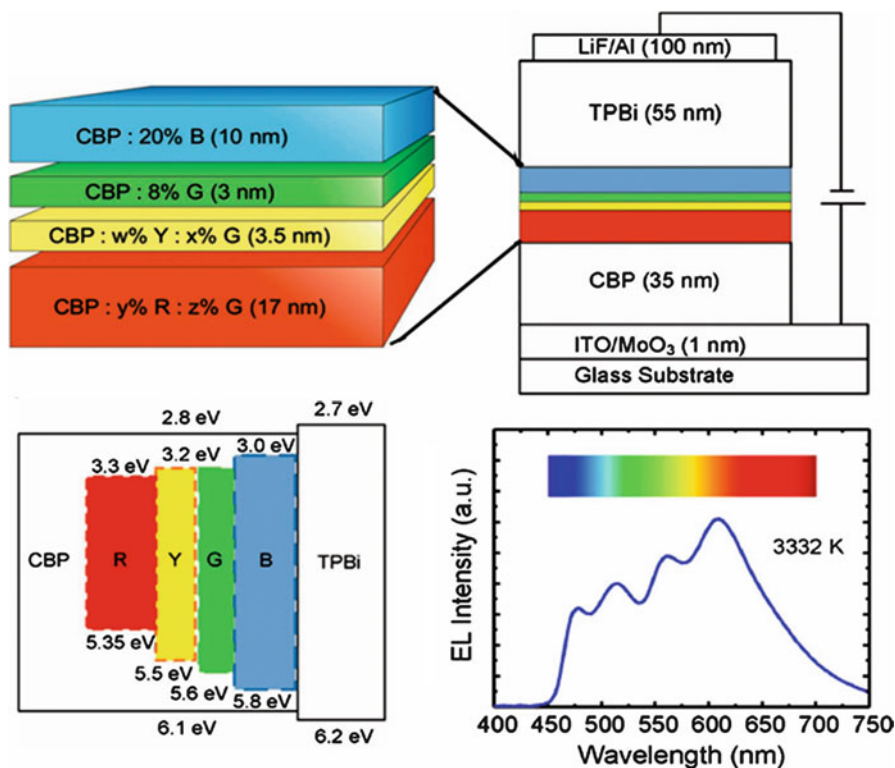


Fig. 12 Schematic illustration (*top*) of four-color white OLED device structure and its corresponding energy level diagram (*bottom left*) and EL spectrum (*bottom right*) (Reproduced from Chang et al. (2013))

Recently, Z. H. Lu's group designed a four-color all-phosphorescence white OLED employing a novel device design principle utilizing molecular energy transfer or, specifically, triplet exciton conversion (TEC) within common organic layers in a cascaded emissive zone configuration (Chang et al. 2013). The schematic illustration of the four-color white OLED device structure and its corresponding energy level diagram are shown in Fig. 12. Here, 2,2',2''-(1,3,5-benzinetriyl)-tris(1-phenyl-1-*H*-benzimidazole) (TPBi) serves as the electron-transport layer (ETL), and 4,4'-bis(carbazol-9-yl)biphenyl (CBP) functions as a hole-transport layer (HTL) and as a triplet host. ITO/MoO₃ anode and LiF/Al cathode are applied. In this configuration, it is clearly seen that the majority of excitons will be generated near the CBP/TPBi interface (on both sides) before being harvested by the emitters (i.e., recombination occurs) on the CBP side. As both CBP and TPBi are wide-energy-gap materials with high triplet energies, the generated excitons can be well confined onto the emitters. Since the blue emitter, FIrpic, has the closest energy levels to both materials, direct exciton formation on the blue dopant is unlikely and it is critical to place the blue emitter closest to the CBP/TPBi interface to harvest excitons first. Other lower

energies like green, yellow, and red emitters are placed sequentially next to blue to harvest excitons in a cascaded fashion. This cascaded design using a single host allows for only a single site for exciton generation and recombination without introducing other barrier layers. It is also important that there is no interlayer between two adjacent emitting layers so that the surplus excitons can readily diffuse into the adjacent layer with an emitter having a lower energy. Because a high-energy green phosphor with excellent exciton trapping capability is incorporated into the yellow and red emissive layers, the significant spectral overlap between the green dopant emission spectrum and the red as well as yellow dopant absorption spectra then leads to a strong exciton energy transfer from the green dopant to yellow and red dopants. As a result, the yellow and in particular red emissions in EL spectrum are greatly increased, leading to a high CRI of 85 at lighting luminance of 5,000 cd/m² with CIE coordinates of (0.44, 0.45) (Fig. 12). This enhancement in emission intensity also yields a superior device external quantum efficiency of 23.3 % and power efficiency of 31.0 lm/W, respectively, at 1,000 cd/m². Aided with a lens-based out-coupling enhancement, a high power efficiency of 76.0 lm/W should be obtained.

From the device engineering point of view, it is desirable to simplify the white OLED structure. The simplest method is the utilization of a single emissive layer within the device. The key feature of this method is the employment of one EML comprised of a large-band-gap host doped with two or more emissive dopants to generate white light. However, in most conventional single-EML white OLEDs, sequential energy transfer initially from the short-wavelength dopant always dominates the main emission mechanism. This not only produces exciton energy losses, leading to low device efficiency, but also causes notorious spectra variations associated with the chromophore saturation. In order to resolve this issue, D. G. Ma's group presented a novel concept in structure to construct high-efficiency all-phosphorescence white OLEDs with a single emissive layer (Wang et al. 2009b). In their devices, as shown in Fig. 13, the EML was formed by co-doping two phosphorescent dyes, namely, iridium(III)[bis(4,6-difluorophenyl)pyridinato-N,C2']picolinate (FIrpic) for blue emission and bis(2-(9,9-diethyl-9H-fluoren-2-yl)-1-phenyl-1*H*-benzoimidazol-N,C³)iridium (acetylacetonate) ((fbi)₂Ir(acac)) for orange emission, into 1,3-bis(9-carbazolyl)benzene (mCP) host. The key feature consists of careful manipulation of two exciton-formation modes, namely, host-guest energy transfer (for the blue dopant) and direct exciton formation (for the orange dopant) within an energetic, well-like, single emissive region. In fact, this unique strategy creates two parallel pathways to channel the overall excitons to both dopants within the EML, leading to an improved charge balance and further reduction of the unfavorable energy losses.

The final optimized single-EML white OLED showed a maximum forward-viewing EQE of 19.3 % at a current density of 0.015 mA/cm² and slightly decreases to 16.1 % at a typical display luminance of 100 cd/m². The maximum forward-viewing power efficiency reached 42.5 lm/W, corresponding to a maximum total power efficiency of 72.2 lm/W, which rolls off to 32.3 lm/W at a high brightness of 500 cd/m². More importantly, the device exhibits a total IQE of almost 96 %,

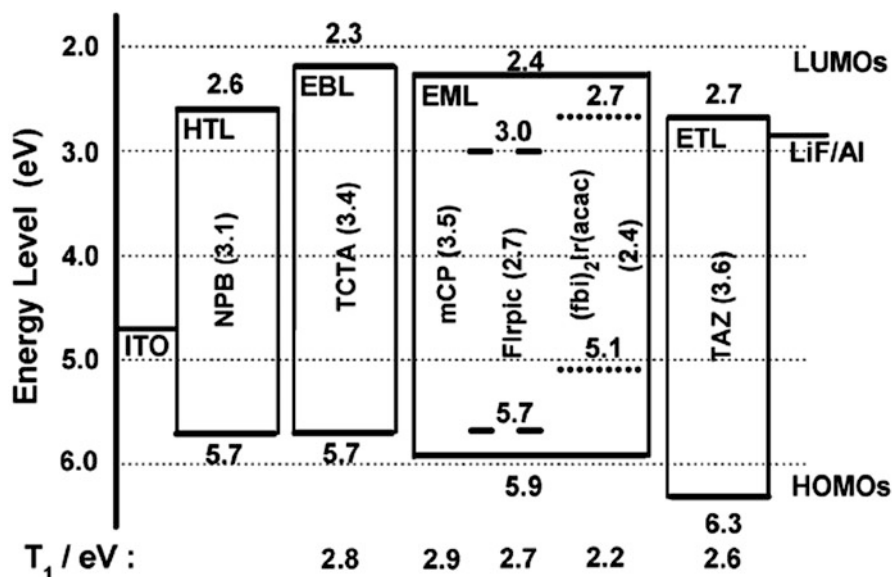


Fig. 13 Energy diagram and device structure of the all-phosphor dual-color white OLED. The numbers in parentheses indicate the energy gaps of materials. T_1 triplet energy. EBL refers to electron-blocking layer (Reproduced from Wang et al. (2009b))

indicating that nearly all the injected carriers are converted into photons. However, it is clearly seen that the emission from (fbi)₂Ir(acac) in the EL spectra slightly decreases with an increase of driven voltage.

To resolve the spectral shift originated from carrier trapping in orange dopant sites in the above single-EML white OLEDs, D.G. Ma's Group designed a spectrally stable all-phosphorescence white OLED, where the EML is composed of a single host by separately doping (fbi)₂Ir(acac) orange dopant and FIrpic blue dopant in adjacent two regions (Zhao et al. 2012a). It is clearly seen that the spectral stability is indeed improved greatly (Fig. 14) but yet keeping higher efficiency. As shown in Fig. 14, in this device, the direct recombination of orange dopant molecules by charge carrier trapping is greatly suppressed by controlling the exciton recombination region within the blue region. The red emission is due to energy transfer.

Based on the above design concept, three-color all-phosphorescence high-efficiency white OLEDs with high CRI and stable spectrum were also developed (Wang et al. 2009a; Wang and Ma 2010). This EML was based on 3-bis(9-carbazolyl)benzene (mCP) as the host and separately dopes red, green, and blue regions. The utilization of single host will greatly reduce structural heterogeneity and facilitate charge injection and transport between different emissive regions. Two types of white OLEDs are proposed: R-G-B (the primary color emitters are separately distributed by red-green-blue sequence) and RG-B (the red and green emitters are combined into one region) devices. For R-G-B white OLEDs, the structure shown in the inset of Fig. 15a, a maximum forward-viewing power efficiency (PE) and external quantum efficiency (EQE) of 41.3

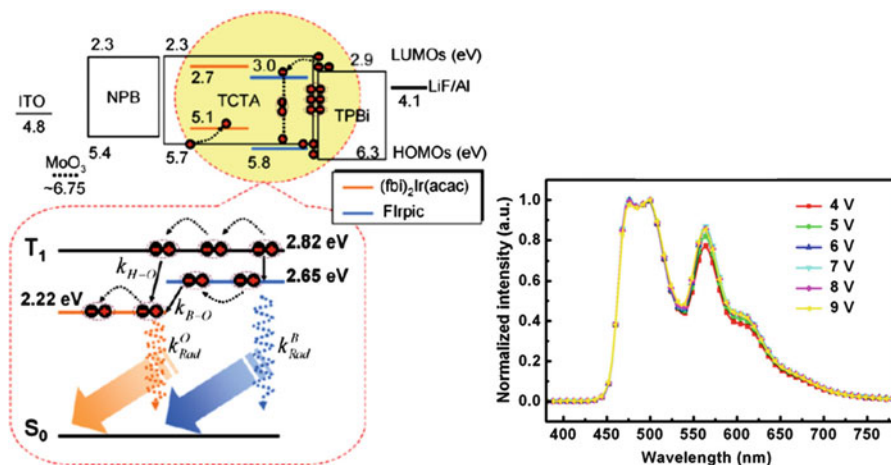


Fig. 14 Energy level diagram and EL processes (*left*) of the proposed two-region single-EML white OLED and its EL spectra at different voltages (*right*) (Reproduced from Zhao et al. (2012a))

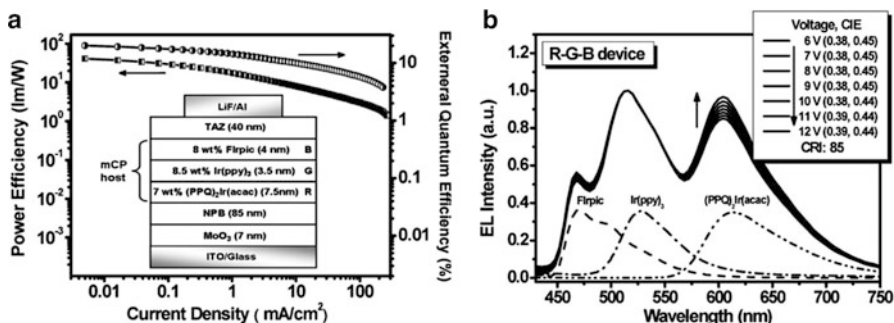


Fig. 15 (a) Power efficiency and EQE versus current density of an R-G-B white OLED. *Inset* shows schematic architecture of the device. (b) EL spectra of the white OLED at different voltages (Reproduced from Wang et al. (2009a))

lm/W and 20.1 % were obtained, respectively (Fig. 15a). As shown in Fig. 15b, the device color rendition is also impressive. The EL spectrum covers all wavelengths from 450 to 750 nm, and the CRI is calculated to reach as high as 85. Notably, this WOLED possesses high color stability. When the luminance changes from 500 to 10,000 cd/m² (which corresponds to an applied voltage of 6–12 V), the variation of the Commission Internationale de L'Eclairage (CIE) coordinates is rather small, CIE (0.38–0.39) and CIE (0.45–0.44), thus revealing superior device efficiency/CRI/chromatic-stability trade-off. For RG-B white OLEDs (see inset of Fig. 16a), as expected, the device efficiency is still maintained at a high level, that is, the maximum forward-viewing PE and EQE are 37.3 lm/W and 19.1 % (Fig. 16a). Surprisingly, the obtained EL spectrum turns out to be

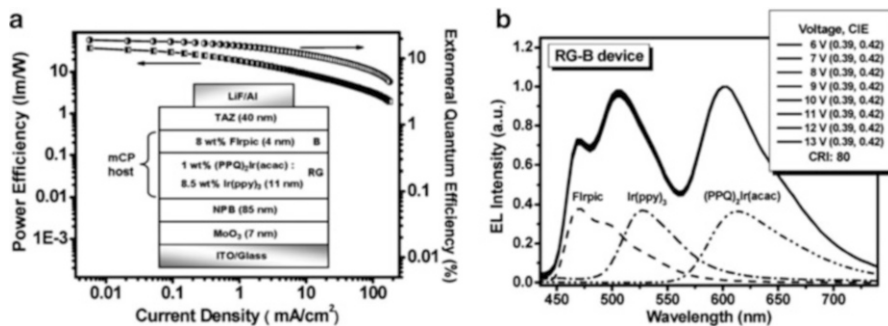


Fig. 16 (a) Power efficiency and EQE versus current density of a RG-B white OLED. *Inset* shows schematic structure of the device. (b) EL spectra of the white OLED at different voltages (Reproduced from Wang et al. (2009a))

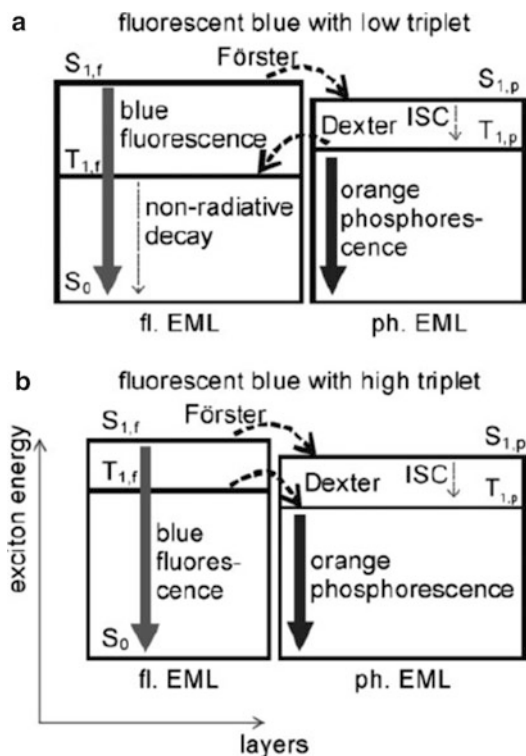
rather stable (Fig. 16b). In a wide range of operational voltages, the CIE coordinates of (0.39, 0.42) are constant, and the CRI remains at 80.

Fluorescence/Phosphorescence Hybrid White OLEDs

Although there are the advantages of long lifetime of fluorescence white OLEDs and high efficiency of phosphorescence white OLEDs, the low efficiency of fluorescence emitters and short lifetime of blue phosphorescence emitters greatly limit their applications. At present a relatively better method is the fluorescence/phosphorescence hybrid structures to realize white OLEDs, from which the white emission is composed of green and red (or orange) phosphorescence and blue fluorescence. Since these approaches avoid using the relatively unstable blue phosphors, yet still realize the harvesting of both singlet and triplet excitons, thus theoretically allowing 100 % exciton use for high efficiency (Schwartz et al. 2007, 2009; Albrecht and Bassler 1995). Obviously, the hybrid devices may be intrinsically superior to the fully phosphorescent approach when it comes to applications where high device lifetime is also required.

As we know, singlet and triplet excitons are always generated together at the same place within OLEDs. Therefore, a proper distribution of those two singlets and triplets to the fluorophore and phosphors needs a mechanism to separate them spatially. The different diffusion of singlet and triplet exciton is the basis of the separating mechanism. Comparatively, triplets can have a much larger diffusion length (tens of nm) than singlets (<10 nm) due to their intrinsically longer lifetime (Segal et al. 2003). Therefore, one has to place the fluorophore in close vicinity of the exciton generation zone, whereas the phosphors are placed at greater distance to realize the harvesting of both triplet and singlet excitons in hybrid devices. This ensures that the singlet excitons are by the majority used for radiative emission from the fluorophore and are not parasitically transferred to the phosphors, whereas the

Fig. 17 Exciton energy diagram for a blue fluorophore and an orange phosphor. Two different scenarios are possible: **(a)** the fluorophore has a large singlet–triplet splitting which results in its triplet state lying below the triplet state of the phosphor; **(b)** the fluorophore has a small singlet–triplet splitting and thus a higher lying triplet state. In this case, its intrinsically non-radiative triplet excitons can still be harvested for light emission when transferred to the phosphor (Reproduced from Schwartz et al. (2009))



majority of triplets still reaches the phosphors where they are harvested for radiative emission.

For the realization of hybrid white OLEDs, there are two cases to be considered in device structure design. If the used blue fluorescent emitters have triplet levels lower than the respective T_1 states of the phosphorescent emitters, the blue triplet level typically becomes a prominent quenching channel. As shown in Fig. 17a, the triplet exciton energy transfer from phosphors to the fluorophore will occur, which is a lost process of non-emission. There are two channels for the exciton loss: (i) The direct formation of triplet excitons on the fluorescent triplet level, which is not emissive for fluorescence. The only way to reduce this channel is to reduce singlets. However, this will decrease the fluorescent intensity at the same time. (ii) The energy transfer from the phosphorescent system to the fluorescent triplet level will greatly reduce the quantum efficiency of the phosphorescent emitter. As shown, obviously, the triplet quenching introduced by non-radiative energy transfer can easily be prevented by introducing a thin interlayer between fluorescent and phosphorescent systems. Because the energy transfer is a Dexter type, requiring orbital overlap, interlayer thicknesses in the range of 2 nm are sufficient (Schwartz et al. 2006). However, if the triplet state of the blue fluorophore lies higher than the triplet state of at least one of the used phosphors, this results in a completely different situation. In this case, there is not only absence of phosphorescence quenching but also the intrinsically

non-radiative triplets on the blue fluorophore can now be harvested for light emission by transferring them to this phosphor (see Fig. 17b). Thus, the singlet excitons are directly used by fluorescence emission without energy loss, whereas the triplet excitons from the blue fluorophore are then completely harvested by the phosphors via diffusing through the bulk fluorescent blue emission layer, truly yielding an internal quantum efficiency of unity. This case has been well realized by incorporating a fluorescence blue emitter with a small singlet–triplet splitting gap (Schwartz et al. 2007).

The hybrid white OLEDs comprising an interlayer was first made by K. Leo's group (Schwartz et al. 2006). This device consisted of a hole-transporting phosphorescent multilayer system for red and green and an electron-transporting 2,2',7,7'-tetrakis(2,2-diphenylvinyl)spiro-9,9'-bifluorene (Spiro-DPVBi) layer for blue emission. By introducing a thin layer of coevaporated TCTA and TPBi between the phosphorescent and the fluorescent region, which is more than enough at 2 nm, both singlet and triplet excitons are confined efficiently in EMLs, whereas charge carriers still pass easily in this interlayer. It can be seen that by introducing the interlayer, the EQE is almost doubled from 4.5 % to 8.0 % at 1,000 cd/m². The corresponding CIE coordinates of the device with interlayer are (0.47, 0.42), the color rendering index is as high as 85, and the luminous efficacy reaches 13.7 lm/W at 1,000 cd/m².

Later, S. R. Forrest's group presented a device concept for hybrid white OLEDs that show improvement in the exciton distribution within the emission layer (Sun et al. 2006). This device is based on the fluorescent blue emitter BCzVBi and the green and red phosphors Ir(ppy)₃ and PQIr, respectively, all embedded into a common host CBP at different spatial locations. This device structure can realize 100 % internal quantum efficiency as a result of a decoupling of singlet and triplet exciton channels, where only a fraction of the 25 % singlets are used for fluorescence whereas the remaining 75 % of the generated triplets are directed to the green and red phosphors. It can be seen that the a maximum forward-viewing EQE of 11.0 % at a current density 1.0 mA/cm² is achieved and decreases only slightly to 10.8 % at a high forward-viewing luminance of 500 cd/m². This device gives a maximum forward-viewing power efficiency of 22.1 lm/W and has maximum total efficiencies of 37.6 lm/W and 18.7 %. At a practical surface luminance of 500 cd/m², the power efficiency reaches 23.8 lm/W. The intrinsic singlet-to-triplet ratio and the separation of the channels in harvesting the two excitonic species gives a well-balanced and largely current-independent color rendition, resulting in a color rendering index of CRI = 85 at all current densities studied. The Commission Internationale d'Eclairage (CIE) coordinates have a negligible shift from (0.40, 0.41) at 1 mA/cm² to (0.38, 0.40) at 100 mA/cm². This device concept is based on the experimental finding that excitons are mainly formed at both EML interfaces to adjacent transport layers, forming a U-shaped exciton generation profile. The unique architecture of the device lies in the harvesting of singlet and triplet excitons along completely independent channels, and hence the transfer from host to dopant for both species can be separately optimized to be nearly resonant, thereby minimizing energy losses while maintaining a unity IQE.

Recently, D. G. Ma's group evaluated the hybrid white OLED with interlayer by broadening the exciton recombination zone to further improve the efficiency, spectral stability, and efficiency roll-off at high luminance (Zhao et al. 2012b, c). One of their devices, the EML, is composed of a red phosphorescent layer and a blue-green layer of a blue light-emitting fluorescent host doped with a green phosphor separated by a bipolar interlayer. This device utilized well the photoluminescence emission property of blue fluorescence host Bepp₂, thus the simultaneous emission of a blue light from the fluorescent host Bepp₂, and a green light from the phosphor was realized by controlling the doping concentration of the green phosphor in the Bepp₂ host. The resulting hybrid white OLEDs achieved a CRI of 90 and kept a rather stable spectral emission with Commission Internationale de L'Éclairage coordinates of (0.42, 0.44) independent of driving voltages. Furthermore, the hybrid white OLED also exhibited a high efficiency that the maximum current efficiency, external quantum efficiency, and power efficiency reached 29.4 cd/A, 13.8 %, and 34.2 lm/W and still remained at 25.4 cd/A, 11.9 %, and 23.0 lm/W at 1,000 cd/m², respectively. The simple design on the high-performance hybrid white OLED with wise control of exciton recombination region may shed light on the practical method of future white OLEDs for lighting.

However, it is clearly seen that the introduction of interlayer in the above hybrid white OLED doubly increases the operational voltage, resulting in the reduction of power efficiency, and the transfer of the triplet exciton energies across the interlayer will also produce additional losses, further decreasing the efficiency. These problems are avoided from the novel design by introducing a blue fluorophore with a small singlet–triplet splitting in the hybrid white OLED structure (Schwartz et al. 2007). By virtue of the relatively higher triplet state of the blue fluorophore, its non-radiative triplet energy can intrinsically transfer to the orange phosphor without losses even without the interlayer. In the hybrid white OLEDs, the blue fluorophore is called *N,N'*-di-1-naphthalenyl-*N,N'*-diphenyl-[1,1':4',1'':4'',1-quateryphenyl]-4,4-diamine (4P-NPD). It not only has a high triplet level of approximately 2.3 eV and small splitting of ~0.6 eV, which is sufficiently high to excite the red phosphor Ir (MDQ)₂(acac) (T₁ = 2.06 eV), but also shows a high quantum yield of 0.94, which effectively emits blue light. The energy level diagram of the proposed hybrid white OLED structure and its working principle are shown in Fig. 18. The device achieved an external power efficiency of 22.0 lm/W, corresponding to 10.4 % external quantum efficiency. In a control measurement in an integrating sphere, the forward direction power efficiency yields 23.3 lm/W as well as color coordinates that are only slightly shifted by (−0.01, +0.01). By applying a microlens out-coupling foil, the light emission in forward direction is continuously enhanced, resulting in 28.0 lm/W (12.9 %) at 1,000 cd/m². Again, the color coordinates are only slightly shifted by (−0.01, +0.01). The measurement of the total external efficiency within an integrating sphere then gives 57.6 lm/W (20.3 %) at 100 cd/m² and 37.5 lm/W (16.1 %) at 1,000 cd/m². The device also has a good color rendering index of 86, making it well suited for lighting applications.

Based on the high-efficiency blue fluorophore 4P-NPB, recently D. G. Ma's group improved the device structure to further enhance the performance of the

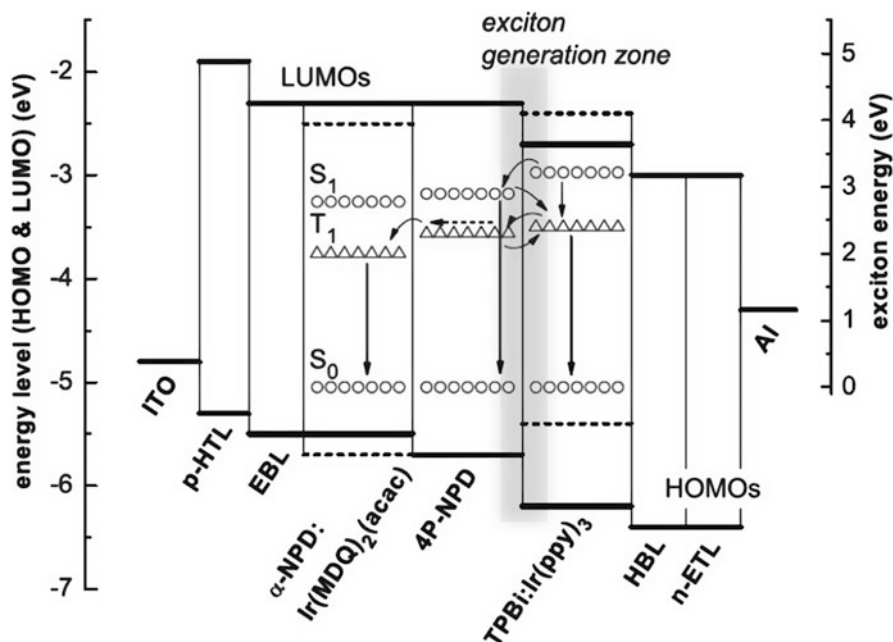


Fig. 18 Energy level diagram [HOMOs and LUMOs (*lines*) and triplet energies (*open symbols*)] of the hybrid white OLED making use of the triplet harvesting concept (Reproduced from Schwartz et al. (2007))

hybrid white OLEDs (Sun et al. 2014). Figure 19 gives the energy level diagram of the proposed hybrid white OLED and its working processes. The importance of constructing the hybrid white OLEDs was the selection of guest emitters and its host. In this structure, the 4P-NPD was used as a blue dopant, which was doped in TCTA: TmPyPB mixed host by a low concentration of 2 %. The role of low 4P-NPB concentration will greatly minimize the negative influence of Dexter-type transfer that favors the formation of the non-luminescent triplet excited state of 4P-NPD, thus significantly suppressing the intrinsically mutual quenching between the fluorescent emitter and phosphorescent emitter(s). It is clearly seen that the optimized device exhibited very impressive EL performance with a turn-on voltage of 3.1 V. The maximum EQE, current efficiency (CE), and PE of the device were 19.0 %, 45.2 cd/A, and 41.7 lm/W, respectively. Furthermore, at the practical brightness of 1,000 cd/m², they yet remained as high as 17.0 %, 40.5 cd/A, and 34.3 lm/W, exhibiting less pronounced efficiency roll-off. The EL spectrum of this device turned out to be rather stable within the investigated brightness range, indicating balanced exciton generation. Furthermore, the device also gave off white light with a high CRI of 82. All these indicate the advantages of the design concept in this hybrid white OLED structure.

As we see, the EML structures for hybrid white OLEDs with and without interlayer described above are relatively complicated yet, which may decrease the

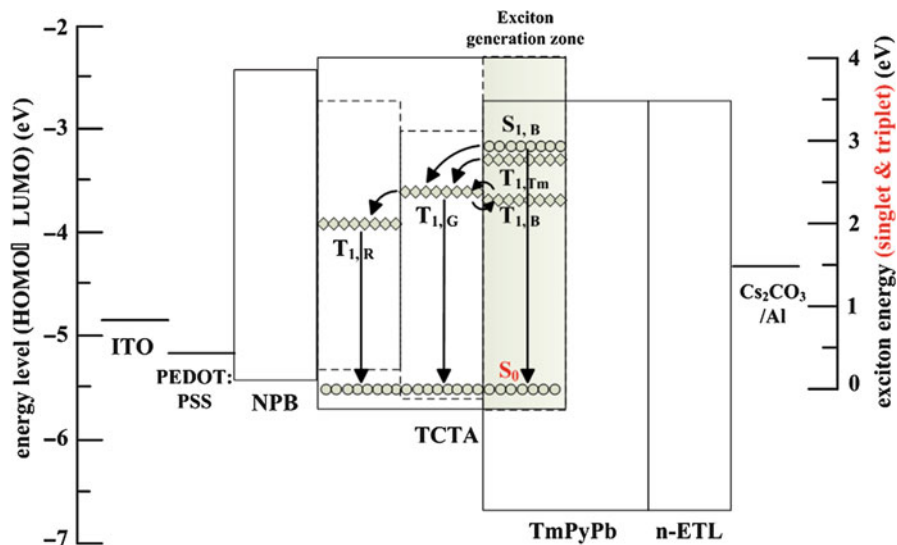


Fig. 19 Energy level scheme for materials used in this hybrid white OLED, and exciton (S_0 , S_1 , and T_1) energy diagram of the emitter layers. The gray-filled rectangle represents the main exciton generation zone. R, G, B, and Tm represent Ir(MDQ)₂(acac), Ir(ppy)₂(acac), 4P-NPD, and TmPyPb, respectively. Solid lines and dashed lines correspond to HOMO and LUMO energy levels, respectively; circles and diamonds refer to exciton (S_0 , S_1 , and T_1) energies, respectively (Reproduced from Sun et al. (2014))

reproducibility and raise the fabrication cost. Therefore, it is urgently expected to develop simple EML structures for the hybrid white OLEDs. One single-EML structure to construct high-efficiency white OLEDs should be the best way (Schwartz et al. 2009). The general way is to dope red (yellow) or red and green phosphorescence dopants directly into the fluorescent blue emitter, where the red and green light are from the phosphors and the blue light is then from the fluorophores. This means that the blue fluorophore should not only possess efficient blue fluorescence but also has to satisfy requirements for a phosphorescent host including high triplet energy, transport property, and good thermal stability. Recently, some advances on the achievement of high-efficiency hybrid white OLEDs with single EML due to the breakout of high efficient blue fluorophores with high triplet level have proven the feasibility of using the simple single-EML structure to construct high-performance hybrid white OLEDs (Ye et al. 2012; Zheng et al. 2013; Chen et al. 2012b). One of best results is the single-EML hybrid white OLED is based on an ideal sky-blue fluorophore, 2,8-di[4-(diphenylamino)phenyl] dibenzothiophene-*S,S*-dioxide (DABDT), as host (Ye et al. 2012). This device was fabricated by doping a common orange phosphorescent dopant, tris(2-phenylquinoline) iridium(III) (Ir(2-phq)₃), in this DABDT as the EML. A double-color white light emission was well realized by controlling the Ir(2-phq)₃ doping concentration. It was experimentally demonstrated that DABDT is a good blue fluorescent emitter with high PL quantum efficiency of 53 %, high triplet level of 2.38 eV, and good thermal stability

($T_g = 123\text{ }^\circ\text{C}$). The finally optimized device was ITO/NPB (30 nm)/TCTA (10 nm)/x 0.1 % Ir(2-phq)₃:DADBT (30 nm)/TPBi (30 nm)/LiF (1.5 nm)/Al. The white device shows excellent EL performance with a low turn-on voltage of 2.4 V and a maximum total EQE, CE, and PE of 26.6 %, 53.5 cd/A/, and 67.2 lm/W, respectively. At a luminescence brightness of 1,000 cd/m², the total EQE and PE remains to be 21.2 % and 33.5 lm/W.

Later, using the same method, three-color single-EML hybrid white OLEDs were successfully fabricated by simultaneously doping green phosphor and red phosphor into fluorescent blue emitters, such as 7-(diphenylamino)-4-methoxycoumarin (DPMC), i[4-(4-diphenylaminophenyl) phenyl]sulfone (DAPSF) (Zheng et al. 2013), and bis[2-(2-hydroxyphenyl)-pyridine] beryllium (Bepp₂) (Chen et al. 2012b). The hybrid white OLED based on DAPSF emitted a maximum power efficiency of 57.3 lm/W and a power efficiency of 24.4 lm/W at 1,000 cd/m² (Zheng et al. 2013). Using Bepp₂ as fluorescent blue host, the fabricated hybrid white OLEDs achieved a maximum forward-viewing power efficiency of 46.8 lm/W. The efficiency roll-off was greatly improved, only dropping to 37.8 and 30.3 lm/W at luminance of 100 and 1,000 cd/m², respectively. The device also showed a high CRI of 90 (Chen et al. 2012b). This indicates that single-EML structures can also achieve high-efficiency white OLEDs by the choice of high-efficiency fluorescence blue emitter and the controlling of low concentration of red and green phosphors in the blue host.

Tandem White OLEDs

Tandem OLEDs are technologically interesting because not only can the luminance and current efficiency be improved linearly with the number of electroluminescent (EL) units in the tandem OLED but also leakage current and breakdown of the electric field can be avoided due to the higher luminance at a low current density and the thicker organic films, resulting in a long lifetime. Importantly, state-of-the-art tandem OLEDs are very easy to vertically stack either individual red, green, and blue emission units or multiple white emission units in series via CGLs to achieve white emission (Tyan et al. 2009; Chen and Ma 2012).

The concept of tandem OLEDs was first proposed by Kido et al. (Matsumoto et al. 2003). They are fabricated by vertically stacking two or more individual EL units and driven entirely by a single power source. All of the EL units in the tandem OLED are electrically connected in a series by inserting a charge generation layer (CGL) between adjacent EL units. Obviously, the CGL plays a critical role in the realization of high-performance tandem OLEDs since it functions as both an internal anode and a cathode to generate intrinsic charge carriers and to facilitate opposite electrons and holes being injected into the adjacent sub-OLED units. Although amounts of CGLs have been developed to use tandem OLEDs (Chen and Ma 2012), the problems still remain including complicated processing and limited material combination. More importantly, since the driving voltage consumed by conventional tandem devices scales linearly with the number of electroluminescent

units, the resulting power consumption would be the same for both the single-unit and tandem OLEDs to obtain the same luminescence, this means that the power efficiency cannot be increased for such tandem devices when using general CGLs. However, it is well known that power efficiency is one key to the commercial realization of a lighting source. Therefore, developing new CGL structures has become an important research topic in this field.

Recently, D. G. Ma's group found (Chen and Ma 2012; Chen et al. 2012c) that intrinsic organic semiconductor heterojunctions (OHJs), a bilayer structure composed of a *p*-type organic semiconductor and an *n*-type organic semiconductor, can be used as the CGL to significantly enhance the power efficiency of the fabricated tandem OLEDs, which was previously suggested to be difficult for tandem devices. It is clearly shown that the novel design concept of OHJ-based CGLs is superior to that of conventional CGLs. The utilization of OHJ CGLs in tandem OLEDs enhances not only doubly the luminance and current efficiency but also significantly the power efficiency (Chen et al. 2011, 2012a; Chen and Ma 2012; Chen et al. 2012c, d). The effect of OHJs as CGL on improving the power efficiency was firstly verified in red, green, and blue tandem OLEDs based on C₆₀/pentacene (H₂Pc, ZnPc, and CuPc) OHJs as CGL. The experimental results clearly showed that the power efficiency of all the devices was enhanced with respect to the single-unit device, indicating the universal method of the design concept of OHJs as CGL for improving efficiency, especially the power efficiency, of tandem OLEDs. The mechanism investigation demonstrated that OHJ allows for an interfacial electron redistribution to supply high-density free charges, thus efficiently decreasing the voltage drop across it. It can be seen that the role of OHJs as CGL is not only related to the mobility of used organic semiconductors but also strongly dependent on energy level position between *p*-type and *n*-type organic semiconductors in OHJ (Chen et al. 2012a).

It is well known that semiconductors are defined by their unique electric conductivity behaviors and can be classified into *p*-type and *n*-type semiconductors. A semiconductor heterojunction is the interface that occurs between *p*-type and *n*-type semiconductors; it is always advantageous to engineer the electronic energy bands in many solid-state devices, including semiconductor light-emitting diodes, lasers, solar cells, and transistors. In fact, the concept of semiconductor heterojunctions had already been proposed, and the energy-band profiles follow the Anderson model (Sharma and Purohit 1974). To date, all inorganic optoelectronic devices are based on this kind of semiconductor heterojunction. The most familiar heterojunction type in inorganic semiconductors is the depletion mode. In this type of heterojunction, a depletion junction is formed on either side of the heterojunction interface; namely, the positive charges are accumulated on the *n*-type side and the negative charges are accumulated on the *p*-type side in the depletion region. In this case, the internal electric field is opposite to the external field and the charges in the depletion region are immovable. However, the case of a heterojunction consisting of two organic semiconductors is somewhat different. The organic semiconductor heterojunction (OHJ) (Yan et al. 2010) (i.e., *n*-type and *p*-type organic semiconductors) is a promising electronic system for charge recombination in OLEDs, for charge

separation in organic photovoltaic cells (OPVs), and for charge accumulation in organic field-effect transistors (OFETs) owing to the energy mismatch between the frontier orbitals of the two organic semiconductors. Because the dielectric constant of the organic semiconductor is usually low and the non-covalent electronic interactions between organic semiconductors are weak compared to inorganic semiconductors, two types of anisotype heterojunctions may be formed: accumulation and depletion heterojunctions. The depletion heterojunction in organic semiconductors is similar to that in inorganic semiconductors. However, the accumulation heterojunction is completely different. In this case, the positive and negative charges are accumulated on the *p*-type and *n*-type sides, respectively, of the organic semiconductor to form the space charge region. We call this phenomenon a heterojunction effect, which has been well demonstrated in OFETs and OPVs exhibiting highly efficient device performance. The direction of the built-in voltage is from the *p*-type region to the *n*-type region. More importantly, the accumulated charges within the charge region are movable. The accumulation of high-density free charge carriers results in a high conductivity along the junction direction.

To be able to effectively achieve and manipulate these processes in OHJs, we select C_{60} /pentacene as an example to elucidate these processes. On the basis of the theory of thermal emission of electrons, the electron transfer from pentacene to C_{60} can be achieved since pentacene has a higher Fermi level than C_{60} in the flat band state (Fig. 20 left). Also noted is that this charge transfer in turn contributes to the interfacial energy level equilibrium. Benefiting from the charge redistribution, the electrons and holes can be accumulated on *n*-type C_{60} and *p*-type pentacene, respectively, in the vicinity of the C_{60} /pentacene interface (Fig. 20 middle). Therefore, high-density free electrons and holes are provided at the C_{60} /pentacene junction (Fig. 20 right), i.e., charge generation occurs. These generated charge carriers can move away from the interface in opposite directions under an external electric field. This process is beneficial to reducing the voltage drop across the CGL and hence the reduction of the overall driving voltage during device operation. Obviously, the relative energy level of both semiconductor components is very important for the CGL construction, which directly determines the charge generation.

As demonstrated above, using OHJs as CGLs can enhance the power efficiency of the fabricated tandem OLEDs. Accordingly, the tandem white OLEDs based on OHJs as CGLs were fabricated (Chen et al. 2011, 2012a) and found that the power

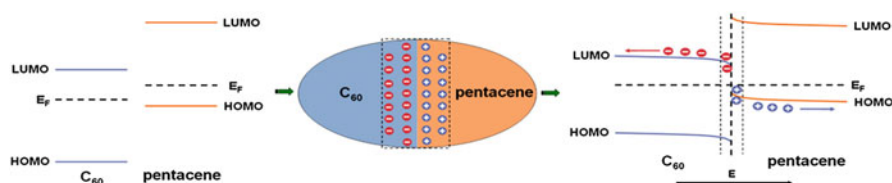


Fig. 20 Proposed working principle of the C_{60} /pentacene OHJ CGL. *EF* Fermi energy level, *LUMO* lowest unoccupied molecular orbital, *HOMO* highest occupied molecular orbital (Reproduced from Chen et al. (2012c))

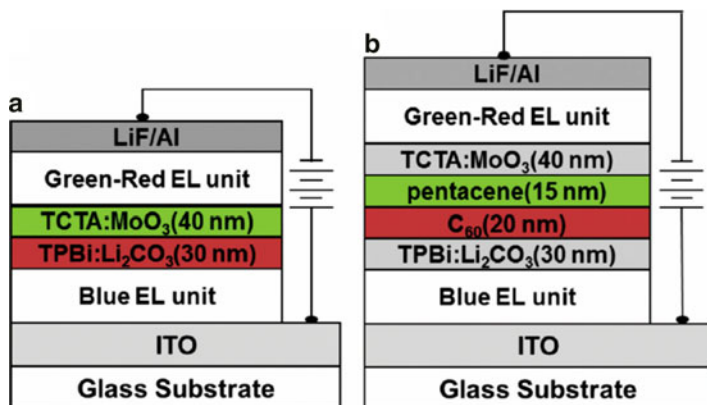
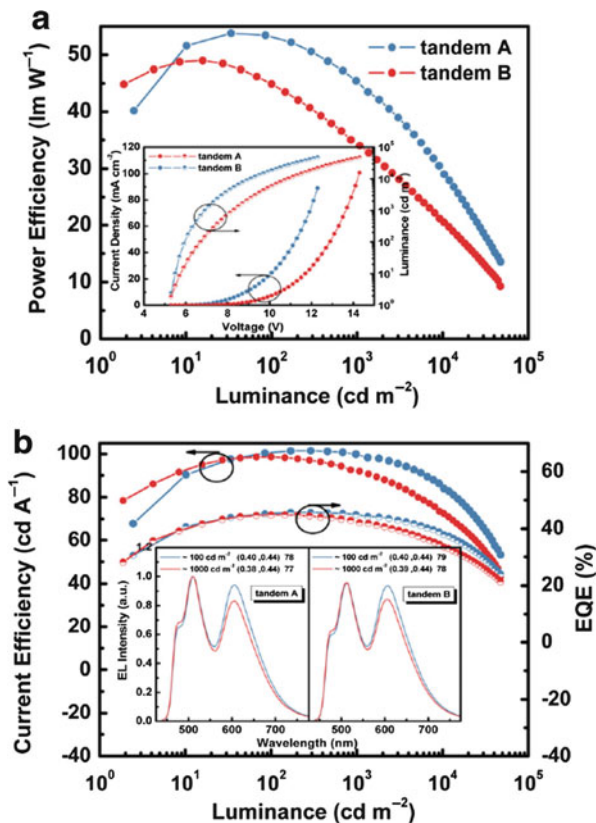


Fig. 21 Schematic diagrams of the tandem white OLEDs based on different CGLs. (a) TPBi:Li₂CO₃/TCTA:MoO₃ as the CGL. (b) C₆₀/pentacene organic heterojunction as the CGL (Reproduced from Chen et al. (2011))

efficiency of the device was indeed enhanced and the efficiency roll-off was greatly reduced compared to the conventional CGL-based tandem WOLEDs. Figure 21 shows the device structures of the tandem WOLEDs based on conventional TPBi:Li₂CO₃/TCTA:MoO₃ CGLs (Fig. 21a) and C₆₀/pentacene OHJ CGL (Fig. 21b). The white emission is realized by stacking a green–red phosphorescence unit and a blue phosphorescence unit via CGLs.

Figure 22 shows the EL characteristics of both tandem white OLEDs based on TPBi:Li₂CO₃/TCTA:MoO₃ (tandem A) and C₆₀/pentacene OHJ (tandem B) as CGLs. As shown in the inset of Fig. 22a, tandem A has a turn-on voltage of 5.5 V and operational voltages of 6.7 V at 100 cd/m² and 8.3 V at 1,000 cd/m². However, the turn-on voltage is reduced to 5.1 V and operational voltage of 6.0 V at 100 cd/m² and 6.9 V at 1,000 cd/m² as using C₆₀/pentacene organic heterojunction as the CGL (tandem B). This indicates that the C₆₀/pentacene organic heterojunction should have further better electrical properties as the internal floating connecting electrode than TPBi:Li₂CO₃/TCTA:MoO₃. As a result, the maximum current efficiency, external quantum efficiency, and power efficiency reach 101.5 cd/A, 45.7 %, and 53.8 lm/W, respectively (Fig. 22a and b). As we see, although the TPBi:Li₂CO₃/TCTA:MoO₃ as CGL indeed also leads to high current efficiency and external quantum efficiency, which reach 98.6 cd/A and 44.9 % (Fig. 22b), respectively, indicating an effective CGL as previously reported, but the maximum power efficiency of 48.7 lm/W is low (Fig. 22a) due to the high operational voltage. More importantly, the utilization of C₆₀/pentacene organic heterojunction as the CGL significantly improves the efficiency roll-off of fabricated tandem white OLEDs. The efficiency is slightly reduced to 101 cd/A, 45.5 %, and 53 lm/W and 99.9 cd/A, 45 %, and 45 lm/W at 100 cd/m² and 1,000 cd/m², respectively. The power efficiency roll-off values are only 1.5 % at 100 cd/m² and 16 % at 1,000 cd/m². Comparatively, the serious efficiency roll-off is given in TPBi:Li₂CO₃/TCTA:MoO₃-based tandem

Fig. 22 (a) Power efficiency versus luminance characteristics. *Inset*: J–V–L characteristics. (b) Current and external quantum efficiency versus luminance characteristics. *Inset*: EL spectra detected at the luminance of 100 cd/m^2 and 1,000 cd/m^2 (Reproduced from Chen et al. (2011))



white OLEDs, where the power efficiency is decreased to 34 lm/W at 1,000 cd/m^2 with the roll-off value of 24 %. As previously demonstrated (Chen et al. 2012c, d), the effective charge transfer from pentacene to C_{60} will result in the accumulation of the holes on the p-side and the electrons on the n-side in the C_{60} /pentacene CGL. The formation of the accumulation-type space charge region not only supplies the amounts of charges used to recombine but also leads to the formation of a high conductance region, thus greatly reducing the voltage drop in connecting layer. The generation of amounts of charges in C_{60} /pentacene CGL should also be favorable to improving charge balance, thus efficiency roll-off. The inset in Fig. 22b shows the EL spectra of two tandem white OLEDs at 100 and 1,000 cd/m^2 . The similar EL spectra of tandem A and B devices indicate that inserting the C_{60} /pentacene organic heterojunction does not lead to additional optical interference and microcavity effect.

Nevertheless, the challenge posed in today's OHJ-based tandem white OLEDs is the fabrication of highly effective charge generation systems, that is, *p*-type and *n*-type organic semiconductors with matched energy levels and charge carrier mobility, which are always related to the operational voltage and the power efficiency. Fortunately, these advancements have shown that the bottleneck has gradually been overcome by reasonably controlling the energy levels of *p*-type and *n*-type

organic semiconductors and by introducing a proper n-doped electron-transporting layer and a p-doped hole-transporting layer to enhance the electron and hole injection, fully showing the effectiveness of OHJs as CGLs in realizing high power efficiency tandem OLEDs.

Future Outlook

OLED technology is presently developing as a promising option for large-area lighting applications with the rapid advances in efficiency, color stability, and lifetime, which are basically satisfied to the requirement of conventional lighting applications, and with various interesting additional complementing features. Particularly, its power efficiency and lifetime reached the level of fluorescence tubes by smart design of device structures. Obviously, the progress in white OLEDs is in part a result of intensive collaborations between research in academia, institutes, OLED-producing companies, and supplying companies, e.g., of chemical components and fabrication tools. It should be predicted that the first applications of OLED lighting are expected in the specialized markets, emphasizing novel design opportunities. Later developments will make it possible to realize general lighting, such as family lighting, office lighting, automotive lighting, etc. Transparent OLEDs are expected in the coming years, whereas flexible OLEDs are expected soon only if flexible encapsulation technology can be resolved.

However, in order to reach the aims, there are many challenges yet that remain to be addressed. Continuous progress in improving the power efficiency and lifetime at illumination luminance and reducing the manufacturing cost are required. Obviously, the development of advanced device structures and simple processing technology and the further improvement of high-performance, low-cost organic EL materials will meet these challenges, finally making the OLED lighting product cost competitive so that OLED panels can really become a premier low-cost lighting technology.

References

- Adachi C, Baldo MA, Forrest SR, Lamansky S, Thompson ME, Kwong RC (2001) High efficiency red electrophosphorescence devices. *Appl Phys Lett* 78:1622
- Albrecht U, Bassler H (1995) Efficiency of charge recombination in organic light-emitting diodes. *Chem Phys* 199:207
- Baldo MA, O'Brien DF, You Y, Shoustikov A, Sibley S, Thompson ME, Forrest SR (1998) Highly efficient phosphorescent emission from organic electroluminescent devices. *Nature* 395:151
- Baldo MA, Adachi C, Forrest SR (2000) Transient analysis of organic electrophosphorescence. II. Transient analysis of triplet-triplet annihilation. *Phys Rev B* 62:10967
- Birks JB (1970) *Photophysics of organic molecules*. Wiley, New York, p 372
- Birnstock J, Wellmann P, Werner A, Romainczyk T, Hofmann M, Limmert M, Grüßing A, Blochwitz-Nimoth J (2005) White OLED structures using molecularly doped charge transport layers. In: *Eurodisplay, proceedings, Edinburgh*, p 192

- Brutting W, Frischeisen J, Schmidt TD, Scholz BJ, Mayr C (2013) Device efficiency of organic light-emitting diodes: progress by improved light outcoupling. *Phys Status Solidi A* 1:44
- Chang CH, Chen CC, Wu CC, Chang SY, Hung JY, Chi Y (2010) High-color-rendering pure-white phosphorescent organic light-emitting devices employing only two complementary colors. *Org Electron* 11:266
- Chang YL, Yin S, Wang ZB, Helander MG, Qiu J, Chai L, Liu ZW, Schles GD, Lu ZH (2013) Highly efficient warm white organic light-emitting diodes by triplet exciton conversion. *Adv Funct Mater* 23:705
- Chen S, Kwok HS (2011) Top-emitting white organic light-emitting diodes with a color conversion cap layer. *Org Electron* 12:677
- Chen YH, Ma DG (2012) Organic semiconductor heterojunctions as charge generation layers and their application in tandem organic light-emitting diodes for high power efficiency. *J Mater Chem* 22:18718
- Chen YH, Chen JS, Ma DG, Yan DH, Wang LX (2011) Tandem white phosphorescent organic light-emitting diodes based on interface-modified C60/pentacene organic heterojunction as charge generation layer. *Appl Phys Lett* 99:103304
- Chen YH, Tian HK, Chen JS, Geng YH, Yan DH, Wang LX, Ma DG (2012a) Highly efficient tandem white organic light-emitting diodes based upon C60/NaT4 organic heterojunction as charge generation layer. *J Mater Chem* 22:8492
- Chen YH, Zhao FC, Zhao YB, Chen JS, Ma DG (2012b) Ultra-simple hybrid white organic light-emitting diodes with high efficiency and CRI trade-off: fabrication and emission-mechanism analysis. *Org Electron* 13:2807
- Chen YH, Wang Q, Chen JS, Ma DG, Yan DH, Wang LX (2012c) Organic semiconductor heterojunction as charge generation layer in tandem organic light-emitting diodes for high power efficiency. *Org Electron* 13:1121
- Chen YH, Tian HK, Geng YH, Chen JS, Ma DG, Yan DH, Wang LX (2012d) Organic heterojunctions as a charge generation layer in tandem organic light-emitting diodes: the effect of interfacial energy level and charge carrier mobility. *J Mater Chem* 21:15332
- Cho SH, Oh JR, Park HK, Kim HK, Lee YH, Lee JG, Do YR (2010) Highly efficient phosphor-converted white organic light-emitting diodes with moderate microcavity and light-recycling filters. *Opt Express* 18:1099
- D'Andrade BW, Forrest SR (2004) White organic light-emitting devices for solid-state lighting. *Adv Mater* 16:1585
- D'Andrade BW, Thompson ME, Forrest SR (2002) Controlling exciton diffusion in multilayer white phosphorescent organic light emitting devices. *Adv Mater* 14:147
- D'Andrade BW, Holmes RJ, Forrest SR (2004) Efficient organic electrophosphorescent white organic light emitting device with a triple doped emissive layer. *Adv Mater* 16:624
- D'Andrade B, Esler J, Lin C, Weaver M, Brown J (2008) Extremely long lived white phosphorescent organic light emitting device with minimum organic materials. *SID 08 Digest*, p 940
- Duan L, Zhang DQ, Wu KW, Huang XQ, Wang LD, Qiu Y (2011) Controlling the recombination zone of white organic light-emitting diodes with extremely long lifetimes. *Adv Funct Mater* 21:3540
- Eom SH, Zheng Y, Wrzesniewski E, Lee J, Chopra N, So F, Xue JG (2009) White phosphorescent organic light-emitting devices with dual triple-doped emissive layers. *Appl Phys Lett* 94:153303
- Fukagawa H, Shimizu T, Ohbe N, Tokito S, Tokumaru K, Fujikake H (2012) Anthracene derivatives as efficient emitting host for blue organic light-emitting diodes utilizing triplet-triplet annihilation. *Org Electron* 13:1197
- Giebink NC, Forrest SR (2008) Quantum efficiency roll-off at high brightness in fluorescent and phosphorescent organic light emitting diodes. *Phys Rev B* 77:235215
- Guo FW, Ma DG (2005) White organic light-emitting diodes based on tandem structures. *Appl Phys Lett* 87:173510
- Hatwar TK, Spindler JP, Ricks ML, Young RH, Cosimbescu L, Begley W, Slyke SV (2004) White OLED structures optimized for RGB and RGBW formats. In: 24th international display research conference ASIA display, Daegu, p 816

- He G, Pfeiffer M, Leo K, Hofmann M, Birnstock J, Pudzich R, Salbeck J (2004) High-efficiency and low-voltage p-i-n electrophosphorescent organic light-emitting diodes with double-emission layers. *Appl Phys Lett* 85:3911
- Jou JH, Shen SM, Chen CC, Chung YC, Wang CJ, Hsu MF, Wang WB, Wu MH, Yang CJ, Liu CP (2008) High-efficiency fluorescent white organic light-emitting diodes using double hole-transporting-layers. *Proc SPIE* 6999:69992S
- Kalinowski J, Stampor W, Mezyk J, Cocchi M, Virgili MD, Fattori V, Di Marco P (2002) Quenching effects in organic electrophosphorescence. *Phys Rev B* 66:235321
- Kalinowski J, Stampor W, Szymkowski J, Virgili D, Cocchi M, Fattori V, Sabatini C (2006) Coexistence of dissociation and annihilation of excitons on charge carriers in organic phosphorescent emitters. *Phys Rev B* 74:085316
- Kido J, Kimura M, Nagai K (1995) Multilayer white light-emitting organic electroluminescent device. *Science* 267:1332
- Lee J, Lee JI, Lee JY, Chu HY (2009) Enhanced efficiency and reduced roll-off in blue and white phosphorescent organic light-emitting diodes with a mixed host structure. *Appl Phys Lett* 94:193305
- Lee SY, Yasuda T, Nomura H, Adachi C (2012) High efficiency organic light-emitting diodes utilizing thermally activated delayed fluorescence from triazine-based donor-acceptor hybrid molecules. *Appl Phys Lett* 101:093306
- Li C, Ichikawa M, Wei B, Taniguchi Y, Kimura H, Kawaguchi K, Sakurai K (2007) A highly color-stability white organic light-emitting diode by color conversion within hole injection layer. *Opt Express* 15:608
- Liao LS, Klubek KP, Tang CW (2004) High-efficiency tandem organic light-emitting diodes. *Appl Phys Lett* 84:167
- Ma YG, Zhang HY, Shen JC, Che CM (1998) Electroluminescence from triplet metal – ligand charge-transfer excited state of transition metal complexes. *Synth Met* 94:245
- Matsumoto T, Nakada T, Endo J, Mori K, Kavamura N, Yokoi A, Kido J (2003) Multiphoton organic EL device having charge generation layer. *SID 03 Digest*, p 979
- Moon J, Joo M, Lee YK, Lee JY, Park M, Choi J, Ham Y, Ahn Y, Kim JD, Lee J, Kim JH, You J, Jeong K, Kim JS, Son S (2013) 80 lm/W White OLED for solid state lighting. *SID 2013 Digest*, p 842
- Murano S, Kucur E, He GF, Blochwitz-Nimoth J, Hatwar TK, Spindler J, Slyke SV (2009) White fluorescent PIN OLED with high efficiency and lifetime for display applications. *SID Intl Symp Dig Tech Papers* 39:417
- Nishimura K, Kawamura M, Jinde Y, Yabunouchi N, Yamamoto H, Arakane T, Iwakuma T, Funahashi M, Fukuoka K, Hosokawa C (2008) The improvement of white OLED's performance. *SID Intl Symp Dig Tech Papers* 39:1971
- Pfeiffer M, Forrest SR, Leo K, Thomson ME (2002) Electrophosphorescent p-i-n organic light-emitting devices for very-high-efficiency flat-panel displays. *Adv Mater* 14:1633
- Reineke S, Walzer K, Leo K (2007) Triplet-exciton quenching in organic phosphorescent light-emitting diodes with Ir-based emitters. *Phys Rev B* 75:125328
- Reineke S, Lindner F, Schwartz G, Seidler N, Walzer K, Lussem B, Leo K (2009) White organic light-emitting diodes with fluorescent tube efficiency. *Nature* 459:234
- Reineke S, Thomschke M, Lu'ssem B, Leo K (2013) White organic light-emitting diodes: status and perspective. *Rev Mod Phys* 85:1245
- Sasabe H, Kido J (2011) Multifunctional materials in high-performance OLEDs: challenges for solid-state lighting. *Chem Mater* 23(3):621–630
- Sasabe H, Takamatsu JI, Motoyama T, Watanabe S, Wagenblast C, Langer N, Molt O, Fuchs E, Lennartz C, Kido J (2010) High efficiency blue and white organic light-emitting devices incorporating a blue iridium carbene complex. *Adv Mater* 22:5003
- Schwab T, Thomschke M, Hofmann S, Furno M, Leo K, Luessem B (2011) Efficiency enhancement of top-emitting organic light-emitting diodes using conversion dyes. *J Appl Phys* 110:083118

- Schwartz G, Fehse K, Pfeiffer M, Walzer K, Leo K (2006) Highly efficient white organic light emitting diodes comprising an interlayer to separate fluorescent and phosphorescent regions. *Appl Phys Lett* 89:083509
- Schwartz G, Pfeiffer M, Reineke S, Walzer K, Leo K (2007) Harvesting triplet excitons from fluorescent blue emitters in white organic light-emitting diodes. *Adv Mater* 19:3672
- Schwartz G, Reineke S, Rosenow TC, Walzer K, Leo K (2009) Triplet harvesting in hybrid white organic light-emitting diodes. *Adv Funct Mater* 19:1319
- Segal M, Baldo MA, Holmes RJ, Forrest SR, Soos ZG (2003) Excitonic singlet-triplet ratios in molecular and polymeric organic materials. *Phys Rev B* 68:075211
- Sharma BL, Purohit RK (1974) *Semiconductor heterojunctions*. Pergamon Press, Oxford
- Spindler JP, Hatwar TK (2009) Fluorescent-based tandem white OLEDs designed for display and solid-state-lighting applications. *J SID* 17(10):861
- Su SJ, Gonmori E, Sasabe H, Kido J (2008) Highly efficient organic blue-and white-light-emitting devices having a carrier- and exciton-confining structure for reduced efficiency roll-off. *Adv Mater* 20:4189
- Sun Y, Forrest SR (2007) High-efficiency white organic light emitting devices with three separate phosphorescent emission layers. *Appl Phys Lett* 91:263503
- Sun YR, Forrest SR (2008) Multiple exciton generation regions in phosphorescent white organic light emitting devices. *Org Electron* 9:994
- Sun Y, Giebink N, Kanno H, Wa B, Thompson ME, Forrest SR (2006) Management of singlet and triplet excitons for efficient white organic light-emitting devices. *Nature* 440:908
- Sun N, Wang Q, Zhao YB, Chen YH, Yang DZ, Zhao FC, Chen JS, Ma DG (2014) High-performance hybrid white organic light-emitting devices without an interlayer between fluorescent and phosphorescent emissive regions. *Adv Mater* 26:1617–1621. doi:10.1002/adma.201304779
- Tang CW, Vanslyke SA (1987) Organic electroluminescence diodes. *Appl Phys Lett* 51:913
- Tyan YS, Rao Y, Wang JS, Kesel R, Cushman TR, Begley WJ (2008) Fluorescent white OLED devices with improved light extraction. *SID Intl Symp Dig Tech Papers* 39:933
- Tyan YS, Rao YQ, Ren XF, Kesel R, Cushman TR, Begley WJ, Bhandari N (2009) Tandem hybrid white OLED devices with improved light extraction. *SID 09 Digest*, p 895
- Uoyama H, Goushi K, Shizu K, Nomura H, Adachi C (2012) Highly efficient organic light-emitting diodes from delayed fluorescence. *Nature* 492:234
- Valeur B (2002) *Molecular fluorescence*. Wiley, Weinheim, p 41
- Wang Q, Ma DG (2010) Management of charges and excitons for high-performance white organic light-emitting diodes. *Chem Soc Rev* 39:2387
- Wang Q, Ding JQ, Ma DG, Cheng YX, Wang LX, Wang FS (2009a) Manipulating charges and excitons within a singlet-host system to accomplish efficiency/CRI/color-stability trade-off for high performance OWLEDs. *Adv Mater* 21:2397
- Wang Q, Ding JQ, Ma DG, Cheng YX, Wang LX, Jing XB, Wang FS (2009b) Harvesting excitons via two parallel channels for efficient WOLED with nearly 100 % internal quantum efficiency: fabrication and mechanism analysis. *Adv Funct Mater* 19:84
- Wang Q, Ding JQ, Zhang ZQ, Ma DG, Cheng YX, Wang LX, Wang FS (2009c) A high-performance tandem white organic light-emitting diode combining highly effective white-units and their interconnection layer. *J Appl Phys* 105:076101
- Wang Q, Ho CL, Zhao YB, Ma DG, Wong WY, Wang LX (2010) Reduced efficiency roll-off in highly efficient and color-stable hybrid WOLEDs: the influence of triplet transfer and charge-transport behavior on enhancing device performance. *Org Electron* 11:238
- Yan DH, Wang HB, Du BX (2010) *Introduction to organic semiconductor heterojunction*. Science Press, Beijing
- Yang Y, Peng T, Ye KQ, Wu Y, Liu Y, Wang Y (2011) High-efficiency and high-quality white organic light-emitting diode employing fluorescent emitters. *Org Electron* 12:29

- Ye J, Zheng CJ, Ou XM, Zhang XH, Fung MK, Lee CS (2012) Management of singlet and triplet excitons in a single emission layer: a simple approach for a high-efficiency fluorescence/phosphorescence hybrid white organic light-emitting device. *Adv Mater* 24:3410
- Yersin H (2004) Triplet emitters for OLED applications: mechanisms of exciton trapping and control of emission properties. *Top Curr Chem* 241:1
- You H, Dai YF, Zhang ZQ, Ma DG (2007) Improved performances of organic light-emitting diodes with metal oxide as anode buffer. *J Appl Phys* 101:026105
- Zhang ZQ, Wang Q, Dai YF, Liu YP, Wang LX, Ma SG (2009) High efficiency fluorescent white organic light-emitting diodes with red, green and blue separately monochromatic emission layers. *Org Electron* 10:491
- Zhang QS, Li J, Shizu K, Huang S, Hirata S, Miyazaki H, Adachi C (2012) Design of efficient thermally activated delayed fluorescence materials for pure blue organic light emitting diodes. *J Am Chem Soc* 134:14706
- Zhao YB, Zhu LP, Chen JS, Ma DG (2012a) Improving color stability of blue/orange complementary white OLEDs by using single-host double-emissive layer structure: comprehensive experimental investigation into the device working mechanism. *Org Electron* 13:1340
- Zhao FC, Zhang ZQ, Liu YP, Dai YF, Chen JS, Ma DG (2012b) A hybrid white organic light-emitting diode with stable color and reduced efficiency roll-off by using a bipolar charge carrier switch. *Org Electron* 13:1049
- Zhao FC, Sun N, Zhang HM, Chen JS, Ma DG (2012c) Hybrid white organic light-emitting diodes with a double light-emitting layer structure for high color-rendering index. *J Appl Phys* 112:084504
- Zheng CJ, Wang J, Ye J, Lo MF, Liu XK, Fung MK, Zhang XH, Lee CS (2013) Novel efficient blue fluorophors with small singlet-triplet splitting: hosts for highly efficient fluorescence and phosphorescence hybrid WOLEDs with simplified structure. *Adv Mater* 25:2205

OLED Optics

Wooram Youn, Sai-Wing Tsang, and Franky So

Contents

Introduction	364
Substrate Mode Extraction	367
Micro lens Array	367
External Scattering Layer	368
Sand Blasting	368
ITO/Organic Mode Extraction	369
Internal Scattering Layer	369
Low-Index Grid	371
High-Refractive-Index Substrate	372
Photonic Crystals	374
Corrugated Structures	376
Conclusion and Outlook	380
References	381

Abstract

Applications of organic light-emitting diodes (OLEDs) have been rapidly developed since the first demonstration of green OLEDs by Tang and Vanslyke (1987). The original structure of an OLED consisted of a layer of organic electron transport layer (ETL)/emitting layer (EML) using 8-hydroxyquinoline aluminum (Alq3) and a hole transport layer (HTL) using an aromatic diamine. In order to inject charges and out-couple the emitted light, the organic layers were

W. Youn (✉) • F. So (✉)

Department of Material Science and Engineering, University of Florida, Gainesville, FL, USA
e-mail: Fso@mse.ufl.edu

S.-W. Tsang

Department of Physics and Materials Science, City University of Hong Kong, Kowloon, Hong Kong

sandwiched by a transparent indium-tin-oxide (ITO) anode and a reflecting metal cathode. Such an OLED can operate at high brightnesses, which can meet the requirement for display and lighting applications. Unfortunately, the poor carrier injection efficiency from electrodes led to the high operating voltage even for standard luminance of $1,000 \text{ cd/m}^2$ for display applications, resulting in undesirable high power consumption and very short lifetime. In 1997, Hung et al. showed that a very thin 1–2 nm lithium fluoride (LiF) layer adjacent to the ETL and aluminum cathode which greatly improved the injection efficiency for OLEDs, resulting in a low operating voltage at high brightness (Hung et al. 1997). The improved injection efficiency is attributed to forming an ohmic contact at the metal/organic interface which facilitates carrier injection. A similar approach at the other side to reduce the injection barrier of holes between the anode and the HTL was demonstrated by simple ultraviolet ozone (UVO) treatment on ITO-coated glass substrates (Sugiyama et al. 2000; Lee et al. 2004). The UVO treatment increases the work function of ITO by removing carbon contaminants and creating a tin-deficient and oxygen-rich surface. Another well-adopted treatment to improve hole injection efficiency is to insert a high work-function conducting polymer poly(3,4-ethylenedioxythiophene):poly(4-styrenesulphonate) (PEDOT:PSS) layer between the anode and the HTL (Jonas and Schrader 1991; Carter et al. 1997; Elschner et al. 2000). Nevertheless, even with the improved carrier injection with those interfacial modifications, the low device efficiency still remained a challenge for commercialization.

Introduction

Applications of organic light-emitting diodes (OLEDs) have been rapidly developed since the first demonstration of green OLEDs by Tang and Vanslyke (1987). The original structure of an OLED consisted of a layer of organic electron transport layer (ETL)/emitting layer (EML) using 8-hydroxyquinoline aluminum (Alq₃) and a hole transport layer (HTL) using an aromatic diamine. In order to inject charges and out-couple the emitted light, the organic layers were sandwiched by a transparent indium-tin-oxide (ITO) anode and a reflecting metal cathode. Such an OLED can operate at high brightnesses, which can meet the requirement for display and lighting applications. Unfortunately, the poor carrier injection efficiency from electrodes led to the high operating voltage even for standard luminance of $1,000 \text{ cd/m}^2$ for display applications, resulting in undesirable high power consumption and very short lifetime. In 1997, Hung et al. showed that a very thin 1–2 nm lithium fluoride (LiF) layer adjacent to the ETL and aluminum cathode greatly improved the injection efficiency for OLEDs, resulting in a low operating voltage at high brightness (Hung et al. 1997). The improved injection efficiency is attributed to forming an ohmic contact at the metal/organic interface which facilitates carrier injection. A similar approach at the other side to reduce the injection barrier of holes between the anode and the HTL was demonstrated by simple ultraviolet ozone (UVO) treatment on

ITO-coated glass substrates (Sugiyama et al. 2000; Lee et al. 2004). The UVO treatment increases the work function of ITO by removing carbon contaminants and creating a tin-deficient and oxygen-rich surface. Another well-adopted treatment to improve hole injection efficiency is to insert a high work-function conducting polymer poly(3,4-ethylenedioxythiophene):poly(4-styrenesulphonate) (PEDOT:PSS) layer between the anode and the HTL (Jonas and Schrader 1991; Carter et al. 1997; Elschner et al. 2000). Nevertheless, even with the improved carrier injection with those interfacial modifications, the low device efficiency still remained a challenge for commercialization.

A key parameter to determine the light-emitting property in OLEDs is the fraction of radiative excitons formed by recombination of electrons and holes. Based on quantum spin statistics, the ratio of singlet and triplet excitons is 1:3. The luminescence from the singlet excitons is known as fluorescence, while the luminescence from the triplet excitons is known as phosphorescence. However, relaxation of triplet excitons to ground state is forbidden. As a result, the triplet excitons lead to non-radiative transition, which eventually generates unwanted heat (Chou et al. 2011). Therefore, the internal quantum efficiency (IQE) of an OLED does not exceed 25 %. In order to tackle the limitation of the radiative efficiency, Baldo et al. discovered that using phosphorescent organic emitters, both singlet and triplet excitons can contribute to light emission, leading to an IQE of 100 % (Baldo et al. 1998). Such forbidden radiative transitions from the triplets can be overcome due to strong spin-orbit coupling (intersystem crossing) in the presence of heavy metal components such as platinum and iridium in the emitting small molecules (Baldo et al. 1998, 1999a, b; O'Brien et al. 1999; Tsutsui et al. 1999; Adachi et al. 2000, 2001a, b; Lee et al. 2000; Lamansky et al. 2001). Along with high-efficiency phosphorescent emitters, the host-guest mixing layer was also suggested for efficient energy transfer from host to guest emitters. With the development of high-efficiency phosphorescent OLEDs, these devices can be a potential candidate for next-generation light sources for lighting applications (Sasabe et al. 2013; Kim et al. 2013a).

Even with an IQE of 100 %, only a small fraction of light can be out-coupled due to the mismatch in indices of refraction resulting in total internal reflection (TIR). Figure 1 shows that light propagation in OLEDs can be categorized into three different optical modes: air mode is the light that directly escapes from the device with an incident angle smaller than the critical angle between air and glass substrate, substrate mode is defined as the light that escapes from ITO/organic layer but is trapped inside the glass substrate, and ITO/organic mode is the light with an incident angle greater than the critical angle determined by the glass-ITO/organic interface. Since only air mode can be collected, the fraction of air mode eventually plays a key role in determining the device efficiency. Thus, the out-coupling efficiency of the device, defined by the ratio of the number of photons escaped to the number of photons generated, is a figure of merit determining the OLED efficiency. For a typical OLED, the maximum out-coupling efficiency is only ~25–30 %.

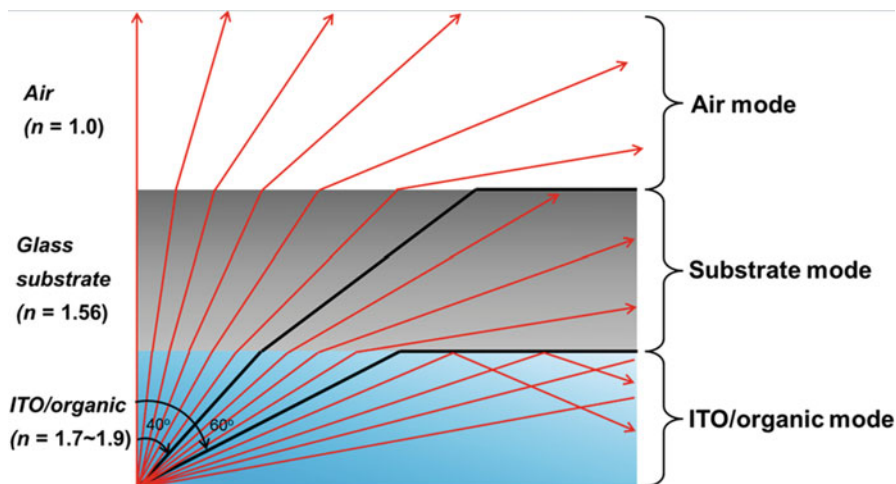


Fig. 1 Schematic illustration of total internal reflection in OLEDs and corresponding three optical modes determined by critical angle at ITO/organic-glass substrate interface of $\sim 60^\circ$ and glass substrate-air interface of $\sim 40^\circ$

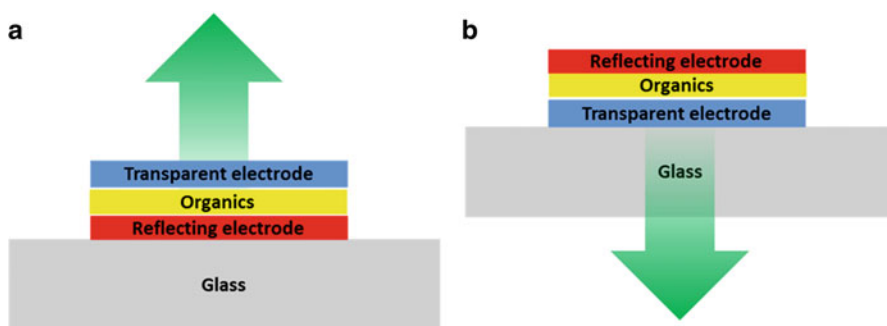


Fig. 2 Schematic OLED structure with light escape direction (a) top-emitting OLED (b) bottom-emitting OLED

Obviously, the out-coupling efficiency is highly influenced by the configuration of the OLED structure. Top-emitting OLEDs, as shown in Fig. 2a, have a higher out-coupling efficiency since generated light will only experience TIR at the interface between air and the transparent electrode. However physical damage on organic layers by sputtering ITO can be problematic to the resulting devices (Bulovic et al. 1996). Although using thin-transparent metal electrodes can be considered, having good transparency and at the same time matching the energy alignment of the metal electrodes with organic layers for charge injection are challenging. For bottom-emitting OLEDs, as shown in Fig. 2b, light with an incident angle greater than the critical angle will undergo TIR at the two interfaces, and only small amount

of light can escape from the glass substrate. Assuming that light emission is totally random oriented (isotropic) in EML of OLEDs and a 100 % reflectance at the back metal electrode, the maximum out-coupling efficiency is only ~30 %. Such poor out-coupling efficiency is the factor limiting device performance. The external quantum efficiency (EQE) of an OLED can be described below:

$$\eta_{\text{ext}} = \chi_{\text{out}} \eta_{\text{out}}$$

where η_{ext} is the external quantum efficiency, χ_{out} is the out-coupling efficiency, and η_{int} is the internal quantum efficiency. Since about 70 % is either trapped in the substrate, lost to the ITO/organic modes, or lost to the surface plasmon, significant efforts have been devoted to light extraction of OLEDs. The detailed characteristics and different extraction techniques will be discussed in the following sections.

Substrate Mode Extraction

As explained in the introduction, TIR at the air-substrate interface limits the light extraction efficiency. Thus, light propagating with an incident angle larger than the critical angle is reflected back at that interface and trapped inside the substrate. This loss is about 20–30 % of total light generated. In order to recover such loss, various approaches such as microlens, texturing, mesh, and sand blasting have been used to extract the substrate modes. These approaches are typically applied on the opposite side of the glass substrate where the OLED is fabricated. Thus, these light extraction schemes do not interfere with the OLED fabrication and can be used to enhance the device extraction efficiency.

Microlens Array

Möller et al. have introduced the use of microlens array on the back side of the glass substrate for OLED light extraction as shown in Fig. 3 (Moller and Forrest 2002). It is done by creating hemispherical lenses on the substrate. Most light rays from the OLEDs enter at angles larger than the critical angle. Thus, light that originally suffers from TIR can escape from the substrate, resulting in the increase of out-coupling efficiency by a factor of up to 1.5.

Based on the microlens approach, Eom et al. suggested the optimized contact angle of 85° of microlens with a diameter of 100 μm can be used to enhance the out-coupling efficiency by a factor of 1.7 (Eom et al. 2011). Further optimizations such as the lens array packing factor, refractive index, and thickness of the substrate have been demonstrated resulting in an increase of out-coupling efficiency up to twofold (Sun and Forrest 2006; Wei and Su 2004).

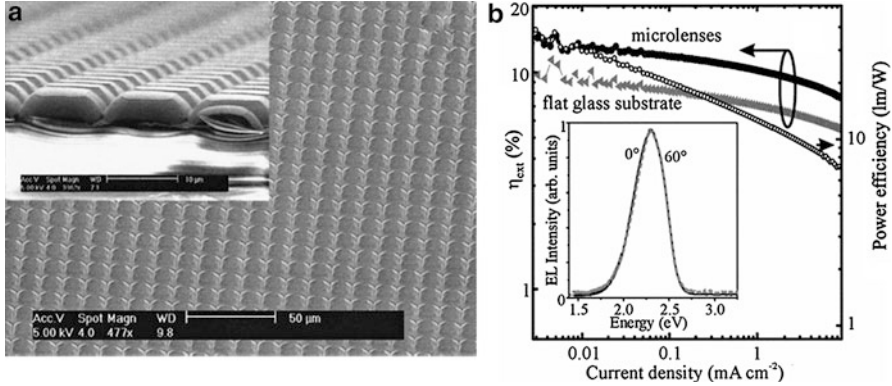


Fig. 3 (a) Scanning electron micrograph (SEM) image of a PDMS microlens array fabricated on the substrate and the detailed side view of the lenses (*inset*). (b) EQE and power efficiency vs. current density of flat glass substrate and microlens OLEDs. (*Inset*) Electroluminescence spectra of microlens OLED with two different incident angles with respect to the surface normal show the angle independency (Moller and Forrest 2002)

The use of microlens arrays has the advantage of simple fabrication process. Typical photolithography, soft-lithography, and wet-etching processes can be used to fabricate microlens arrays with various sizes and shapes (Moller and Forrest 2002; Eom et al. 2011; Sun and Forrest 2006; Wei and Su 2004).

External Scattering Layer

Another approach for substrate mode extraction is done by texturing the air-glass interface as shown in Fig. 4a. Cheng et al. demonstrated that the meshed-scattering layer can diffuse the trapped substrate mode, resulting in an enhancement of 46 % in out-coupling efficiency (Cheng et al. 2007). Figure 4b shows the scattering layer made of poly(dimethyl siloxane) (PDMS). Using molding and etching processes with porous anodic aluminum oxide templates, such porous scattering film can be formed. With refractive index-matching gel between the scattering PDMS layer and the glass substrate, light extraction of the substrate mode was demonstrated.

Sand Blasting

Having a similar working principle as the external scattering layer described earlier, instead of adding a functional layer through complicated fabrication process, sand-blasting techniques create the scattering center by simply roughening the glass substrate surface. As shown in optical microscope images in Fig. 5, the scattering effect is attributed by the randomly distributed particles with ~ 100 μm in sizes, which is similar to the size of sand particles to blast the substrate.

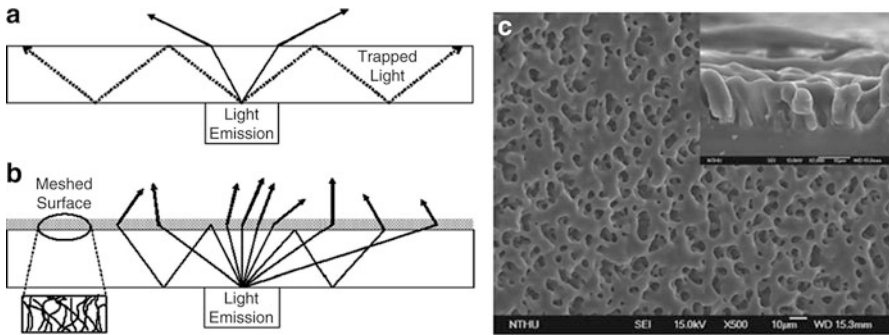


Fig. 4 (a) Schematic view of propagating light in the substrate mode in (a) conventional flat glass and (b) scattering layer deposited on the glass. (c) SEM image of scattering layer made by PDMS (*Inset*). The detailed view of the cross section of the layer (Cheng et al. 2007)

Chen et al. suggested that additional roughening on the edge of the glass substrate, as shown in Fig. 6, can increase light extraction by about 10 % in the forward direction (Chen and Kwok 2010). Further, sand blasting on both the edge and surface of the glass substrate resulted in about a 20 % enhancement in light extraction efficiency. Although the enhancement is not large compared to other techniques, the simple and low-cost processes are beneficial to OLED lighting applications (Zhou et al. 2011; Mulder et al. 2007).

ITO/Organic Mode Extraction

The refractive index mismatch between ITO/organic ($n = 1.7\text{--}1.9$) and glass ($n \sim 1.5$) substrate also results in TIR at that interface. Light suffering from TIR eventually is trapped in the ITO/organic layer, which is known as ITO/organic thin-film waveguide mode. Considering a conventional structure of an OLED consisting of a 200 nm thick-ITO and organic layers, the transverse electric (TE) mode is concentrated at the ITO layer, whereas the transverse magnetic (TM) mode which eventually is lost to the surface plasmon (SP) mode is concentrated at the organic/metal interface (Fuhrmann et al. 2003). Such waveguide modes are responsible for total 50–60 % loss in the emitted light (Fujita et al. 2005). Below are the various approaches that have been used to reduce the loss due to the waveguide modes.

Internal Scattering Layer

Internal scattering layers are normally placed between the substrate and the transparent electrode in order to recover the light trapped in the ITO/organic layers. In contrary to methods for extracting the substrate mode, methods to extract waveguide mode require additional processing on the OLED glass substrates which can affect the device fabrication because the ITO electrode and the organic layers are directly

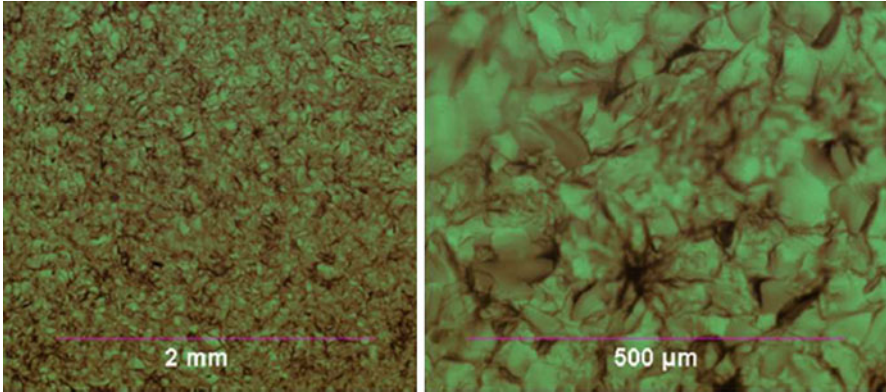


Fig. 5 Optical microscope images of the substrate surface after sand blasting with low (*left*) and high (*right*) resolution (Chen and Kwok 2010)

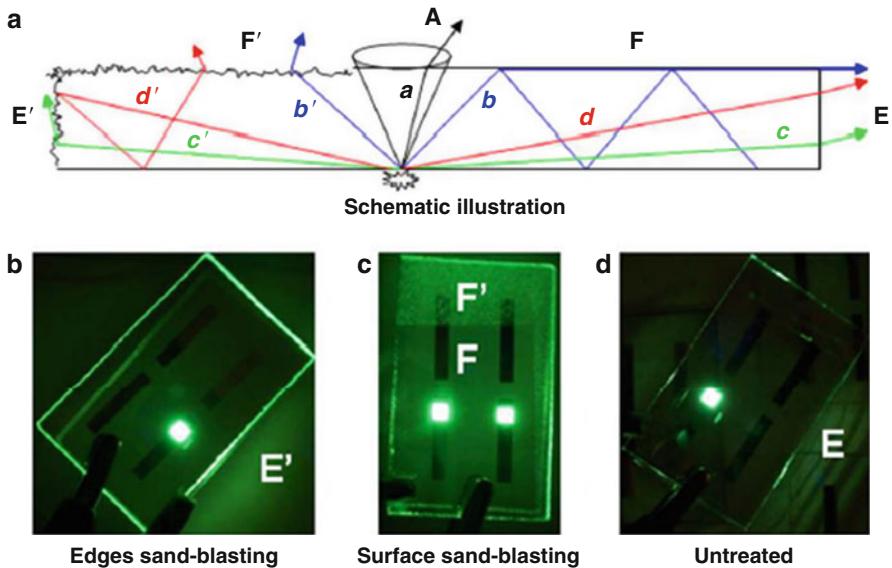
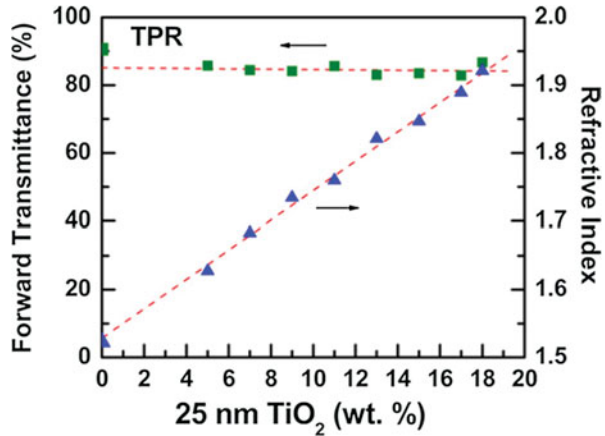


Fig. 6 (a) Schematic view of light with different incident angles in scattering glass substrate. (b, c, and d) Comparison of light propagation between sand blasted on edges, on surface, and planar substrates (Chen and Kwok 2010)

deposited on the internal scattering layer. Chang et al. demonstrated that using simple solution processing to fabricate nanocomposite thin films, the internal scattering layer can be obtained, resulting in a twofold enhancement in out-coupling efficiency (Chang et al. 2011). While the internal scattering layer can be rough which can affect the surface quality, it was demonstrated that smooth scattering layers can be made by

Fig. 7 The forward transmittances and the refractive indices of the nanocomposite films with different concentrations of 25 nm NPs at the incident wavelength of 475 nm (Chang et al. 2012)



spin-coating of a blend of 25 and 250 nm TiO₂ nanoparticles (NPs) in a photoresist (Chang et al. 2012). Since the 25 nm NPs are too small to interact with the light with visible wavelength, it mainly works for controlling the refractive index and flattening the resulting film surface, as shown in Figs. 7 and 8. On the other side, the 250 nm NPs are very strong scattering centers in the mixture. Thus, by balancing the ratio, they were able to fabricate such scattering layers with refractive index similar to the underlying ITO layer such that more light can enter the scatter layer. This approach results in a strong scattering effect which gives about two times enhancement in light out-coupling.

Low-Index Grid

Forrest group first introduced a way to extract waveguide mode by embedding low-index grids (LIG) between ITO and organic layers, as shown in Fig. 9 (Sun and Forrest 2008). Light that would be trapped as waveguide mode enters the low-index region and is redirected toward the substrate normal in direction, which eventually enhances the out-coupling efficiency. In addition, the LIG does not affect the light that originally escapes from the substrate. In another following work, they also demonstrated that the efficiency was enhanced up to 2.3 times using an ultralow-index grid ($n = 1.10\text{--}1.15$) (Slootsky and Forrest 2010). Figure 10 shows the SEM images of an ultralow-index grid made with porous SiO₂ by glancing angle vapor deposition (Xi et al. 2005, 2007). Using this technique, a low-refractive-index LIG is obtained because of the presence of small pores in range of ~ 10 nm inside of SiO₂ layer, as shown in Fig. 10. In contrast to other waveguide extraction technique such as photonic crystals, which will be discussed in the following section, it does not distort the emission spectrum of the OLED since the width of the grid is not in the range of the wavelength of the emitting light. However, the micro-size patterning process for the LIG is not simple and difficult to be used in large area manufacturing for lighting applications.

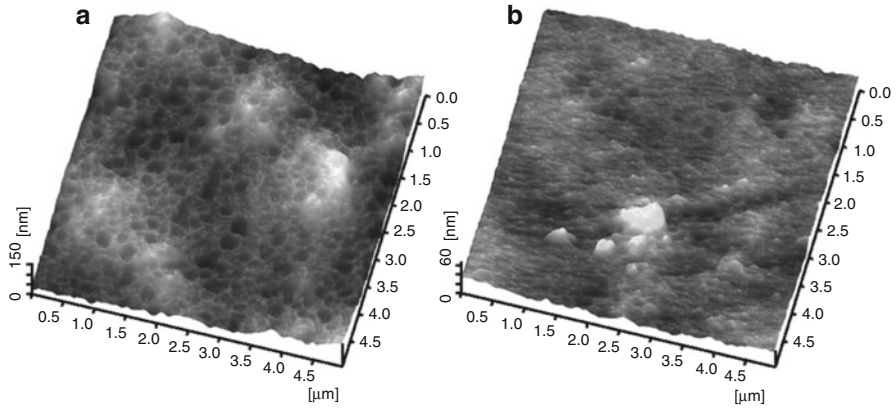


Fig. 8 The atomic force microscope (AFM) images of (a) single-sized NPs and (b) dual-sized NPs of the surface of the substrate: the 25 nm NPs show much better surface condition for device fabrication (Chang et al. 2012)

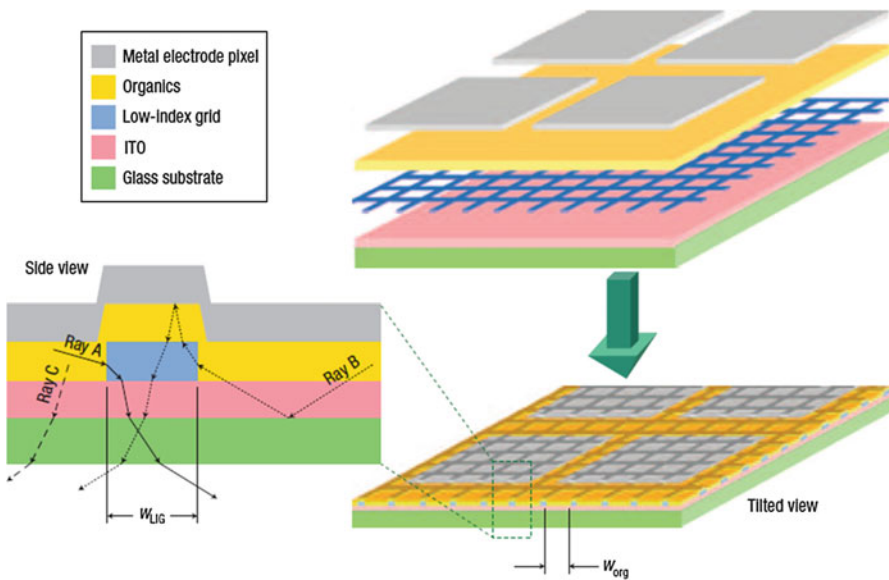
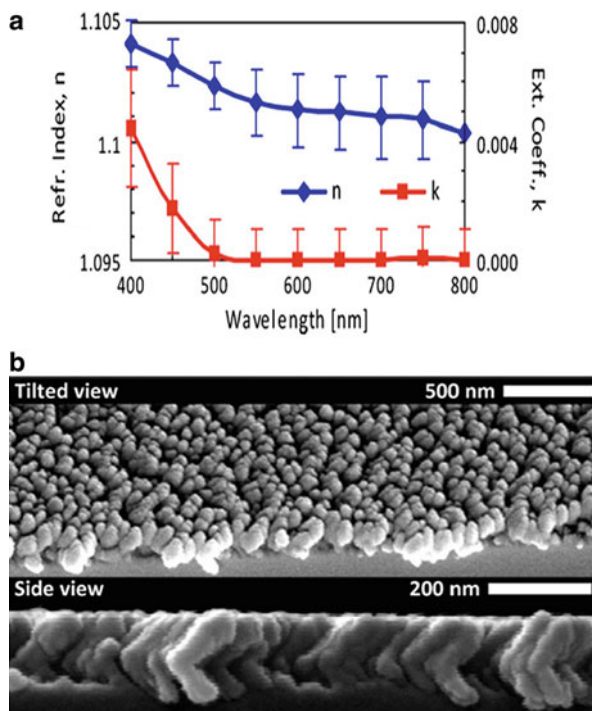


Fig. 9 Schematic diagram of the OLED with the embedded low-index grid (LIG) in the organic layers (Sun and Forrest 2008)

High-Refractive-Index Substrate

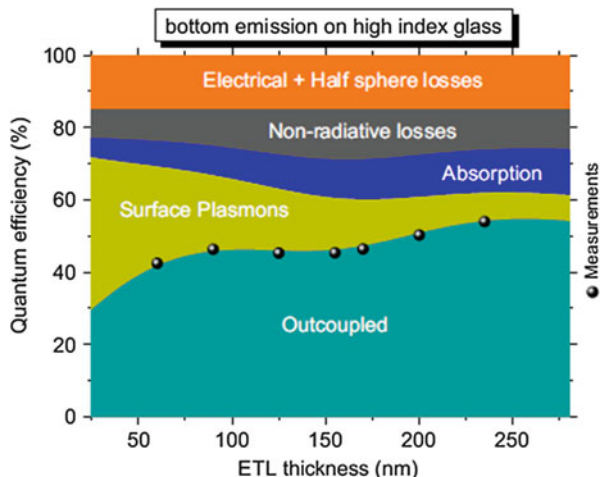
The relatively low refractive index of the conventional glass substrate compared to that of the ITO/organic layer is the boundary confining the waveguide mode. This problem can easily be solved by replacing the glass substrate with a high-refractive-index

Fig. 10 (a) Optical constants n and k measured by ellipsometry of porous SiO_2 with an ultralow-refractive index. (b) Tilted and side view of SEM images of the porous SiO_2 deposited on the substrate (Slootsky and Forrest 2010)



substrate. Leo group successfully demonstrated that waveguide modes can be completely eliminated in OLEDs fabricated on such a high-refractive-index glass substrate (Reineke et al. 2009; Meerheim et al. 2010). They also employed a macro extractor in order to evaluate the redistributed intensity of glass mode, showing a maximum EQE of 45 % using iridium(III)bis[2-methylidibenzo-(f, h)quinoxaline] (acetylacetonate) $[\text{Ir}(\text{MDQ})_2(\text{acac})]$ as a phosphorescent emitter, and the results are in good agreement with simulated results, as shown in Fig. 11 (Meerheim et al. 2010). However, this approach results in 20–30 % loss of the light generated by surface plasmon (SP) mode due to the interaction between the free electrons and the electromagnetic field at the organic/metal interface. By placing the emitting zone further away from the metal layer, the intensity of SP mode drops exponentially. Using this approach, the EQE increased further up to 55 % with a macro extractor. One of the key factors that relates to the SP mode is the absorption coefficient of the metal layer. Compared to Ag electrode, the coefficient of commonly used Al metal electrode is five times larger: $1.5 \times 10^5 \text{ cm}^{-1}$ and $7.6 \times 10^5 \text{ cm}^{-1}$, respectively. Reduced SP mode loss by employing the silver electrode was consequently confirmed by OLED fabricated on the high-index substrate by Mladenovski (Adawi et al. 2006). While high-index substrates are effective extracting thin-film waveguided mode, they are much more expensive than conventional glass substrates. However, the price of high-refractive-index substrates has dropped recently,

Fig. 11 EQE as function of ETL thickness by both simulation and experiment. It is worth noting that the intensity of SP mode decreases as locating the emitting zone away from the metal electrode (Meerheim et al. 2010)



and perhaps this can be a viable approach for OLED lighting extraction in the near future.

Photonic Crystals

Photonic crystals (PCs) are an efficient and expedient tool to extract waveguide modes by forming forbidden optical band gaps generated by a periodic variation of refractive index on a substrate. Light that would propagate as waveguide mode can be Bragg-scattered out toward the substrate normal as described by Bragg diffraction law:

$$k_{//} = k_o \sin \theta = k_{wg} \pm mk_{PC}, \tag{1}$$

where k_o denotes the wave vector in free space, $k_{//}$ the in-plane component of out-coupled wave, k_{wg} the in-plane waveguide component, k_{PC} the grating vector of the PC that is inversely proportional to the periodicity, θ is polar angle with respect to the surface normal, and m is an integer.

As shown in Eq. 1, the in-plane component of waveguide mode can be modified using the PC structure such that the diffracted waveguide mode can escape to the air mode. Figure 12 shows a schematic view of how the Bragg scattering is employed on a PC.

Considering the effective location in terms of out-coupling waveguide mode, it has been shown that the PC layer should be inserted between the glass substrate and the ITO electrode in an OLED to enhance the out-coupling efficiency (Adawi et al. 2006; Do et al. 2003, 2004; Riedel et al. 2010). A schematic diagram of an OLED incorporating a PC layer is shown in Fig. 13b. As calculated by the Bragg

Fig. 12 Schematic diagram of Bragg scattering based on PC layer (Adawi et al. 2006)

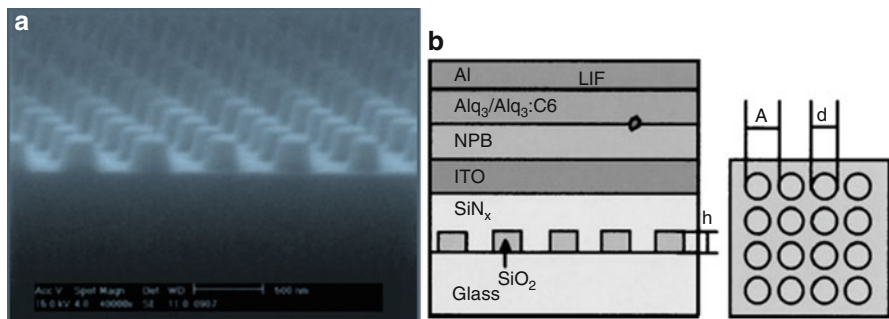
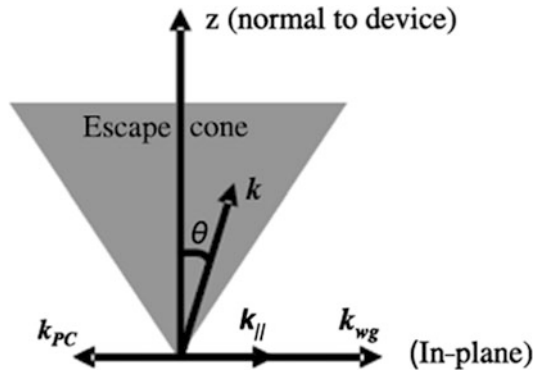
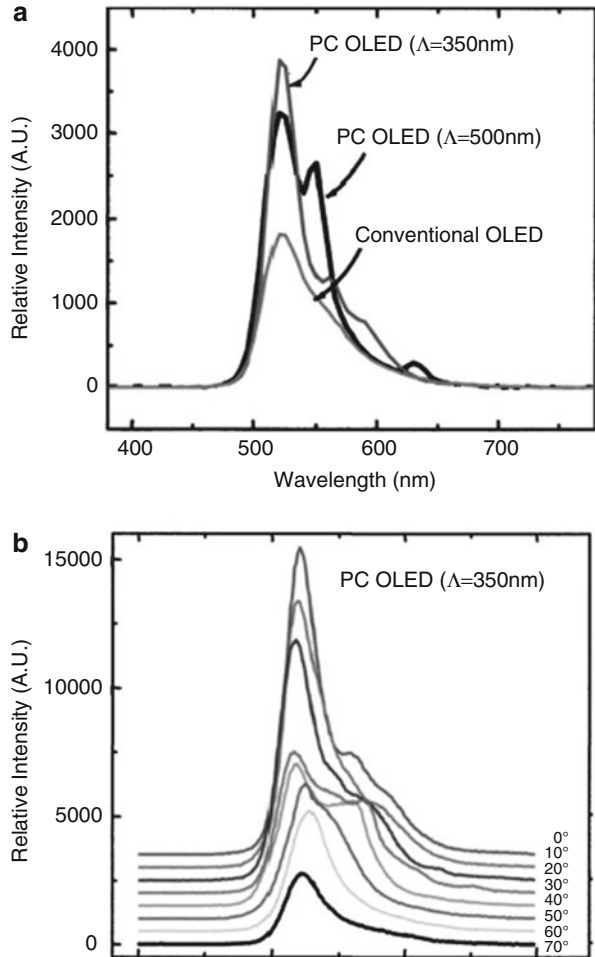


Fig. 13 (a) SEM image of PC layer fabricated on the substrate. (b) Schematic diagram of PC OLED (Do et al. 2004)

diffraction law as described above, the optimum periodicity is 0.3–1 μm depending on the wavelength of the emitting light in an OLED. In order to have diffraction-assisted out-coupling, the PC structure should consist of two dielectric materials that have a low and a high refractive index. For low-refractive-index materials, SiO_x ($n = 1.5$) and PEDOT:PSS ($n = 1.5$) are good examples, and for high-refractive-index materials, TiO_2 ($n = 2.5$), Ta_2O_5 ($n = 2.1$) and SiN_x ($n = 1.95$) are used as shown in Fig. 13a. Do et al. showed a two-dimensional (2D) PC made by two-beam laser interference method (Do et al. 2004). The out-coupling efficiency increased up to 50 % by utilizing a PC structure with a periodicity of 350 nm. As a very distinct characteristic of the PC, Fig. 14 shows that the enhancement due to PC strongly depends on direction and emission wavelength, indicating strong PC effect. Such a wavelength and directional selectivity depends on the periodicity and the modulation of the refractive index of the PC. As shown in Fig. 15, the poor emitting pixel uniformity due to the presence of the PC is problematic for lighting applications, and this approach is now not considered viable for lighting applications.

Fig. 14 (a)
Electroluminescence (EL) spectra of conventional flat OLED and PC OLED with 350 and 500 nm periodicity. **(b)** EL spectra of PC OLED as function of viewing angle (Do et al. 2004)



Corrugated Structures

Similar to photonic crystals, light extraction in OLEDs fabricated on a corrugated substrate is by Bragg diffraction. Diffraction occurs at the interfaces of metal/organic and ITO/glass due to corrugation: the difference in refractive index at corrugated interface of ITO ($n = 1.7\text{--}1.9$)/glass ($n \sim 1.5$) results in diffraction depending on the specific pattern of the corrugated structure. In corrugated OLEDs, diffraction also occurs at the metal/organic interface due to the reflection of the corrugated metal surface. As shown in Fig. 16, the corrugated substrate should result in corrugated organic layers as well as a corrugated metal electrode. It is important that the corrugated structure is maintained throughout the entire OLED stack up to the top metal electrode.

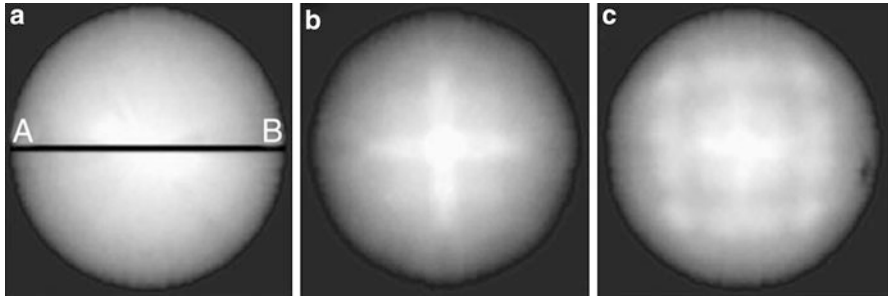


Fig. 15 Measured 2D far-field intensity profile of (a) conventional (b) PC OLED with 350 nm periodicity and (c) PC OLED with 500 nm periodicity (Do et al. 2004)

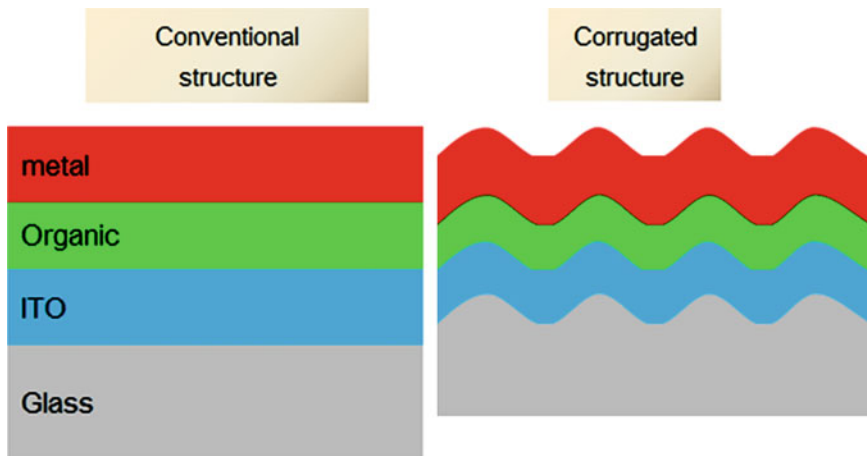


Fig. 16 Schematic diagram of conventional flat OLED vs. corrugated OLED

According to the work by Ishihara et al., OLEDs fabricated on a periodic corrugated structure showed an enhancement of 1.5 times in light extraction (Ishihara et al. 2007). It should be noted that in addition to the extraction of the waveguide mode, out-coupling of SP mode can be also realized. The electric field distribution calculated by transfer matrix formalism shown in Fig. 17b indicates that generated light is coupled to the waveguide and SP modes excited inside of organic/ITO layers. And by corrugated structure, the extraction of the waveguide and SP modes is confirmed by the EL spectrum as shown in Fig. 17c.

Figure 17 shows the optical modes, the electric field distribution, and the output spectrum of an OLED fabricated on a corrugated structure with a 300 nm periodicity. The enhanced, out-coupled optical mode into normal direction is in good agreement with calculation as shown in Fig. 17a, c. Obviously, similar to photonic crystals, the strong wavelength and angular dependence of the emitted light in OLEDs fabricated on a periodic corrugated substrate is not desirable for lighting. In 2010,

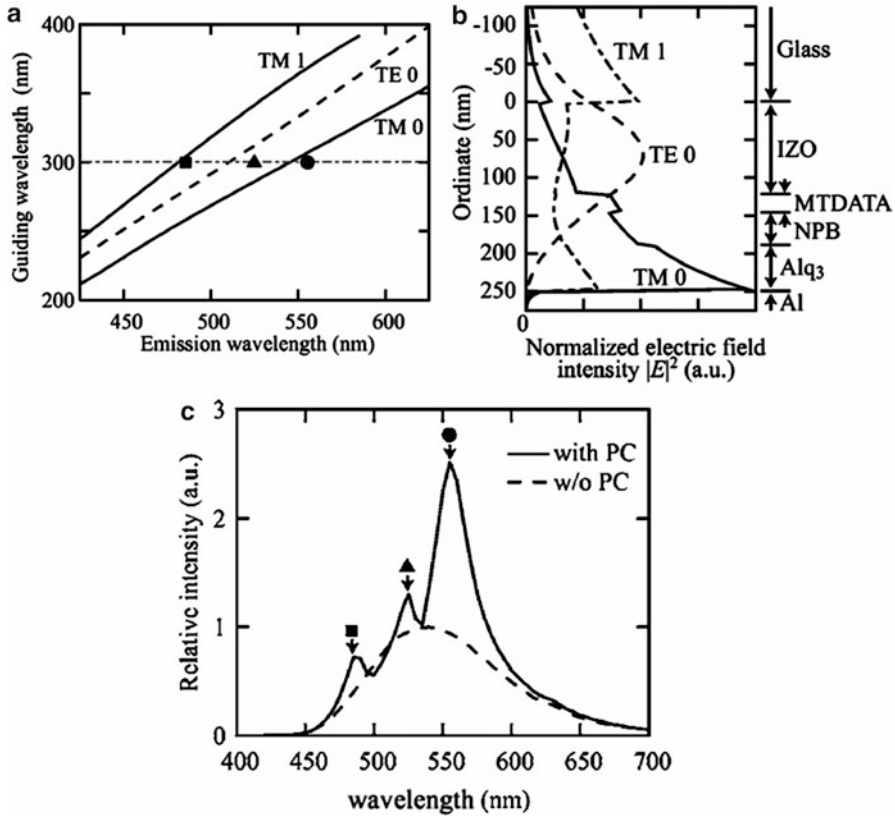


Fig. 17 (a) Calculated periodicity for corresponding wavelength of each optical mode for diffraction in normal direction. (b) Normalized electric field distribution of each optical mode within the device. (c) EL spectra of both conventional and corrugated OLED (Ishihara et al. 2007)

Koo et al. introduced a simple fabrication method to achieve a quasi-periodic structure such that light extraction is relatively uniform over all emission wavelength and emission angle (Koo et al. 2010). This quasi-periodic structure is formed from spontaneous buckling when a thin Al film is deposited on a pre-cured PDMS substrate. The driving force to form the buckling structure is the compressive stress induced by the differential thermal expansion coefficients of the PDMS and Al films. The buckling structure can be further controlled by introducing extra compressive stress through further deposition of Al, resulting in an increase in surface area and the depth of corrugation. Using the buckled PDMS as a mold, the structure was successfully transferred onto the glass substrate by refractive index-matching UV curable epoxy. Such a simple process is not limited by size and does not require complicated process like nano-imprint, laser interference, and electron-beam lithography. The best buckled OLEDs showed an enhancement of 120 % in current efficiency from fluorescent green OLEDs. As shown in periodicity distribution of

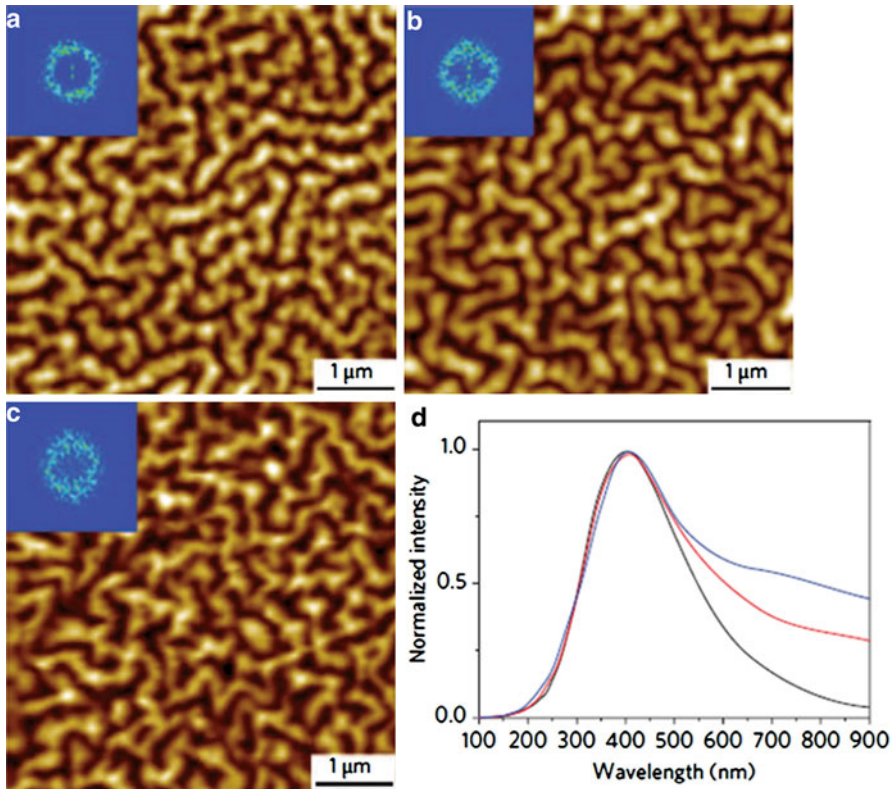


Fig. 18 AFM analysis of buckling patterns. (a) Buckled structure formed by a 10 nm thick aluminum layer (b),(c) Buckled structures formed by deposition of a 10 nm thick aluminum layer twice and three times, respectively. *Inset:* FFT patterns of each image. (d) Power spectra from FFTs as a function of wavelength for buckled patterns obtained with deposition of a 10 nm thick aluminum layer once (*black*), twice (*red*), and three times (*blue*) (Koo et al. 2010)

buckling structure from Fig. 18d, the distinctive characteristic of distributed periodicity from buckling structure indicates that the Bragg diffraction is independent of wavelength (400–700 nm) which is ideal for light extraction of white OLED with a broad emission spectrum. Also shown in the insets from Fig. 18a–c, the circular fast Fourier transform (FFT) pattern further confirmed that light extraction occurs in all azimuthal direction which is important for lighting applications.

More recently, Koo et al. also demonstrated another simple approach without using lithography process to fabricate the corrugated structure (Koo et al. 2012). The structure consists of hexagonal-close-packed (HCP) silica arrays embedded into a thin film of polystyrene (PS) on the glass substrate prepared by rapid convective deposition technique. As seen in the AFM image from Fig. 19a, the defects introduced in HCP array distort the long-range hexagonal symmetry, resulting in a ring pattern in the FFT image, similar to that with the corrugated structure shown

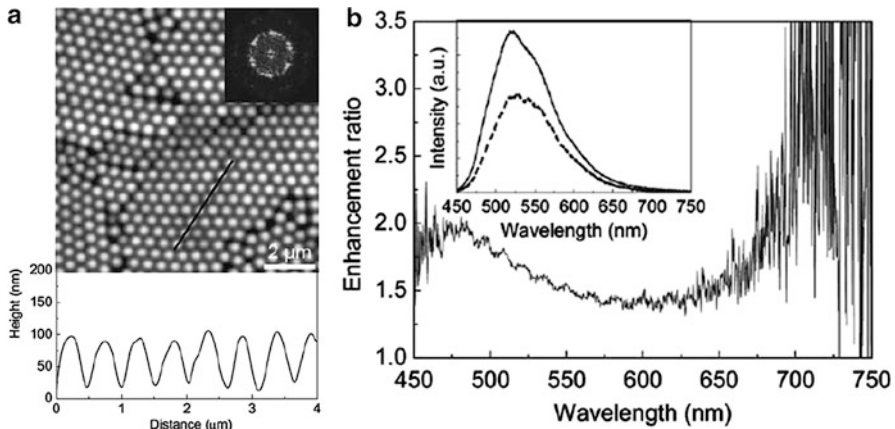


Fig. 19 (a) AFM image of “defective” HCP array of corrugated structure. *Inset*: ring FFT pattern indicates that all azimuthal direction can have diffraction. (b) The enhancement ratio as function of wavelength of OLED. *Inset*: the EL spectrum of both conventional and corrugated OLED (Koo et al. 2012)

previously formed by spontaneous buckling. Consequently, the extraction enhancement in current and power efficiency from corrugated OLEDs with a nominal 500 nm periodicity was increased by 70 % and 90 % with a broadband enhancement in all wavelengths as shown in Fig. 19b.

There is another technique demonstrating similar random corrugated structures with different process schemes (Choi et al. 2012). It is important that the fabrication process must be low cost and can be applied to large area processing for lighting applications.

Conclusion and Outlook

In conventional bottom-emitting OLEDs, typical out-coupling efficiency is less than 30 %, while the rest of the photons generated in OLEDs are confined in the devices as the substrate mode, ITO/organic mode, and SP mode. Despite of the success of different light extraction techniques to extract the substrate and ITO/organic waveguide modes as described previously, the extraction of the SP mode has not been so successful. Although the corrugated structure was demonstrated for the SP mode extraction, the out-coupling of the SP mode into air mode was still interrupted by inherent, strong absorption of metallic electrode. Recently, Kim’s group proposed to replace the metal electrode with a transparent oxide electrode to suppress the SP mode loss and simultaneously enhance the EQE. This approach might be one of the strong candidates for light extraction (Lee et al. 2012; Kim et al. 2013b). By combining the different light extraction techniques, an out-coupling efficiency of 60–70 % has been reached. However, these light extraction schemes must be low cost and scalable to large area manufacturing for general lighting applications.

Recently, 80 lm/W white lighting panels employing ITO/organic mode out-coupling technique have been commercialized by LG Chem, and 114 lm/W lighting panels have also been announced by Panasonic (<http://www.lgchem.com/global/green-energy/oled-lighting>; <http://www.osadirect.com/news/article/970/>), indicating that OLED lighting is indeed coming. Full commercialization of OLED lighting requires further enhancements in light extraction using novel schemes which are low cost and compatible with large area manufacturing.

References

- Adachi C et al (2000) High-efficiency organic electrophosphorescent devices with tris (2-phenylpyridine)iridium doped into electron-transporting materials. *Appl Phys Lett* 77 (6):904–906
- Adachi C et al (2001a) High-efficiency red electrophosphorescence devices. *Appl Phys Lett* 78 (11):1622–1624
- Adachi C, Kwong R, Forrest SR (2001b) Efficient electrophosphorescence using a doped ambipolar conductive molecular organic thin film. *Org Electron* 2(1):37–43
- Adawi AM et al (2006) Improving the light extraction efficiency of polymeric light emitting diodes using two-dimensional photonic crystals. *Org Electron* 7(4):222–228
- Baldo MA et al (1998) Highly efficient phosphorescent emission from organic electroluminescent devices. *Nature* 395(6698):151–154
- Baldo MA, Thompson ME, Forrest SR (1999a) Phosphorescent materials for application to organic light emitting devices. *Pure Appl Chem* 71(11):2095–2106
- Baldo MA et al (1999b) Very high-efficiency green organic light-emitting devices based on electrophosphorescence. *Appl Phys Lett* 75(1):4–6
- Bulovic V et al (1996) Transparent light-emitting devices. *Nature* 380(6569):29
- Carter SA et al (1997) Polymeric anodes for improved polymer light-emitting diode performance. *Appl Phys Lett* 70(16):2067–2069
- Chang HW et al (2011) Organic light-emitting devices integrated with internal scattering layers for enhancing optical out-coupling. *J Soc Inf Disp* 19(2):196–204
- Chang CH et al (2012) Fourfold power efficiency improvement in organic light-emitting devices using an embedded nanocomposite scattering layer. *Org Electron* 13(6):1073–1080
- Chen SM, Kwok HS (2010) Light extraction from organic light-emitting diodes for lighting applications by sand-blasting substrates. *Opt Express* 18(1):37–42
- Cheng YH et al (2007) Enhanced light outcoupling in a thin film by texturing meshed surfaces. *Appl Phys Lett* 90(9):091102
- Choi CS et al (2012) Improved light extraction efficiency in organic light emitting diodes with a perforated WO₃ hole injection layer fabricated by use of colloidal lithography. *Opt Express* 20 (6):A309–A318
- Chou PT et al (2011) Harvesting luminescence via harnessing the photophysical properties of transition metal complexes. *Coord Chem Rev* 255(21–22):2653–2665
- Do YR et al (2003) Enhanced light extraction from organic light-emitting diodes with 2D SiO₂/SiNx photonic crystals. *Adv Mater* 15(14):1214
- Do YR et al (2004) Enhanced light extraction efficiency from organic light emitting diodes by insertion of a two-dimensional photonic crystal structure. *J Appl Phys* 96(12):7629–7636
- Elschner A et al (2000) PEDT/PSS for efficient hole-injection in hybrid organic light-emitting diodes. *Synth Met* 111:139–143
- Eom SH, Wrzesniewski E, Xue JG (2011) Close-packed hemispherical microlens arrays for light extraction enhancement in organic light-emitting devices. *Org Electron* 12(3):472–476

- Fuhrmann T et al (2003) Guided electromagnetic waves in organic light emitting diode structures. *Org Electron* 4(4):219–226
- Fujita M et al (2005) Optical and electrical characteristics of organic light-emitting diodes with two-dimensional photonic crystals in organic/electrode layers. *Jpn J Appl Phys Part Lett Express Lett* 44(6A):3669–3677
- Hung LS, Tang CW, Mason MG (1997) Enhanced electron injection in organic electroluminescence devices using an Al/LiF electrode. *Appl Phys Lett* 70(2):152–154
- Ishihara K et al (2007) Organic light-emitting diodes with photonic crystals on glass substrate fabricated by nanoimprint lithography. *Appl Phys Lett* 90(11):111114
- Jonas F, Schrader L (1991) Conductive modifications of polymers with polypyrroles and polythiophenes. *Synth Met* 41(3):831–836
- Kim YH et al (2013a) Achieving high efficiency and improved stability in ITO-free transparent organic light-emitting diodes with conductive polymer electrodes. *Adv Funct Mater* 23(30):3763–3769
- Kim JB et al (2013b) Highly enhanced light extraction from surface plasmonic loss minimized organic light-emitting diodes. *Adv Mater* 25(26):3571–3577
- Koo WH et al (2010) Light extraction from organic light-emitting diodes enhanced by spontaneously formed buckles. *Nat Photonics* 4(4):222–226
- Koo WH et al (2012) Light extraction of organic light emitting diodes by defective hexagonal-close-packed array. *Adv Funct Mater* 22(16):3454–3459
- Lamansky S et al (2001) Highly phosphorescent bis-cyclometalated iridium complexes: synthesis, photophysical characterization, and use in organic light emitting diodes. *J Am Chem Soc* 123(18):4304–4312
- Lee CL, Lee KB, Kim JJ (2000) Polymer phosphorescent light-emitting devices doped with tris (2-phenylpyridine) iridium as a triplet emitter. *Appl Phys Lett* 77(15):2280–2282
- Lee KH et al (2004) Mechanism for the increase of indium-tin-oxide work function by O-2 inductively coupled plasma treatment. *J Appl Phys* 95(2):586–590
- Lee JH et al (2012) A high performance transparent inverted organic light emitting diode with 1,4,5,8,9,11-hexaazatriphenylenehexacarbonitrile as an organic buffer layer. *J Mater Chem* 22(30):15262–15266
- Meerheim R et al (2010) Quantification of energy loss mechanisms in organic light-emitting diodes. *Appl Phys Lett* 97(25):253305
- Moller S, Forrest SR (2002) Improved light out-coupling in organic light emitting diodes employing ordered microlens arrays. *J Appl Phys* 91(5):3324–3327
- Mulder CL et al (2007) Saturated and efficient blue phosphorescent organic light emitting devices with Lambertian angular emission. *Appl Phys Lett* 90(21):211109
- O'Brien DF et al (1999) Improved energy transfer in electrophosphorescent devices. *Appl Phys Lett* 74(3):442–444
- Reineke S et al (2009) White organic light-emitting diodes with fluorescent tube efficiency. *Nature* 459(7244):234–U116
- Riedel B et al (2010) Enhancing outcoupling efficiency of indium-tin-oxide-free organic light-emitting diodes via nanostructured high index layers. *Appl Phys Lett* 96(24):243302
- Sasabe H et al (2013) Extremely low operating voltage green phosphorescent organic light-emitting devices. *Adv Funct Mater* 23(44):5550–5555
- Slotsky M, Forrest SR (2010) Enhancing waveguided light extraction in organic LEDs using an ultra-low-index grid. *Opt Lett* 35(7):1052–1054
- Sugiyama K et al (2000) Dependence of indium-tin-oxide work function on surface cleaning method as studied by ultraviolet and x-ray photoemission spectroscopies. *J Appl Phys* 87(1):295–298
- Sun Y, Forrest SR (2006) Organic light emitting devices with enhanced outcoupling via microlenses fabricated by imprint lithography. *J Appl Phys* 100(7):073106
- Sun Y, Forrest SR (2008) Enhanced light out-coupling of organic light-emitting devices using embedded low-index grids. *Nat Photonics* 2(8):483–487

- Tang CW, Vanslyke SA (1987) Organic electroluminescent diodes. *Appl Phys Lett* 51(12):913–915
- Tsutsui T et al (1999) High quantum efficiency in organic light-emitting devices with iridium-complex as a triplet emissive center. *Jpn J Appl Phys Part Lett Express Lett* 38(12B):L1502–L1504
- Wei MK, Su IL (2004) Method to evaluate the enhancement of luminance efficiency in planar OLED light emitting devices for microlens array. *Opt Express* 12(23):5777–5782
- Xi JQ, Kim JK, Schubert EF (2005) Silica nanorod-array films with very low refractive indices. *Nano Lett* 5(7):1385–1387
- Xi JQ et al (2007) Optical thin-film materials with low refractive index for broadband elimination of Fresnel reflection. *Nat Photonics* 1(3):176–179
- Zhou JH et al (2011) Roughening the white OLED substrate's surface through sandblasting to improve the external quantum efficiency. *Org Electron* 12(4):648–653

White OLED Lighting Panel Manufacturing Process

Jeffrey P. Spindler, John W. Hamer, and Marina E. Kondakova

Contents

Introduction	385
Panel Design Considerations	386
Manufacturing Process for OLED Lighting Panels	389
White OLED Device Configurations	393
Process Challenges for White OLEDs	403
Yield and Reliability	410
Future Directions	411
References	413

Abstract

This chapter describes the typical manufacturing processes used to fabricate white OLED lighting panels, with an emphasis on OLEDs produced by dry manufacturing methods such as VTE. OLED panel fabrication is typically classified into three categories: front-end fabrication, sometimes referred to substrate or backplane fabrication, OLED device fabrication, and back-end fabrication including encapsulation and packaging. This chapter mainly focuses on the substrate and OLED device manufacturing processes.

Introduction

OLED display manufacturing began in the late 1990s when Pioneer commercialized the first passive-matrix OLED (PMOLED) displays for aftermarket car stereo applications. The first active-matrix OLED (AMOLED) displays were commercialized in 2001 in a joint venture between Kodak and Sanyo called SK Display

J.P. Spindler (✉) • J.W. Hamer • M.E. Kondakova
OLEDWorks LLC, Rochester, NY, USA
e-mail: jspindler@oledworks.com

(Hamer et al. 2005). Since then, many other players have entered into the OLED display market, and the manufacturing processes for small OLED displays have become somewhat standardized. In recent years, the push toward large AMOLED TV displays has forced the main players, Samsung and LG Display, to develop and commercialize new technologies to enable scaling of the backplane and display manufacturing technologies. OLED lighting, on the other hand, has gained momentum only in the last few years as the efficiency of white OLED technology has improved to the point where the energy efficiency is now a compelling factor, with commercial OLED lighting panels from LG Chem having efficacies of 60–80 lm/W. The largest barrier to widespread commercial adoption of OLED lighting remains the high manufacturing costs. Although an OLED lighting panel is a simpler device than an AMOLED display because it avoids the thin film transistor (TFT) backplane and fine patterning requirements, the manufacturing costs are still much too high to justify a large market and to compete with incumbent low-cost lighting technologies such as linear fluorescent lighting and now increasingly efficient LED lighting. An OLED lighting panel is more similar to a PMOLED display in that it is a passive device with a simple backplane. In this chapter we will start by discussing the typical design considerations for OLED lighting panels and the impact the design choice can have on the manufacturing process. Then we will describe how patterning is achieved for the substrate and OLED device layers and describe the typical process flow for manufacturing of OLED lighting panels. Next we will focus on the specific issues pertaining to white OLEDs, namely, the device configurations and process challenges encountered in the manufacturing of OLED lighting panels. Finally, we will describe the yield issues that can impact OLEDs and discuss future trends and considerations for the manufacturing of white OLED lighting panels.

Panel Design Considerations

Although an OLED light is a relatively simple device by construction, there are a number of important design parameters that must be considered up front. To produce a large-area OLED lighting panel with high efficacy and uniform light emission across the panel, the electrical and photometric characteristics of the OLED stack must first be known. The relationship of current and luminance as a function of voltage, referred to as I-V-L or J-V-L characteristics, needs to be measured for the targeted brightness which is usually in the range of 1,000–5,000 cd/m² for typical lighting applications. The efficacy of the lighting panel is mostly determined by the efficiency of the OLED stack, and to that end high-efficiency phosphorescent OLED materials are required (Ma et al. 2011). The luminance uniformity is largely determined by the resistivity of the transparent anode electrode (Neyts et al. 2006) which is usually ITO or another transparent conductive oxide (TCO). The resistance of the metal cathode is negligible compared to that of the ITO anode in the case of a typical bottom-emitting OLED. The limited conductivity of ITO causes a voltage drop across the anode plane, which in turn causes a luminance inhomogeneity by reducing the current delivered to the OLED. The choice of OLED architecture also has a large

impact on the uniformity. Higher voltage and stacked OLED architectures effectively increase the vertical resistance through the organic stack, which allows the current to be distributed further through the anode electrode and improves the luminance uniformity (Park et al. 2009). The slope of the I-V-L curve is reduced such that the voltage drop across the anode produces a smaller change in luminance. Higher-efficiency OLED stacks can produce the desired luminance at a lower current density, so not only is the slope of the I-V-L curve reduced, but the panel requires less current which further reduces the resistance through the anode electrode. Luminance nonuniformity is not just a cosmetic or visual defect. It is also a reliability concern for several reasons. Joule heating due to locally high current density further increases the nonuniformity (Garditz et al. 2007) and creates a thermal gradient across the panel. Organic materials exposed to locally higher temperatures will age faster potentially causing color gradients as well as brightness gradients and may be more susceptible to shorting. OLED lifetime is a sensitive function of temperature, so it is always beneficial to limit the panel temperature. It has been reported that a $1.65\times$ improvement in lifetime can be expected for every 10 K reduction in panel temperature (Levermore 2011a, b).

Typical ITO films used in OLEDs are in the range of 50–150 nm thick and have a sheet resistance of 10–30 Ω /square. The ITO layer can be made thicker in order to reduce the resistance to below 10 Ω /square; however, the transparency becomes an issue and too much absorption occurs. The problem of limited conductivity of the ITO has been addressed by introducing metallic grids in contact with the ITO, which reduces the effective resistance of the anode electrode. The overall panel design determines the requirements for the metal grids. The size of the panel and the OLED I-V-L characteristics determine the overall current required to produce the desired lumen output. The next consideration is how the current will be distributed around the perimeter of the panel. Typically, the current is distributed evenly around the perimeter of the panel by the use of wide metal bus bars or other perimeter conductors such as flexible printed circuits that are attached around the edges. This ensures the least distance that the current must flow from the edge toward the center of the panel. If the current is fed from only one side and not distributed around the perimeter, then the total current must flow from one edge to the other, which places much greater requirements on the conductivity of the anode electrode and may limit the effective panel width. A typical grid design is a simple square mesh design that has been used by some OLED lighting panel manufacturers such as LG Chem. Other grid designs have been investigated, and the hexagonal layout was found to be the most efficient in terms of minimizing power loss (Neyts et al. 2008). This design has been employed by Osram in their Orbeos OLED lighting panels. The width, thickness, and density of the grid structures determine the effective resistance of the anode. The goal is to maximize the fill factor or utilization of the emitting area by minimizing the coverage of the metal. Higher conductivity metals allow for narrower linewidths and reduced thickness of the metal film, and for that reason Al or Cu are good choices (Park et al. 2012). Multilayer stacks such as Mo/Al/Mo are commonly used and are adopted from the LCD and AMOLED display industry, since the Mo layer makes good ohmic contact with ITO and stabilizes the Al

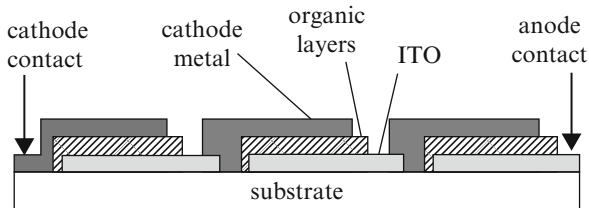
interface. Linewidths of the metal traces are typically kept below 100 μm , such that they are not visible from a reasonable viewing distance of around 1 m. Diffusing light extraction films used by most OLED lighting panel manufacturers further reduce the visibility of the metal grid lines. The density or pitch between metal lines is typically 1–2 mm. Since the metal grid networks distribute the current throughout the panel, the ITO requirements can be relaxed somewhat. The ITO only needs to spread the current uniformly between the metal grids, so its sheet resistance can be higher and thickness can be reduced, which in turn may reduce absorption and improve the light outcoupling efficiency. The metal film thickness is typically in the range of 200–500 nm. Since the total organic stack thickness is usually in the same range, the metal traces must be planarized with an insulator material to ensure good step coverage and prevent shorting between the anode and cathode. The organic layers alone are not sufficient, since the evaporation process used to deposit them is directional in nature; therefore, step coverage along metal lines or other topography is insufficient. An insulator material such as a positive photoresist around 1 μm thick is an effective approach to the planarization of patterned metal and ITO edges. The additional cost associated with the metal and insulator layers must be considered.

Another design approach that has been considered is the serial connection technique, where the ITO electrode is segmented into small squares or stripes, and the organic layers are patterned to allow the cathode from one segment to connect serially to the ITO anode of the next segment, as depicted in Fig. 1.

Since it avoids the need for metallic grids, this method can result in lower manufacturing costs. One drawback to this method is that visible non-emitting borders will be present between segments to allow for the serial connection. Using practical shadow mask techniques, it is difficult to reduce the border width to less than 0.5 mm. With the serial approach, the overall voltage is multiplied by the number of segments, while the current remains the same as for a single segment. If a tandem structure is used with a nominal drive voltage of 6 V, then a 4-segment design would require 24 V. Voltage limitations should be considered so that products comply with low-voltage safety standards, and some applications requiring low-voltage lighting may not be possible.

As mentioned before, heat generation and dissipation are very important considerations in the design of a large-area OLED lighting panel. The typical operating current density for state-of-the-art OLED lights is in the range of 1–5 mA/cm^2 , which is orders of magnitude lower than the current density for high-brightness

Fig. 1 Schematic of serially connected OLED panel design (Reproduced from Duggal et al. (2004))



inorganic LEDs. However, the heat produced by the OLED, although relatively small compared to LEDs, must still be contended with. The choice of encapsulation system can impact how the heat is removed from the OLED panel. The traditional approach to OLED encapsulation is to attach a glass or metal lid having a recessed pocket to the OLED substrate using a perimeter adhesive. The recessed portion of the lid allows space for a getter material to be placed inside the package in order to absorb any moisture that may permeate through the edge seal over time. Although effective, this technique leaves a gas-filled space between the metal cathode of the OLED and the lid. The gas, usually nitrogen, is not a very effective thermal transfer medium; therefore, the space impedes the heat transfer out of the OLED (Bergemann et al. 2012). Solid-state encapsulation methods have been developed to enable large AMOLED televisions and flexible OLED displays and lighting (Han et al. 2014; Hong et al. 2014). Typically a thin film barrier layer or multiple layers are formed directly over the OLED cathode, and then an adhesive layer is used to attach the OLED substrate to a metal foil, thin glass, or barrier-coated plastic film. The adhesive can have multiple components that function as getters of moisture and as thermal transfer media. Solid-state encapsulation has the benefit of eliminating the gas void inside the OLED package, which aids in dissipating heat. A further benefit is that differential pressure, which can be a yield issue with high-altitude shipping, is no longer an issue.

The tolerance of all the manufacturing process steps should be known prior to designing the panel. The design rules should account for tolerances in the glass cutting process, mask alignment processes for the organics and cathode layers, and encapsulation tolerances in the positioning and width of the perimeter adhesive seal. The overall goal for the panel design is to minimize the unlit area, most of which is around the perimeter of the panel, in order to maximize the emitting area. A higher lumen output per panel reduces the area-related cost, expressed as cost per lumen. Although the OLED can be driven at higher current density to increase the lumen output, the negative impacts on uniformity, heating, and lifetime must be understood before considering this approach.

Manufacturing Process for OLED Lighting Panels

This section describes the typical process flow, materials, and patterning methods used in manufacturing OLED lighting panels. All commercial rigid OLED panels are currently made on a glass substrate with typical thickness in the range of 0.5–1.1 mm. Many OLED producers have chosen to use soda-lime glass for lower cost, while others have adopted alkali-free borosilicate glass, the standard in displays. Soda-lime glass is more cost effective; however, thickness is limited due to the more fragile nature of the soda-lime float glass. Borosilicate and other versions of display glass can be made with thickness in the range of 0.3–0.7 mm for rigid panels and in the range of 0.05–0.2 mm for flexible panels. Flexible OLEDs can also be constructed on plastic substrates coated with moisture barrier layers. Recently Konica Minolta has invested in a roll-to-roll production line for flexible OLED

lighting using their own barrier layer technology coated on a plastic film. OLEDs have also been demonstrated on metal foil substrates, which may allow for more effective heat removal from the OLED. The metal foil must be effectively planarized with smoothing layers to ensure low roughness in order to avoid shorting and leakage problems. Due to the opaque substrate, the OLED must be a top-emitting configuration, which places greater requirements on the transparent cathode and transparent encapsulation. So far, OLED lighting panels on metal substrates have not been commercialized. Glass substrates are typically purchased in the desired production size with a ground or beveled edge which eliminates the sharp edge and reduces likelihood of breakage or particulate formation during the OLED manufacturing process. Currently, the common production substrate sizes are Gen2 (370 × 470 mm) or Gen2.5 (400 × 500 mm). When the OLED lighting market grows, it is expected that larger generation production facilities will be established to meet the demand, as in the case of display production. One of the leading manufacturers, LG Chem, has announced intentions to build a Gen5 (1.25 × 1.1 m) OLED lighting factory. Similar to the display industry, multiple lighting panels or tiles are arranged on the motherglass substrate. For instance, a common OLED panel size of 100 × 100 mm can be fit into a 3 × 4 array on the Gen2 motherglass, while leaving an unused border around the perimeter of the substrate where there are excessive nonuniformities and defects due to handling. This exclusion zone is typically a minimum of 10–15 mm. The motherglass substrate is processed as a whole through all of the front-end and OLED device fabrication steps. After encapsulation of the individual panels, the panels are then separated from the motherglass for further processing and testing.

Improving the light extraction efficiency of OLEDs is one of the key technological development areas required for OLEDs to reach over 100 lm/W efficacy. Many different light extraction approaches have been investigated (Saxena et al. 2009). It is well known that only about 20–25 % of the photons generated within the organic layers escape into the air, while the remaining photons are lost due to absorption, waveguiding in the substrate or within the organic layers and ITO, or energy transfer to metal surface plasmon polaritons at the cathode interface (Lu 2013). One of the most promising techniques to improve extraction efficiency is to apply internal extraction structures on the glass substrate, which contain scattering particles and high-index materials that more closely match the index of refraction of ITO and organic materials. Several glass manufacturers are developing integrated OLED substrates that contain embedded scattering particles, high-index coatings, or combinations of layers to improve the outcoupling of light trapped inside the ITO and organic layers (Nakamura et al. 2013; Lecamp 2013; Hung et al. 2014; Taylor 2015). Others are developing integrated substrate technologies based on nanoimprint lithography techniques that are more compatible with flexible substrate processing (Slafer 2014; Kobrin 2014), while some are developing the high-index materials that can be coated onto the glass substrate (Cooper 2015). At least one OLED panel manufacturer claims to have developed their own internal light extraction technology and demonstrated white OLEDs with 80 lm/W efficacy (Moon et al. 2013). To be effective, the internal extraction layer (IEL) is applied between the glass substrate

having an index of 1.5 and the ITO/organic materials having an index of 1.8 or higher. If the IEL is an inorganic-type material with good moisture resistance, then it may not need to be patterned. However, certain polymer-based IELs may need to be patterned within the encapsulated area of the OLED device to prevent a potential pathway for moisture to enter underneath the adhesive seal.

Most OLED lighting panel manufacturers currently purchase glass substrates with patterned ITO and in some cases with patterned metal and insulator layers. Some manufacturers are more vertically integrated and have their own large-area lithography and wet processing facilities for internal production of substrates, particularly larger display companies that may have depreciated equipment obsoleted for display production. Due to the large capital investment, it is cost prohibitive to establish a new photolithography line for production of OLED substrates, and therefore it is more economical to purchase substrates from an established production facility. The ITO, metal, and insulator layers are patterned by photolithographic techniques developed for the flat panel display industry. The ITO film is deposited by physical vapor deposition (PVD) techniques such as DC magnetron sputtering or RF sputtering. In order to pattern the ITO, a photoresist layer is coated onto the substrate by either spin-coating or slit-coating techniques, and then the layer is exposed through a photomask on a large-area stepper or step-and-scan system. These large-area optical lithography tools are very complex and expensive, as they are designed with large-area exposure fields that can produce features with 2–3 μm resolution and can align subsequent patterned layers with less than 1 μm accuracy. The pattern design on the photomask is imaged into the photoresist layer during the exposure process, and then the substrate is developed to form the physical pattern in the photoresist. The substrate is then baked at a moderate temperature around 130 $^{\circ}\text{C}$ to harden the photoresist material and make it resistant to acids that are used to etch the ITO. The ITO film is typically etched in a mixture of hydrochloric acid, water, and sometimes nitric acid. After transferring the pattern into the ITO, the photoresist layer is stripped in a solvent solution. The metal layer is also deposited by sputtering and then patterned by a similar photolithography and etch process cycle. The metal layer is etched in an acidic solution, and the chemistry depends on the type of metal used. Aluminum is most often etched in a mixture of phosphoric, acetic, and nitric acids diluted with water. The insulator layer, usually a photosensitive polyimide or novolak resin material, is also patterned by photolithography. Since this material remains on the substrate, it is not stripped off but rather cured at high temperature above 200 $^{\circ}\text{C}$ to polymerize the material. The high-temperature baking process also allows the bulk material to flow somewhat and create a tapered edge profile, which will allow for good step coverage of the thin organic layers over the insulator features and minimize the chances of the OLED device shorting. Because the cost of photolithography patterned substrates is relatively high, lower cost patterning methods are being investigated. For patterning ITO, laser ablation techniques have been developed and demonstrated on large scale. Direct printing methods such as inkjet printing, screen printing, and aerosol jet printing have been demonstrated for printing of metal lines and insulator materials.

Before deposition of the organic materials, the substrate undergoes several preparation processes such as cleaning, baking to remove moisture, and plasma or UV ozone treatment to remove organic contaminants and increase the work function of the ITO surface. In some cases, a surface modification layer such as a plasma-deposited fluorocarbon thin film buffer layer can be beneficial (Kondakova et al. 2010). The baking and surface preparation processes are typically done just prior to coating the organic layers, and precautions are taken to avoid exposure to moisture and oxygen. Although this chapter is focused on dry manufacturing methods, a wet-coating method for the first organic layer, the hole-injection layer (HIL), is worth mentioning since some OLED manufacturers have adopted this hybrid process (Komoda 2011). A solution-coated HIL has been shown to be effective at planarizing defects and roughness that may cause shorting or leakage currents. The wet HIL can be applied by direct patterning methods, such as inkjet printing, or by slot-die coating with selective patterning or removal (Gibson and Snodgrass 2011). The HIL film is baked to remove residual solvents or moisture.

Substrates are loaded into the VTE system through a load lock chamber, which is pumped down to pressure equivalent to the main vacuum chamber before transfer. Load locks allow the main vacuum chamber to be kept under high vacuum, except when it needs to be vented for periodic maintenance purposes. Substrates are typically transferred onto a mask frame that holds the organic shadow mask used for in situ patterning of the organic materials. The mask is made of a thin metal to minimize shadowing of the deposition patterns and can have magnetic properties as sometimes magnets are employed to keep the mask in close contact with the substrate during deposition. The substrate is loaded process side down onto the mask, and the whole frame travels through the VTE system for deposition of the organic layers. Mechanical alignment of the substrate to the mask is achieved by nesting the substrate toward one corner against pins located in the mask frame. Active alignment used for precision masking of AMOLED displays is costly and not necessary for OLED lighting devices. After all organic layer depositions are complete, the substrate is transferred to another mask frame containing the cathode shadow mask. After cathode deposition, the substrate is removed from the mask frame and transferred to an unload chamber, which is also a load lock chamber. Since the OLED must be kept under vacuum or inert atmosphere until it is encapsulated, the unload chamber is often attached to a drybox where encapsulation is performed. The unload chamber can be vented with nitrogen directly into a nitrogen-purged drybox.

Traditional encapsulation methods using a pocketed cover glass lid or metal lid are accomplished with automated equipment that mates the OLED substrate, which is still process side down, with the prepared lid. In the case of glass encapsulation, pockets are preformed in the glass by acid etching. The pocket depth is typically 100–300 μm , and the pocket design matches the design of the OLED panel. The cover glass is cleaned by plasma treatment or UV ozone, which is often integrated as part of the encapsulation process. The desiccant is applied inside the pockets in the form of a dispensable desiccant or precut getter tape which can be purchased commercially. The adhesive seal is then dispensed around the perimeter of each

pocket, and then the substrate is precisely lowered onto the prepared cover glass, usually under a slight vacuum to allow for better control of the adhesive seal quality and to maintain a slight negative pressure inside the cavity. The adhesive is usually a UV curable material, to allow for rapid curing while the substrate and cover are still pressed together. Often spacer beads are employed in the adhesive material for precise control of the gap between substrate and cover. The process is much the same when metal cans are used, except that individual cans must be held in mechanical fixtures, and UV curing must be done through the substrate since the cans are opaque. Once the encapsulation process is complete, the packaged substrate is unloaded from the drybox for post processing. Each individual panel is “singulated” from the motherglass substrate by either mechanical scribing or laser scribing. Mechanical scribing is more cost effective, but laser scribing produces defect-free edges and can increase the yield due to elimination of breakage. When a single cover glass sheet is used for encapsulation, both the substrate and the cover glass must be scribed before separation. Once the scribing process is complete, individual panels are separated manually or with the use of an automated machine.

At this point, panels are inspected and performance tested to determine yield and quality. Panels that pass initial quality checks are typically further tested for reliability. Since OLEDs are susceptible to electrical short formation where particle defects may have occurred during the production process, it is necessary to operate the panels for a period of time to eliminate shorted panels or ones that may potentially short circuit. Infrared imaging can be used to identify hotspots where high amounts of leakage current occur around a defect (Zhou et al. 2000). Most often panels are set on “burn-in” stations and operated at normal driving conditions or slightly elevated conditions for a period of days or weeks, to allow for infant failures to be sorted out. Most OLED panel manufacturers apply an external extraction film (EEL) to the emitting side of the glass, which is either a diffusive-type film containing scattering particles or a structured film having microlenses or other features. This film allows most of the light trapped inside the glass to escape and can boost the integrated light output by 50 % or more. The EEL film is also a form of protection against damage to the glass surface and can be effective at reducing the amount of color shift observed off angle. Since the OLED lighting market is in the very early stages, there are no standard methods for electrical connection and packaging (Spindler et al. 2011). The same panel manufacturer can offer several different levels of packaging, anywhere from the bare panel to a fully integrated module with drivers or anywhere in between. Most commonly, panels are requested by luminaire manufacturers with a simple 2-wire connection or with a plug-in type connector attached to the panel.

White OLED Device Configurations

OLED lighting is a topic that is attracting a lot of interest from governments, academia, and industry because of its promise for a green, environmentally friendly technology that creates novel lighting shapes and forms. It is unique as it naturally offers a pleasant, uniform, diffused light, minimizing the need for fixtures.

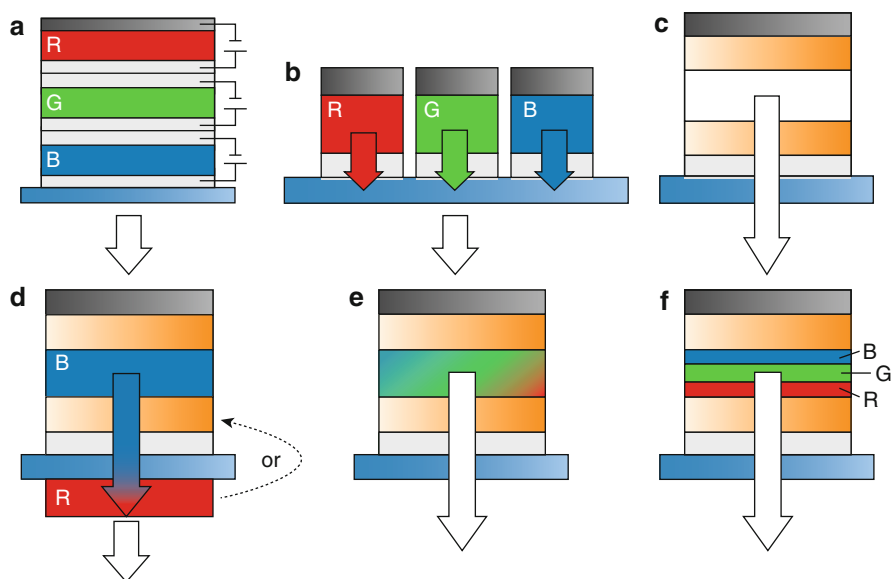


Fig. 2 Various device layouts to realize white-light emission. (a) Vertically stacked OLEDs, (b) pixilated monochrome OLEDs, (c) single-emitter-based white OLEDs, (d) blue OLED with downconversion layers, (e) multiple-doped emission layers (EMLs), and (f) single OLEDs with sublayer EML design. (c–f) Shaded layers represent optional functional layers, e.g., transport layers [not shown for (a) and (b) for better visibility]. *R*, *G*, and *B* stand for red, green, and blue, respectively (Reproduced from Reineke et al. 2013)

The efficacy values measured on OLED devices are therefore very close to what end users experience in actual applications. OLED panels have the further advantages of being extremely thin (<2 mm), lightweight, and operating at temperatures near ambient. During the recent years, white OLED devices have undergone a very fast development as they emerge as next-generation lighting sources.

OLEDs are current-driven devices that utilize emission from the electronically excited molecules. The operation of OLEDs involves charge injection from the electrodes into the adjacent organic layers, transport of injected charge carriers through organic layers, and exothermic recombination of holes and electrons to generate excited states of molecules (excitons), followed by their deactivation by emitting either fluorescence or phosphorescence, which is taken out of the device as electroluminescence (EL) (Tang and VanSlyke 1987; Kalinowski 1999; Brutting and Frischeisen 2012).

Different approaches used for generation of white emission are shown in Fig. 2 (Reineke et al. 2013). The architectures include multiple emitters in a single emission region, stacked OLEDs with multiple emissive regions, single-emitter-based white OLEDs, and single color-emitting OLED in combination with down-conversion layer. The details of these structures have been discussed in various publications.

Each architecture has its advantages and disadvantages. Single-stack devices generally provide lower voltage compared to stacked structures, have reduced number of layers, require less materials, and, in general, are easier to fabricate. However, it may be difficult to adjust color and maintain color stability during the operation of such devices as various emitters can degrade with different rates.

Stacked (tandem) OLED structures were developed to improve luminous efficiency and increase lifetime. This was accomplished by vertically stacking several individual EL units and driving the entire device with a single power source (Matsumoto et al. 2003; Liao et al. 2004). The EL units are electrically connected in series by inserting an intermediate connector (p-n junction) between adjacent units. Each time a hole/electron (h^+/e^-) pair is generated at the electrodes, each p-n junction also generates a (h^+/e^-) pair. Thus, for an N-stack tandem OLED, N hole/electron pairs are formed for every injection event generated by the electrodes. Each of the N (h^+/e^-) pairs can generate a photon of light, one per electroluminescent unit, resulting in luminous yield and external quantum efficiency that are N times higher than the analogous single-stack OLED. The major advantage of tandem OLEDs is that they are N times brighter than a single-stack device at the same current density and that they can achieve the same brightness at a lower current density. Because lifetime is usually superlinear with current density, this leads to an increase in lifetime of the tandem OLED that is greater than a factor of N times that of the single-stack OLED at the same brightness. The balancing factor is that the voltage of a tandem OLED is also about N times larger than that of the single-stack OLED, resulting in a similar power efficiency for the tandem and single-stack architectures at a given current density. However, because the tandem devices need to operate at lower current densities to achieve the same brightness as single-stack devices, they typically consume less power and have considerably longer lifetimes.

There are other important features of the stacked structure, such as flexibility in color adjustment, reduced short circuit defects, and improved uniformity of light emission. As mentioned earlier in the chapter, the use of the stacked structure also helps to reduce resistive power losses and Joule heating, thus extending device lifetime. All these features are particularly important for the operation of large-size lighting panels.

Figure 3 shows performance data, such as voltage–current dependencies, efficiency versus current density, and brightness uniformity for single-stack and tandem phosphorescent amber devices and hybrid white stacked OLEDs. Amber OLEDs are defined as devices emitting light with the color coordinates of $CIE_x = 0.56$ and $CIE_y = 0.43$ (wikipedia.org). OLED devices combining fluorescent and phosphorescent emitters are termed hybrid. Data show that luminous yield of stacked amber devices increases in proportion to the number of stacks. Drive voltage also increases as tandem structure becomes more complex. However, as the number of stacks rises, slope of I-V curves becomes less steep. During OLED operation, since resistivity of ITO is relatively high, a significant voltage drop occurs as the current passes through the electrode. In devices with steep I-V curves, a small voltage difference can result in a large difference of the operating current, creating large nonuniformity in luminance. Figure 3d shows that the brightness of the white OLED panel is

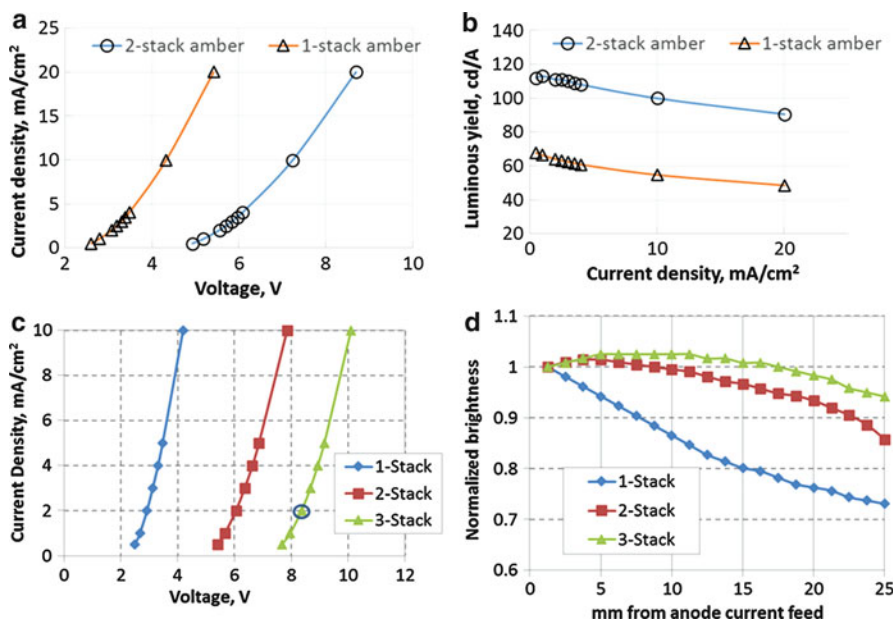


Fig. 3 (a, c): Current density vs voltage dependencies for single-stack and double-stack phosphorescent amber devices (a) and multiple stack white-emitting devices (c); (b): luminous yield shown with external light extraction layer as a function of current density for phosphorescent amber devices; (d): brightness measured at various distances across anode for white devices

measured at various distances across a 25 mm long OLED device, starting at the point where the anode current is fed. Devices with triple-stack structure are the most uniform; single-stack white devices demonstrate a sharp decrease in luminous efficiency as a function of distance from anode edge.

One of the first triple-stack white devices with separate EL unit (fluorescent blue, phosphorescent green, and red) was reported by Novald (Birnstock et al. 2008). The red and green phosphorescent EMLs were placed at the first and second optical maximum, respectively, and the blue fluorescent EML was placed in the third maximum (the blue EL stack is the closest to the anode). Measured with an outcoupling enhancement film at 1,000 cd/m², the devices showed 38 lm/W, CIE_x, *y* of (0.43, 0.43), and CRI of 90. Most importantly, the operational stability of the white OLED (*T*₅₀, i.e., the time to half initial luminance) exceeded 100,000 h.

Performance of stacked devices has improved significantly since that time. LG Chem, one of the leading companies in OLED lighting, recently developed 3-stack hybrid white devices showing 80 lm/W at 3,000 cd/m², CRI of 84, and color temperatures of 2,900 K as shown in Table 1; the devices include both internal and external light extraction layers and extraction efficiency is 1.8×. Lifetime (*T*₇₀, time to 70 % of initial luminance) of the panels is reported to be 30,000 h (Moon et al. 2013). Advances in the device encapsulation process further improved lifetime to 40,000 h (LG Chem roadmap, LGchem.com). Hybrid devices will be discussed in

Table 1 Performance of the state-of-the-art stacked white devices at 3,000 cd/m²

Company/ emitters	Device structure	Light Extraction	Lm/W	CCT, K	CRI	L70, Kh	References
UDC/ phosphorescent	RG-B//RG-B	1.68X external	54	2,990	83	19.5	Adamovich et al. 2012
UDC/ phosphorescent	Y-B//Y-B	1.72X external	60	2,882	<80	25	Xu et al. 2013
Panasonic/ phosphorescent	B-R//G-R	2X internal	102	2,550	80	10	Yamae et al. 2013
LG Chem/ hybrid 3-stack	Y-R//B//Y-R	1.8X internal	82	2,900	84	40	Moon et al. 2013
LG Chem/ hybrid	B//YR	External	60	3,000	>90	20	LGChem.com
Philips/hybrid 6-stack	Y//B//Y// Y//B//Y	External	40–50 ^a	3,000	80	>10 ^a	Dotter 2015

^aData are given at 8,300 cd/m²

more detail later this chapter. Stacked architectures also can be applied in flexible substrates, such as thin-glass or plastic substrates. Recently, LG Chem announced that the company developed truly flexible OLED panel on plastic substrate with bending radius flexibility of 30 mm. The new plastic-based OLED light panel shows 60 lm/W, 3,000 K CCT, and CRI over 85.

There is an increasing convergence of lighting that is mutually energy efficient and healthy, promoting well-being, rest, and productivity. Both of these needs find roots in the natural lighting cycle of the sun. In sync with nature, the human body responds physiologically to the daily patterns of changing wavelengths. Consequently, there is significant interest in dynamically tuned OLED lighting. These factors are important for OLED market adoption.

Color-tunable OLEDs can be fabricated by several methods: (1) voltage-controlled color tuning can be realized in a single-stack OLEDs; (2) vertically stacking of several OLED units, each having its own emitter material in each emission layer; and (3) fine lateral structuring of monochrome OLEDs and mixing of the produced light with an optional combination of a scattering film on top, for better homogeneity. The latter design is called “striped,” and this type of devices is relatively well developed (Weaver et al. 2014, Verbatimlighting.com). In contrast, vertically stacked color-tunable OLEDs, which might be more suitable for use as large-area lighting sources than striped OLEDs (as they are generally less expensive to fabricate and do not require scattering foils to achieve uniform light), are just in the beginning of research. To the best of our knowledge, there are very few reports on vertically stacked tunable OLEDs with practical performance. In such devices an intermediate electrode is interposed between the vertically stacked light-emitting diodes with different emission colors, and the diodes are connected to the outside. By applying voltage to the individual OLED stack(s), the device can emit any mixture of color hues. However, achieving necessary transmission properties of the electrode presents a fundamental challenge (Burrows et al. 1996; Shen et al. 1997).

Green-red and green-blue color-tunable OLED devices with an Al/Au intermediate electrode were demonstrated (Liang and Choy 2009; Zheng and Choy 2008). The color can be tuned along a line from green to red or green to sky blue, respectively. Transmission of the optimized Al/Au electrode is reported to be over 65 % (Zheng and Choy 2008).

Another interesting approach to obtaining color change in OLEDs is to develop voltage-controlled color-tunable devices (Kalinowski et al. 1996; Huang et al. 2004; Kohnen et al. 2008). In such devices the recombination zone shifts from one light-emitting layer (EML) to the adjacent EML resulting in shifting color. Jou et al. developed single-stack devices with color temperature tunable from 1,700 K to 5,200 K (Jou et al. 2012). The devices contain phosphorescent blue and orange EMLs separated by a layer made of electron-transport materials with various triplet and HOMO–LUMO energies. Changing properties of the electron-transporting interlayer allows to influence charge injection and transport into the emissive layers which, in turn, results in wide range of CCT variation. The devices can be driven with pulse-width-modulation mode which allows for independently control brightness.

As was mentioned before, OLED SSL can provide energy-efficient and environmentally friendly lighting. White light is preferred for general lighting purposes. White color is often given in terms of chromaticity coordinates (either x-y or u' - v'), but this is not intuitive. As depicted in the 1976 u' - v' chromaticity diagram in Fig. 4,

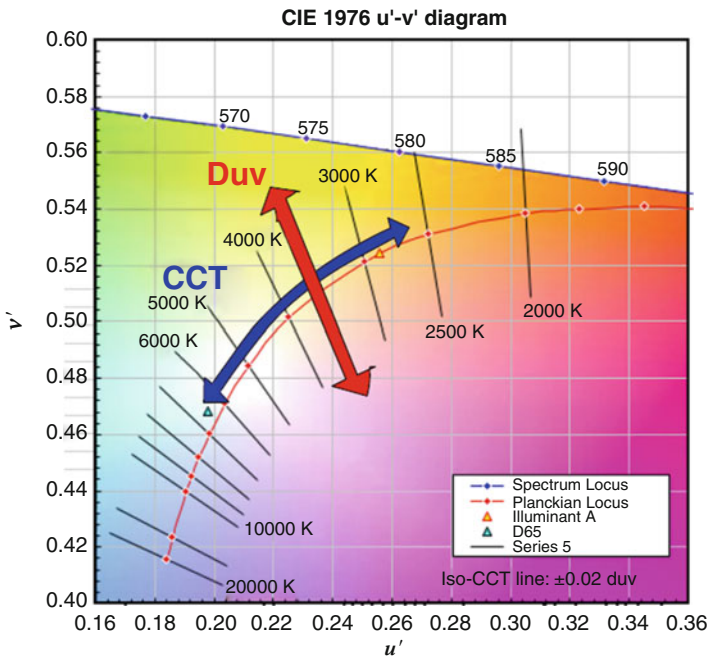


Fig. 4 CIE 1976 u' - v' diagram showing CCT and Duv metrics (Ohno 2014)

white light is distributed along the Planckian locus and defined by color temperature. If a light source emits with a spectrum which is displaced from the Planckian curve, then its chromaticity is defined by correlated color temperature (CCT) (Hunt 1988). Duv is also important to describe the magnitude and direction of the white color deviation from the Planckian locus; CCT and Duv combined can provide color information intuitively (Ohno 2014). Standard illuminant A is a Planckian radiator with color temperature of 2,856 K and is called the warm-white point. 1931 CIE_{x,y} coordinates for Illuminant A is (0.448, 0.408); the D65 point has CIE _{x,y} of (0.31, 0.33) and CCT of 6,500 K (en.wikipedia). White OLEDs for display application require “cool” white color, e.g., white emission with CCTs of 6,500–10,000 K CCT. Lighting applications call for warm-white emission with lower color temperatures of 2,700–5,000 K. To ensure that the emitted light has the correct color for lighting, the Department of Energy published Energy Star specifications for lighting sources requiring that color coordinates of light falls within one of the eight tolerance quadrangles (CIE color coordinates fall within 0.005 of Planckian locus) (energystar.gov).

Another important parameter of a lighting source is its color rendering index (CRI). The CRI represents how satisfactory an illuminant renders the true colors of different objects and can be varied from 0 to 100. (For more details, see Hunt 1988.) For Energy Star certification, lighting sources are required to have a CRI above 80. Due to their naturally broad emission spectrum, OLED light sources can easily achieve CRI over 80, and with the correct choice of emitters and spectral tuning, CRI over 90 is realizable.

In OLED devices light is formed when the electrons and holes recombine in the light-emitting layer, creating an excited state on an organic molecule, which then relaxes to its ground state, providing a photon during relaxation. Several mechanisms to energy loss exist in the conversion from electricity to light. Firstly, the emitters that were originally used were fluorescent emitters, which permitted light production only from singlet excited states within the organic molecules, theoretically implying that only 25 % of the electron–hole pairs can be involved in light production (yielding devices with an internal quantum efficiency (IQE) of less than 25 %). Phosphorescent emitters overcome this limitation, making it possible to convert all of the electron–hole pairs to photons, providing a 100 % internal quantum efficiency. Secondly, while the very thin layers of organic materials are semitransparent, these materials absorb a portion of the emitted light. Thirdly, the index of refraction for the organic materials can be around 1.8, and the anode is often formed on a glass layer having an index of refraction around 1.5. As such, light that is emitted at angles substantially off normal from the display surface can become trapped within the device. It is common that without optical modification, only about 25 % of the light that is created in the device will be emitted by the device. As such, external quantum efficiencies (EQE) can be expected to range from about 8 % for a reasonably good fluorescent device to theoretically as high as 25 % for an ideal phosphorescent device without specialized optical structures.

At the beginning of OLED development, a majority of emitting dopants were fluorescent materials. Fluorescent materials emit light by relaxing from their lowest

singlet excited state, S_1 , to their ground state, S_0 , which is also a singlet. S_1 on the dopant can be created when excitons formed on a host molecule during recombination transfer their energy to the dopant or by direct recombination on a charge-trapping or charge-transporting dopant molecule. Transition from any triplet state T_x to the ground state, S_0 , is formally spin forbidden and in typical organic molecules involves a nonradiative process. When it does occur, emission from a triplet state is referred to as phosphorescence. Materials with high quantum yield of fluorescence are known across the entire visible spectrum, enabling the utilization of red, green, and blue colors for display and solid-state lighting applications, and are discussed in detail in the “► [White OLED Materials](#)” chapter.

Blue fluorescent OLEDs continue to be the center of intense research to develop color that emits in the deep-blue spectrum and has sufficiently high operational stability, which is particularly important for display applications but also important for lighting since long-lived white OLEDs require a stable blue emitter. For display applications, it is essential that a blue-emitting dopant shows narrow emission (close to the long wavelength of the blue region of the spectrum $\sim 460\text{--}470$ nm). Emission at shorter wavelengths would improve the color quality but also reduce the luminous yield and operational lifetime (due to high-energy exciton formation and its chemistry). For lighting applications, dopants with broad emission spectra are preferable as they can improve the CRI of the lighting source. Table 2 shows the development and performance improvement of advanced blue fluorescent OLEDs. Molecular structures of the blue emitters and blue host materials are trade secrets of the companies, which makes it difficult to discuss why or how significant improvement of performance were achieved. It should be noted that the performance of blue OLEDs depends very strongly not only on the blue emitter but also on properties of the host and charge-transporting materials.

Phosphorescent emitters can capture and radiatively emit more of the excitons formed in the organic layers of the OLED than fluorescent materials, theoretically four times more due to the nature of the excited states in an OLED device (Baldo et al. 1999; Baldo and Forrest 2000). This in turn could result in higher OLED peak luminescent efficiencies, allowing for reduced operating currents, and possibly

Table 2 Performance of blue fluorescent emitters developed by various companies

Company	cd/A	EQE, %	CIE x,y	L50, Kh (brightness, cd/m^2)	References
IDK	8.4		0.13, 0.21	50 (1,000)	Nishimura et al. 2008
CDT	12.4		0.14, 0.12	16 (1,000)	Roberts et al. 2013
IDK		11.5	0.14, 0.09	16.5 (500)	Ogiwara et al. 2013
Merck	5.1		0.14, 0.08	45 (500)	Heil et al. 2014
Merck	4.7	7.5	0.14, 0.06	>10 (500)	Heil et al. 2014

longer device lifetimes. Phosphorescent-based OLEDs typically use phosphorescent organometallic materials doped at low concentrations (~8–10 % by weight) into a fluorescent organic host material. Most of the phosphorescent dopants reported to date have been organometallics of heavy metals such as iridium- and platinum-based compounds, but osmium- and europium-based compounds have also been demonstrated (Yersin and Finkenzeller 2008; Yersin et al. 2012). It is thought that phosphorescent devices work by the following mechanism: (a) the excited states in these devices are initially formed on the host and then transferred to lower energy states on the dopant, creating an excited triplet on the phosphorescent compound that can radiatively decay; (b) it is also possible for the dopant to trap electrons and/or holes directly to form the exciton and then radiatively decay (Baldo et al. 1999). However, transfer of the excited state between the host and dopant could still occur in this case, depending on the relative energy levels of each of the excited states, and in general, creation of highly efficient phosphorescent-based OLEDs requires a proper matching of the energy levels of the dopant and host species.

Significant progress has been made in the last several years toward successful OLEDs using phosphorescent materials. Phosphorescent red, yellow, and green emitters developed by Universal Display Corporation show excellent performance and are being used widely in display and lighting applications. Device performance is summarized in Table 3 (UDC website).

Blue (fluorescent and phosphorescent) OLEDs degrade more rapidly than green and red devices. High-energy dopants require higher-energy hosts and materials adjacent to the light-emitting layer to confine excitons and charge carriers to the EML. High-energy excitons undergoing excited state chemistry are the main reason of instability of blue fluorescent and phosphorescent OLEDs. Green and red emissive materials form lower-energy excited states, thus, acting as “energy sinks” for high-energy excited states of host and hole-transporting materials, and are less subject to excited state reactions (Kondakov et al. 2007, 2010; Kondakov 2008). The red and green emitters better compete with deep traps (e.g., nonradiative recombination centers) formed during degradation of charge carriers and recombination events than the blue ones. Additionally, high-energy blue emitters are more likely to be quenched by degradation products than green and red dopants. Thus, the blue OLEDs are much more sensitive to the presence of luminescence – quenching

Table 3 Commercial Universal PHOLED materials in bottom-emitting devices

PHOLED Performance (at 1000 cd/m ²)	1931 CIE Color Coordinates	Luminous Efficiency (cd/A)	Operating Lifetime (hrs)	
			LT 95%	LT 50%
DEEP RED	(0.69, 0.31)	17	14,000	250,000
RED	(0.66, 0.34)	29	23,000	600,000
RED	(0.64, 0.36)	30	50,000	900,000
YELLOW	(0.44, 0.54)	81	85,000	1,450,000
GREEN	(0.31, 0.63)	85	18,000	400,000
LIGHT BLUE	(0.18, 0.42)	50	700	20,000

The results are for bottom-emitting structures (with no cavities) fabricated by vacuum thermal evaporation. Lifetime data are based on accelerated current drive conditions at room temperature without any initial burn-in.

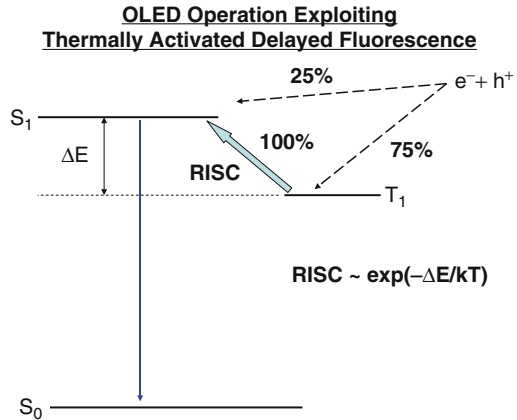
and charge-trapping degradation species – than green or red-emitting devices (Kondakov and Young 2010).

The development of sufficiently stable deep-blue phosphorescent OLEDs is a challenging task. In a phosphorescent OLED, the lowest triplet state (T_1) of the host is energetically above that of the dopant so that triplet excitons reside preferentially on the dopant. As the triplet energy defined as the difference in energy between the ground state and T_1 of the dopant increases, it becomes more difficult to find suitable stable hosts with even higher triplet energy levels. For typical molecules, the bandgap – the difference between the HOMO and LUMO energies – is larger or comparable to S_1 , the lowest singlet excited state. S_1 in turn is higher in energy than T_1 . This means that a molecule with a given triplet energy will usually have a larger bandgap than a molecule with an S_1 state of the same given energy. Thus, S_1 energy levels of phosphorescent blue host materials are considerably higher compared to that of fluorescent blue host materials. As was mentioned above, higher-energy materials degrade very fast. The blue component is the key element of white OLED architectures. Because of the absence of long-lived blue triplet emitters, white OLEDs for SSL often combine blue fluorescent emitters with longer-wavelength emitting phosphorescent dopants. As was previously mentioned, such architectures are called hybrid. White OLED for SSL can use sky-blue phosphorescent emitters, and such devices with all-phosphorescent emitters have been reported. As expected the devices show higher efficiency compared to hybrid white devices. However, it should be noted that in order to achieve practical lifetimes and acceptable color shift with aging in panels with all-phosphorescent materials, stacked architecture is used where the phosphorescent blue EML has to be placed adjacent to a lower-energy green or red EML. Data show that presently lifetime of hybrid white devices remains superior to that of all-phosphorescent-based devices.

A new class of fluorescent emitters that thermally promote their triplet states formed by recombination to the higher-energy excited singlet states by “reverse intersystem crossing” (RISC) is now the focus of OLED material research. This phenomenon, most commonly referred to as TADF or E-type delayed luminescence, is well known and was observed long ago, but interest in it was revived recently by the work of Adachi (Yoyama et al. 2012). The group reported developing a new class of materials where the RISC was shown to be nearly 100 % efficient and where green fluorescent OLEDs with EQE of 19.5 % were demonstrated. That work has been extended to similar efficiencies for blue devices in 2014 (Zhang et al. 2014). The general scheme by which highly efficient OLEDs can be realized in this way is illustrated in Fig. 5.

TADF emitters can have significant potential advantages over the current state-of-the-art phosphorescent materials in that reducing the triplet energy by several tenths of an eV enables the use of a wider variety of host materials and leaves less energy available for excited triplet reactions that have been shown to cause degradation of organic emitters. Furthermore, TADF materials can address the expense associated with the use of rare materials, such as iridium and platinum that are toxic and predominantly available from foreign sources.

Fig. 5 General mechanism for using TADF to circumvent limitations imposed on fluorescent OLEDs by spin statistics



Process Challenges for White OLEDs

White OLED architectures are often developed on small devices fabricated using laboratory equipment with the focus mainly on achieving superior performance. However, transferring the technology from the laboratory to the manufacturing stage presents a number of challenges, some of which are robustness of OLED device formulation with respect to fabrication tools and processes, thermal stability of materials, cost reduction, and many others. In this section we will focus on the difficulties associated with the manufacturing of white OLED lighting panels.

Stacked hybrid white OLED devices can bridge the performance gap between fluorescent and all-phosphorescent OLEDs as this architecture provides devices with high power efficiency and sufficiently long operational stability. As was described earlier, the hybrid devices incorporate a fluorescent blue dopant and phosphorescent emitters to complement blue emission and create white light. Today, deep-blue phosphorescent dopants with reasonable lifetime do not exist; however, sky-blue phosphorescent dopants are available. Figure 6 illustrates the range of white color temperatures that can be achieved with typical red and green dopants combined with either a deep-blue fluorescent dopant or a sky-blue phosphorescent dopant. Any color within the triangles is possible. With the deep-blue dopant, any correlated color temperature (CCT) along the Planckian locus from 1,000 K to 20,000 K is possible. With the sky-blue dopant, the color triangle is truncated such that any CCT above 4,000 K is not possible; however, warmer color temperatures from 4,000 K and below are still possible. The OLED light source with the sky-blue emitter can still have a high CRI, but it will not be able to render deep-blue and magenta colors.

To achieve warm-white color with CRI above 80, the phosphorescent stack of a white OLED usually contains two emitters – yellow and red. To increase the CRI further (CRI ~ 90), green and deep-red phosphorescent emitters can be included into the device structure. To increase the R9 value, which is important for skin tone

Fig. 6 Effective color temperatures possible with white OLED light sources containing either *sky-blue* or *deep-blue* emitters

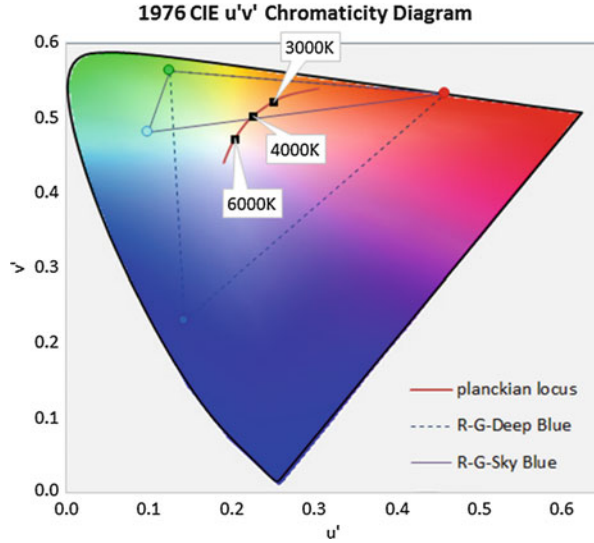
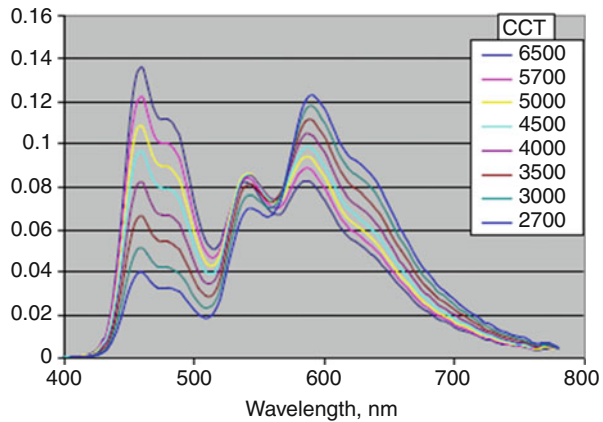


Fig. 7 Dependence of CCT of white OLED devices on their emission spectra (Reproduced from Tyan 2013)



reproduction, deeper red emission is required. It is important to note that the white color is very sensitive to the relative amount emitted from each stack (Tyan 2011, 2013). As shown in Fig. 7, a white device color temperature can be changed from 6,500 K to 2,700 K depending on contribution to emission from each emitting material. Yellow and red components of the white spectrum are particularly sensitive to the concentration of the emitters as well as formulation and placements of the light-emitting layers. Therefore, in order to obtain a required white color, precise control of material ratios is necessary which might not be possible to achieve with certain material combinations and device architectures on a manufacturing line.

An example of a two-stack hybrid white device is shown in Fig. 8 (Hatwar et al. 2010). The device contains 11 organic layers and three emitters. A fluorescent

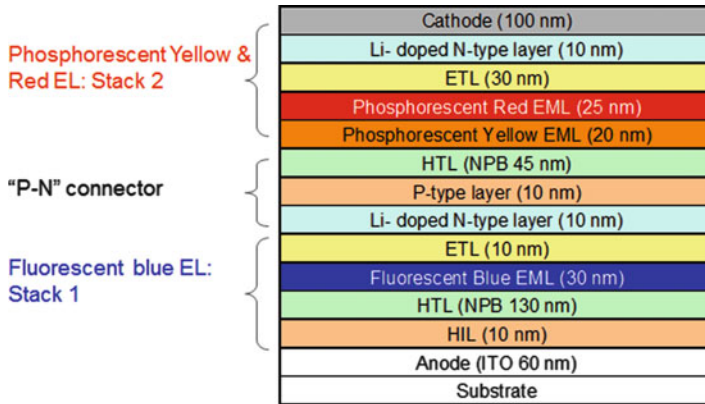


Fig. 8 Schematic of the hybrid tandem for SSL (Reproduced from Hatwar et al. 2010)

blue stack was placed near the anode. Emission of warm-white color was achieved by the placement of a red-emitting layer adjacent the yellow-emitting layer in the phosphorescent stack. Similar results were also obtained by co-doping of the EML with both yellow and red emitters at concentrations that balance the ratio of red to yellow emission. In the latter case, the concentration of the red emitter should be kept very low; otherwise, yellow emission significantly decreases resulting in unacceptable hue of white color.

In a stacked device, the charge generation layer (CGL), or connector layer, is of critical importance to achieve low voltage, high efficiency, and long lifetime. An effective and manufacturable CGL that was developed by Kodak includes a Li-doped N-type layer and an electron-accepting p-type layer (Liao et al. 2004). The lithium is co-doped into an electron-transporting host at a low concentration in the 1–4 % range. Alternately, Li or another metal can be deposited as a very thin pure layer typically around 0.5 nm thick. It can be challenging to uniformly deposit such a thin layer, so co-doping the metal throughout a thicker organic layer can provide more process tolerance. An alternate non-Li connector was developed by Kodak and Novaled which consists of a p-doped layer and an n-doped layer with a buffer layer in between (Hatwar et al. 2009). Stable and repeatable production of this type of CGL was first demonstrated in a Gen2 pilot manufacturing system by the Fraunhofer IPMS group (Eritt et al. 2010).

Light extraction technology is one of the most effective methods to improve the power efficacy of OLED devices. The internally generated light is distributed between several different modes, and details of the emission loss and light-trapping mechanism have been described elsewhere (Saxena et al. 2009; Meerheim et al. 2010). Light extraction methods can essentially be divided into two types of structures: external and internal. An external extraction layer (EEL) is applied on the outside emitting surface of the glass substrate, and the internal extraction layer (IEL) is inserted between the substrate and the cathode, usually between the substrate and the transparent anode. The EEL is formed after the OLED is complete and is simple

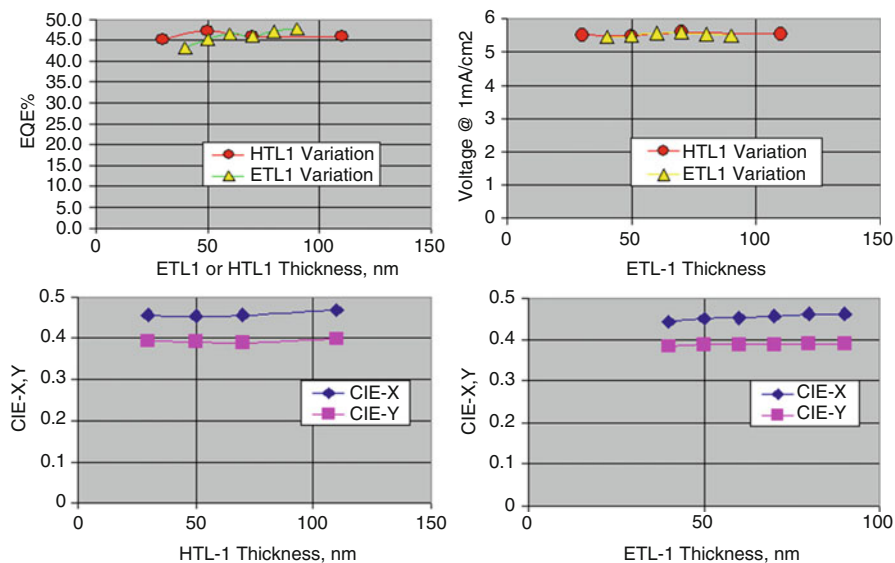


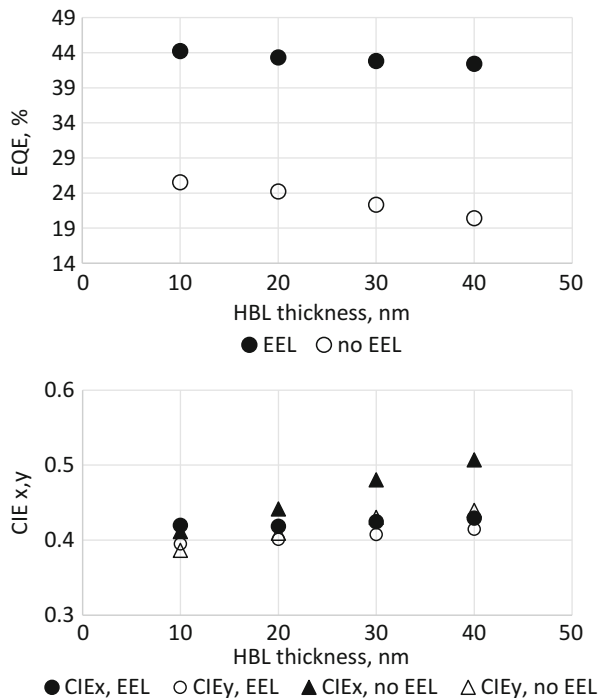
Fig. 9 Dependence of EQE, voltage, and CIE color coordinates on thickness of ETL-1 and HTL-1 (Reproduced from Tyan et al. 2009)

to apply. IEL technology is more effective than EEL, since it can access more of the trapped light inside the OLED device. However, it is much more difficult to implement, as it can damage the organic layers of the OLED. If the IEL surface is too rough, microshorts, leakage, or nonuniformities can occur, and if any residual moisture, solvent, or other materials outgas, chemical degradation of OLED materials can occur.

The white color of an OLED can also be affected by the use of light extraction layers. Tyan et al. have shown that one advantage of the IEL is the improved tolerance of device parameters to layer thickness variation. As shown in Fig. 9, efficiency, color, and voltage of a hybrid white 2-stack OLED are practically unchanged by variation of the hole transport layer (HTL-1) and the electron-transport layer (ETL-1) both used in the phosphorescent stack. The phosphorescent stack is placed near the cathode and blue fluorescent stack is near anode. Thickness variation of both HTL-1 and ETL-1 changes distance between emitters and the reflector (cathode). It is known that light generated inside the OLED spreads through 4 modes, such as air mode, substrate, wave guided, and surface plasmon mode. Light modulation in all modes is caused by optical interference between the directly emitted light and reflected light and, therefore, depends strongly on the distance from the emission zone to the cathode. Insensitivity of emission of devices with IEL to this distance can result in improved fabrication yields.

EEL enhances efficiency by redistributing light trapped in the substrate mode to the air mode. The distribution of light in the various modes depends on the device structure. High extraction efficiency can be obtained if the device structure is

Fig. 10 Dependence of EQE and CIE color coordinates on thickness of a hole-blocking layer in the blue stack of white device with and without EEL. Device structure is ITO/HIL1/HTL1/fluorescent BEML/HBL/EIL1/HIL2/HTL-2/phosphorescent EML/ETL/EIL2/Ag



designed for this purpose. Figure 10 shows the dependence of EQE and color in a hybrid white device with and without EEL. The fluorescent blue stack includes a hole-blocking layer (HBL) placed between the blue EML and ETL. Variation of the hole-blocking layer thickness changes the distance between the blue emitter and the cathode. As the results show, EQE decreases as HBL thickens in devices without EEL. Color coordinates show noticeable change with the thickness as well. The use of EEL reduces variation in EQE and color change. Extraction efficiency varies depending on distance between blue EML to cathode; light is extracted most efficiently in the device with 40 nm of HBL.

Even with light extraction layers, the white structure still needs to be tuned to achieve the best trade-offs between efficiency and color. Figure 11 shows the variation in CRI (both R_a and R_9) and efficacy as a function of the HTL thickness nearest the anode in a 3-stack white device with EEL. CRI R_a can vary between 80 and 90, while R_9 can vary between 4 and 42 over the same range of HTL thickness. The highest R_9 values are achieved when the emission spectrum broadens in the deep-red wavelength region. The highest efficacy is achieved at the lowest CRI values.

Today's warm-white OLED panels use 3-stack architectures to achieve practical lifetimes at $3,000 \text{ cd/m}^2$ (Moon et al. 2013), or even 6-stack architectures that can achieve good lifetime at very high brightness above $8,000 \text{ cd/m}^2$ (Dotter 2015). Using multi-stacked architectures increases the level of production complexity and

Fig. 11 Efficacy and *CRI* as a function of HTL thickness

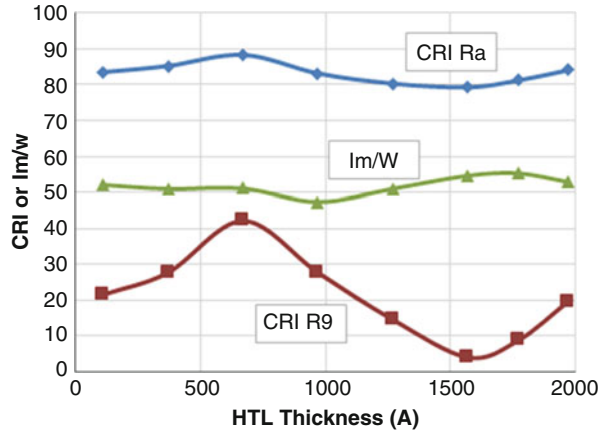
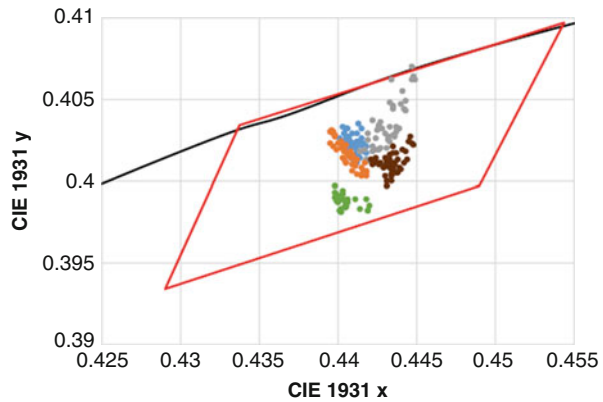


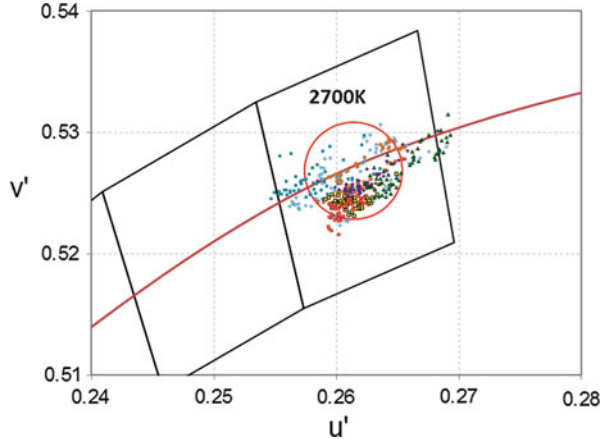
Fig. 12 White OLED color reproducibility over several production runs (Reproduced from Dotter 2015)



cost; however, it is currently the most effective method to achieve long operating lifetimes. To minimize color shift over the life of the OLED light source, each emitting unit should have approximately the same luminance decay characteristics. Each emitting unit is usually first constructed as a single-stack device in order to optimize the performance, before placing multiple units into a stacked configuration. Each unit should be characterized at the same current density at which the final stacked device is expected to operate.

In the manufacturing of white OLED panels, one of the largest yield loss factors is due to white color-point variations between panels. LEDs have long had color-point variation and have been classified in narrow color-range bins where the color range is imperceptible. In LED lights that contain multiple LEDs, LEDs from many bins can be combined to achieve an overall illuminant color that is within the tight lamp-to-lamp color specification. With OLED panels, however, the panels are in direct view, and it is not possible to combine panels with different white color points without the customer noticing. Thus, for high OLED manufacturing yield, it is important that the variability of the white color point fits within one “bin.” Figures 12

Fig. 13 White OLED color reproducibility over several production runs (OLEDWorks data). Shown is the 7-step Energy Star quadrangle for 2,700 K, along with a delta u'/v' circle with radius 0.004



and 13 show the variation in color point seen in the Philips production machine and the OLEDWorks production machine, respectively. It is important to note that this level of color reproducibility has been achieved only after many years of effort and improvements to the equipment, process, and materials.

In general, all layers need to be controlled to within 5 % of the intended thickness. To keep a very tight distribution of white color, some layers, particularly the emission layers, need to be controlled within 1–3 % of the target. A major factor causing the white color-point variation is the ability to control the composition of the emitter layers, especially when there is a high sensitivity to the level of a dopant which is present at a low concentration of less than 2 %. This is caused by the difficulty in accurately and precisely measuring and controlling evaporation rates at very low rates. The root of this problem is due to the general-purpose nature of the vapor generation sources for multicomponent layers combined with the limitations of the QCM rate monitoring methods.

Multicomponent layer sources must be designed so that the minor components can be in the range of 1–50 % of the total layer. However, this dynamic range is larger than practical for QCM systems. A practical maximum deposition rate on a QCM crystal is 10 Å/s. At that rate, a crystal may have a lifetime of less than one day – which may be acceptable if the machine holds a large number of spare quartz crystals. Controlling the rate to within $\pm 5\%$ of the target level is no problem under this condition. If this is the design for a 50 % dopant, then such a system will operate with a deposition rate of 0.2 Å/s on the QCM crystal when used for a 1 % dopant. At these low rates, QCM noise and precision limitations become significant factors. The noise can be removed by filtering; however, the filtering introduces lag or delay into the measurement signal. As the lag increases due to the increased filtering, the control system must be detuned to prevent it from becoming unstable, which in turn diminishes the system’s ability to keep the error within the allowed limit. Many vendors are working to improve the precision and reduce the noise on QCM system. With the use of “in situ optical process control capability,” Philips has demonstrated

color-point yield of >95 % (Dotter 2015). Another approach to this problem is to develop robust-for-manufacture device formulation with high concentration of emitting components.

Mass production of OLED devices requires high-processing temperatures to achieve high throughput and high yield. The OLED materials are exposed to high temperatures in the linear sources and other parts of the evaporation systems, and often the temperature is held well above the evaporation temperature of the material to prevent condensation anywhere within the vapor generation system. This can cause thermal degradation and decomposition of sensitive materials; therefore, material manufacturers strive to develop organic materials with high thermal stability (Murano et al. 2014). Materials with a larger gap between their evaporation temperature and decomposition temperature are generally more robust in a mass production environment.

Yield and Reliability

Like any emerging technology, OLED lighting has yield and reliability challenges that must be overcome. The major AMOLED display manufacturers claim to have greater than 90 % yield. OLED lighting manufacturers claim to have yields of 50–80 % (Dotter 2015), so there is still much room for improvement. Because OLED lighting panels are essentially one large pixel, unlike a display which can have millions of tiny pixels, they are susceptible to shorting and reliability issues (Park et al. 2011). The organic layer stack is the only means of insulation between the conductive electrodes. The total organic stack thickness is typically only 200–400 nm, so a small particle can easily cause a nonuniformity in the organic coating or even a void where shadowing may occur along an edge of a particle. The void or thin area creates an opportunity for a direct short circuit of the metal cathode to the anode or at least a current leakage pathway due to a higher electric field. Since a leakage path will have a locally higher current density, greater heating occurs which accelerates the thermal degradation of the OLED (Lee et al. 2007). When the local temperature exceeds the sublimation temperature of the organic materials, gas bubbles can form underneath the cathode and cause the cathode metal to stretch or even explode (Kolosov et al. 2001). Sometimes a leakage path can heal itself, but most likely it will progress into a greater leakage path and eventually a short circuit will occur. Thicker organic stacks can improve the situation, and often the hole-transporting layer (HTL) is chosen to be made thicker since it is a single component layer and generally less expensive material. Multi-stack architectures naturally make the stack thicker. Some manufacturers use a solution-coated hole-injection layer (HIL) to planarize defects and improve yield (Hamer 2015). The use of a high-resistance short-reduction layer (SRL) has also been shown to be effective in limiting the current that can flow through the leakage path (Tyan et al. 2007). Particles can occur anywhere in the manufacturing process. One of the largest sources of particles is the organic deposition chambers, where material builds up on masks, shields, and chamber walls and can flake off over time. Periodic maintenance is important to thoroughly clean the deposition chambers. Masks are

typically cleaned more often, and it is a common practice to have multiple sets of masks that can be cleaned and rotated into the production process.

Additional yield loss can come in the form of visual defects, such as bright or dark spots, scratches, chipped glass, or brightness nonuniformities. Small dark spots less than 100–200 μm can usually be tolerated, as they are likely to be hidden if a diffusive external extraction film is used. However, if larger dark spots are visible, it may be an indication that the integrity of the encapsulation has been compromised. The encapsulation system is designed for a storage life of 10–20 years and usually tested through accelerated heat and humidity conditions like 85C/85 %RH. This represents nearly a $100\times$ acceleration factor (Dotter 2015). However, defects in the seal quality or thin film encapsulation quality can result in a short time to failure.

The other important category of yield loss is associated with accuracy and consistency of the white color (Hamer 2015). This is especially important for applications where multiple OLED panels will be tiled into a larger luminaire, and minor color differences between panels will be easily visible. Not only does the initial color have to be correct, but it cannot change much over the life of the product. If panels have to be replaced in a multi-panel luminaire, the replacements cannot look different than the originals. Typical specifications for color groupings are within a radius of 0.002 $\Delta u'v'$ or less initially, and color cannot shift more than 0.004 over life. Product specifications for electrical performance usually require the voltage and luminous flux to be within 5–10 % of a nominal value.

The operational lifetime of an OLED lighting panel is determined by the OLED materials and architecture and by the operating conditions. Today's high-efficiency stacked white OLEDs can produce 3,000 cd/m^2 at a current density of 2 mA/cm^2 or less, with lifetimes in the range of 20,000–40,000 h to T70. Reducing the drive current by a factor of two results in an increase in lifetime by approximately a factor of three. Lower heat generation and better heat removal can further increase the lifetime. Lifetime measurements are typically done on a sampling basis, since it is impractical to measure every panel. Lifetime tests are usually accelerated to some degree but cannot be accelerated too much on large-area panels due to extreme heating. Methods have been developed to indirectly assess lifetime (Pang et al. 2014). Until OLED reliability can achieve lower than 100 ppm failure rate, it will likely be necessary to perform burn-in operations to eliminate early failures.

Future Directions

For OLED lighting to become a large market, the US DOE has suggested that the panel manufacturing costs have to continue to decrease to about $\$100/\text{m}^2$ in 2025 or about $\$1$ for a 100 cm^2 size panel. In 2015, OLED lighting panels are being sold for $\$200/\text{klm}$ (LG Chem and Philips) or the equivalent of $\$2,000/\text{m}^2$ at a brightness of 10 klm/m^2 . These are not sustainable prices at current sales volumes, indicating that manufacturing costs must continue to fall by at least an order of magnitude over the next decade. At the same time, panel efficacy is expected to rise to 190 lm/W with LT70 lifetimes reaching 50 kh (US DOE 2014). Cost reduction trends that have

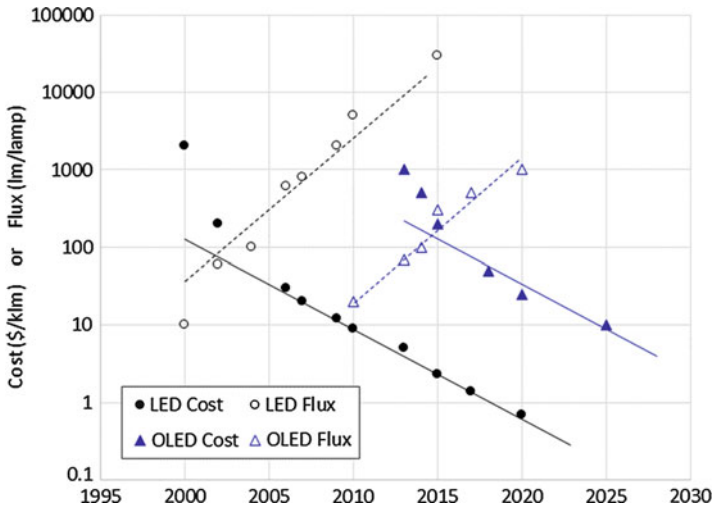


Fig. 14 SSL trends for flux/lamp and cost per thousand lumens

occurred with LED SSL over the past decade should also apply to OLED SSL. Figure 14 shows a graphical extension of Haitz' Law that includes both LED and OLED SSL lighting. This "law" was an observation of the historical SSL trends showing a $10\times$ cost reduction per decade and a $20\times$ flux increase per decade and was first published in 1999 then updated in 2010 (Haitz and Tsao 2011). It is important to note that this is the LED package cost, not the lamp. In 2015, a 60 W Cree A19 LED lamp that produced 800 lm cost \$10, resulting in a lamp cost of \$12.50/klm. An OLED panel is essentially the lamp; in 2015 OLED panels cost around \$200/klm. Assuming that OLED cost trends parallel historical LED trends, OLEDs should reach the cost target of \$10/klm around 2025.

Achieving these cost and performance goals will require improvements in fundamental OLED stack design (e.g., long-lifetime blue phosphorescent emitter systems), great improvements in light outcoupling, and improvements in manufacturing processes and equipment. This will require a combination of evolutionary improvements of existing materials and processes and revolutionary introduction of new ideas. One of the critical components is the need to practice manufacturing and sales – both to help define what the customers want and to help define OLED lighting's best market opportunities – and to provide opportunity for manufacturing equipment and production process development in order to reduce costs by reducing TAC times, increasing yields, and reducing the capital required to reduce depreciation costs.

Low costs cannot be achieved without high yields. Today's color variation can be reduced by several approaches: new sensor approaches, multicomponent layer formulations that are less sensitive to dopant concentration variations, or the elimination of the dopant entirely (Wang et al. 2014). Customer satisfaction cannot be achieved without excellent reliability on the order of 1 ppm failure rate. This may require performance along several fronts: improved short-reduction layers, improved methods

to detect nascent weaknesses within a panel, improved formulations to impede short circuit development, and improved production methods and equipment to reduce particles associated with the defects that can form during operation in the field.

Since OLED SSL cost and performance will continue to lag behind LED SSL cost and performance, it will be important to find the right applications that allow the OLED market to grow, while costs are still somewhat high. One application where OLED may differentiate from LED is in the form of flexible OLED lighting, which may open up a multitude of new market opportunities. To that end, technologies that enable flexible lighting will be important, including flexible substrates, transparent flexible conductors, lower cost patterning and printing technologies, and barrier layer and encapsulation technologies.

References

- Adamovich VI, Levermore PA, Xu X, Dyatkin AB, Elshenawy Z, Weaver MS, Brown JJ (2012) High-performance phosphorescent white-stacked organic light-emitting devices for solid-state lighting. *J Photonics Energy* 2(1):021202
- Baldo MA, Forrest SR (2000) Transient analysis of organic electrophosphorescence: I. Transient analysis of triplet energy transfer. *Phys Rev B* 62:10958
- Baldo MA, O'Brien DF, Thompson ME et al (1999) Excitonic singlet-triplet ratio in a semiconducting organic thin film. *Phys Rev B* 60:14442
- Bergemann KJ, Krasny R, Forrest SR (2012) Thermal properties of organic light-emitting diodes. *Org Electron* 13:1565–1568
- Birnstock J, He G, Murano S et al (2008) White stacked OLED with 35 lm/W and 100,000 hours lifetime at 1000 cd/m² for display and lighting application. *SID Int Symp Digest* 39:822
- Brutting W, Frischeisen J (2012) Device efficiency of organic light-emitting diodes. In: Brutting W, Adachi C (eds) *Physics of organic semiconductors*. Wiley-VCH, Weinheim
- Burrows PE, Forrest SR, Sibley SP, Thompson ME (1996) Color-tunable organic light-emitting devices. *Appl Phys Lett* 69:2959
- Cooper G (2015) Light extraction for OLEDs. Presentation at 2015 DOE solid-state lighting manufacturing R&D workshop, San Francisco. http://www.energy.gov/sites/prod/files/2015/02/f19/cooper_oled-integration_sanfrancisco2015.pdf
- Dotter W (2015) OLED lighting technology and applications. Presented at lighting Japan 2015
- Duggal AR, Foust DF, Nealon WF, Heller CM (2004) OLEDs for lighting: new approaches. *Proc SPIE* 5214:241–247
- Eritt M, Toerker M, Jahnel M, May C, Leo K (2010) Up-scaling of OLED manufacturing for lighting applications. *SID Int Symp Digest* 41:699–702
- Garditz C, Winnaker A, Schindler F, Paetzold R (2007) Impact of Joule heating on the brightness homogeneity of organic light emitting devices. *Appl Phys Lett* 90:103506
- Gibson G, Snodgrass S (2011) Selective coating and removal technologies to produce patterned films for printed electronics. *Proc LOPE-C* 2011:31–35
- Haitz R, Tsao JY (2011) Solid-state lighting: 'the case' 10 years after and future prospects. *Phys Status Solidi A* 208(1):17–29
- Hamer J (2015) Problems and opportunities in OLED lighting manufacturing. Presentation at 2015 DOE solid-state lighting R&D workshop, San Francisco. http://www.energy.gov/sites/prod/files/2015/02/f19/hamer_oled-mfg_sanfrancisco2015.pdf
- Hamer J, Yamamoto A, Rajeswaran G, Van Slyke SA (2005) Mass production of full-color AMOLED displays. *SID Int Symp Digest* 36:1902–1907
- Han CW, Park JS, Shin YH, Lim MJ, Kim BC, Tak YH, Ahn BC (2014) Advanced technologies for large-sized OLED TV. *SID Int Symp Digest* 45:770–773

- Hatwar TK, Spindler JP, Begley WJ, Giesen DJ, Kondakov DY, Van Slyke S, Murano S, Kucur E, He G, Blochwitz-Nimoth J (2009) High-performance tandem white OLEDs using a Li-free “P-N” connector. *SID Int Symp Digest* 40:499–502
- Hatwar TK, Spindler JP, Kondakova M, Giesen D, Deaton J, Vargas JR (2010) Hybrid tandem white OLEDs with high efficiency and long life-time for AMOLED displays and solid-state lighting. *SID Int Symp Digest* 41:778
- Heil H, Rodriguez L, Burkhart B, Meyer S, Riedmueller S, Darsy A, Pflumm C, Buchholz H, Boehm E (2014) High-performance OLED materials. *SID Int Symp Digest* 44:495
- Hong S, Jeon C, Song S, Kim J, Lee J, Kim D, Jeong S, Nam H, Lee J, Yang W, Park S, Tak Y, Ryu J, Kim C, Ahn B, Yeo S (2014) Development of commercial flexible AMOLEDs. *SID Int Symp Digest* 45:334–337
- http://apps1.eere.energy.gov/buildings/publications/pdfs/ssl/ssl_mypp2014_web.pdf
- http://en.wikipedia.org/wiki/Amber_%28color%29#UNECE_Amber. Accessed 03 Feb 2015
- <http://www.lgchem.com/global/green-energy/oled-lighting>. Accessed 03 Mar 2015
- <http://www.udcoled.com/default.asp?contentID=604>. Accessed 03 Dec 2015
- <http://www.verbatimlighting.com/article/oled/>. Accessed 03 May 2015
- https://www.energystar.gov/ia/partners/prod_development/new_specs/downloads/SSL_FinalCriteria.pdf
- Huang CC, Meng HF, Ho GK, Chen CH, Hsu CS, Huang JH, Horng SF, Chen BX, Chen LS (2004) Color-tunable multilayer light emitting diodes based on conjugated polymers. *Appl Phys Lett* 84(7):1195
- Hung CH, O’Shaughnessy D, Ganjoo A, McCamy J, Bhandari A (2014) Large-area integrated substrate for OLED lighting. Presentation at 2014 DOE solid-state lighting workshop, San Diego. http://apps1.eere.energy.gov/buildings/publications/pdfs/ssl/hung_integrated-substrate_sandiego2014.pdf
- Hunt RWG (1988) *Measuring color*, 3rd edn. Fountain Press, London
- Jou JH, Wang HC, Shen SM, Peng SH, Wu MH, Chen SH, Wu PH (2012) Highly efficient color temperature tunable organic light-emitting diodes. *J Mater Chem* 22:8117
- Kalinowski J (1999) Electroluminescence on organics. *Phys D Appl Phys* 32:R179
- Kalinowski J, Di Marco P, Cocchi M, Fattori V, Camaioni N, Duff J (1996) Voltage-tunable-color multilayer organic light emitting diode. *Appl Phys Lett* 68(17):2317
- Kobrin B (2014) Advanced manufacturing of nanostructured transparent conductors and light extraction structures for OLED lighting devices. Presentation at 2014 DOE solid-state lighting manufacturing R&D workshop, San Diego. http://apps1.eere.energy.gov/buildings/publications/pdfs/ssl/kobrin_nanoweb_sandiego2014.pdf
- Kohnen A, Meerholz K, Hagemann M, Brinkmann M, Sinzinger S (2008) Simultaneous color and luminance control of organic light-emitting diodes for mood lighting applications. *Appl Phys Lett* 92:033305
- Kolosov D, English DS, Bulovic V, Barbara PF, Forrest SR, Thompson ME (2001) Direct observation of structural changes in organic light emitting devices during degradation. *J Appl Phys* 90(7):3242–3247
- Komoda T (2011) High quality white OLEDs and resource saving fabrication: processes for lighting application. Presentation at 2011 printed electronics and photovoltaics, Dusseldorf
- Kondakov DY (2008) Role of chemical reactions of arylamine hole transport materials in operational degradation of organic light-emitting diodes. *J Appl Phys* 104:084520
- Kondakov DY, Young RH (2010) Variable sensitivity of organic light-emitting diodes to operation-induced chemical degradation: nature of the antagonistic relationship between lifetime and efficiency. *J Appl Phys* 108:074513
- Kondakov DY, Lenhart WC, Nichols W (2007) Operational degradation of organic light-emitting diodes: mechanism and identification of chemical products. *J Appl Phys* 101:024512
- Kondakov DY, Brown CT, Pawlik TD et al (2010) Chemical reactivity of aromatic hydrocarbons and operational degradation of organic light-emitting diodes. *J Appl Phys* 107:024507

- Kondakova ME, Young RH, Prosperi DA, Miller RL, Comfort DL (2010) Effect of organic hole-injecting buffer layer on stability of organic light-emitting diodes. *SID Int Symp Digest* 41:1808–1811
- Lecamp G (2013) Light outcoupling for OLED: double the efficiency while keeping the dark current low. *SID Int Symp Digest* 44:597–599
- Lee YJ, Lee H, Byun Y, Song S, Kim JE, Eom D, Cha W, Park SS, Kim J, Kim H (2007) Study of thermal degradation of organic light emitting device structures by X-ray scattering. *Thin Solid Films* 515:5674–5677
- Levermore PA, Dyatkin AB, Elshenawy ZM, Pang H, Kwong RC, Ma R, Weaver MS, Brown JJ (2011a) Phosphorescent OLEDs: enabling solid state lighting with lower temperature and longer lifetime. *SID Int Symp Digest* 42:1060–1063
- Levermore PA, Pang H, Dyatkin AB, Elshenawy Z, Mandlik P, Rajan K, Silvernail J, Kwong RC, Ma R, Weaver MS, Hack M, Brown JJ (2011b) Phosphorescent OLEDs: enabling energy-efficient lighting with improved uniformity and longer lifetime. *J Soc Inf Disp* 19(12):943–949
- Liang CJ, Choy WCH (2009) Tunable full-color emission of two-unit stacked organic light emitting diodes with dual-metal intermediate electrode. *J Organomet Chem* 694:2712
- Liao LS, Klubek KP, Tang CW (2004) High-efficiency tandem organic light-emitting diodes. *Appl Phys Lett* 84(2):167–169
- Lu MH (2013) Outcoupling efficiency enhancement strategies in OLED lighting panels. *SID Int Symp Digest* 44:912–915
- Ma R, Levermore P, Pang H, Mandlik P, Rajan K, Silvernail J, Hack M, Brown JJ (2011) Challenges and opportunities in scaling up OLED lighting devices. *SID Int Symp Digest* 42:983–986
- Matsumoto T, Nakada J, Endo N, Mori A, Kavamura K, Yokoi A, Kido J (2003) Multiphoton organic EL device having charge generation layer. *SID Int Symp Digest* 34:979
- Meerheim R, Furno M, Hofmann S, Lussem B, Leo K (2010) Quantification of energy-loss mechanisms in organic light-emitting diodes. *Appl Phys Lett* 97:253305
- Moon J, Joo M, Lee Y, Lee J, Park M, Choi J, Ham Y, Ahn Y, Kim J, Lee J, Kim J, You J, Wonik J, Kim J, Son S (2013) 80 lm/W white OLEDs for solid state lighting. *SID Int Symp Digest* 44:842–844
- Murano S, Gilge K, Ammann M, Werner A (2014) AMOLED manufacturing – challenges and solutions from a material makers perspective. *SID Int Symp Digest* 45:403–406
- Nakamura N, Domercq B, Billet S, Roquiny P, Wada N, Fukumoto N, Tanida M, Aoki Y, Ohgawara M (2013) Advanced glass substrate for the enhancement of OLED lighting out-coupling efficiency. *SID Int Symp Digest* 44:803–806
- Neyts K, Marescaux M, Nieto U, Elschner A, Lovenich W, Fehse K, Huang Q, Walzer K, Leo K (2006) Inhomogeneous luminance in organic light emitting diodes related to electrode resistivity. *J Appl Phys* 100:114513
- Neyts K, Real A, Marescaux M, Mladenovski S, Beekman J (2008) Conductor grid optimization for luminance loss reduction in organic light emitting diodes. *J Appl Phys* 103:093113
- Nishimura K, Kawamura M, Jinde Y et al (2008) The improvement of white OLED's performance. *SID Int Symp Digest* 39:1971
- Ogiwara T, Ito H, Mizuki Y, Naraoka R, Funahashi M, Kuma H (2013) Efficiency improvement of fluorescent blue device by molecular orientation of blue dopant. *SID Int Symp Digest* 44:515
- Ohno Y (2014) Practical use and calculation of CCT and Duv. *Leukos* 10(1):47–55
- Pang H, Michalski L, Weaver MS, Ma R, Brown JJ (2014) Thermal behavior and indirect life test of large-area OLED lighting panels. *J Solid State Light* 1:7
- Park J, Lee J, Shin D, Park S (2009) Luminance uniformity of large-area OLEDs with an auxiliary metal electrode. *J Disp Tech* 5(8):306–311
- Park JW, Shin DC, Park SH (2011) Large-area OLED lightings and their applications. *Semicond Sci Tech* 26:034002
- Park J, Lee J, Noh YY (2012) Optical and thermal properties of large-area OLED lightings with metallic grids. *Org Electron* 13:184–194

- Reineke S, Thomschke M, Lussem B, Leo K (2013) White organic light-emitting diodes: status and perspective. *Rev Mod Phys* 8(3)
- Roberts M, Akino N, Asada K, Benzie P, Hamamatsu H, Hatcher M, King S, Snedden E, Strevens A, Tanaka S, Toner J, Wilson R, Young K, Yamamoto K, Yamada T (2013) High efficiency polymer OLEDs – analysis and progress. Presentation at SPIE 2013 optics and photonics, <https://www.cdtltd.co.uk/pdf/spie2013-mroberts-cdt.pdf>. Accessed 03 Dec 2015
- Saxena K, Jain VK, Mehta DS (2009) A review on the light extraction techniques in organic electroluminescent devices. *Opt Mater* 32:221–233
- Shen Z, Burrows PE, Bulovic V, Forrest SR, Thompson ME (1997) Three-color, tunable, organic light-emitting devices. *Science* 276:2009
- Slafer WD (2014) Developing enhanced substrate for OLED SSL. Presentation at 2014 DOE Solid-State Lighting R&D Manufacturing Workshop, San Diego. http://apps1.eere.energy.gov/buildings/publications/pdfs/ssl/slafer_olesubstrates_sandiego2014.pdf
- Spindler J, Govindarajan VS, Newman D, Rajeswaran G (2011) Packaging Technologies for OLED Displays and Lighting Products. *Proc SMTA Intl 2011*, Fort Worth Tx
- Tang CW, VanSlyke SA (1987) Organic electroluminescent diode. *Appl Phys Lett* 51(12):913
- Taylor M (2015) Integrated OLED substrates. Presentation at 2015 DOE Solid-State Lighting R&D Workshop, San Francisco. http://energy.gov/sites/prod/files/2015/02/f19/chowdhury-taylor_substrates_sanfrancisco2015_0.pdf
- Tyan YS (2011) Organic light-emitting-diode lighting overview. *J Photonics Energy* 1(1):011009
- Tyan YS (2013) OLED lighting. Seminar M-11, presented at 2013 SID conference, Vancouver
- Tyan YS, Farruggia G, Cushman TR (2007) Reduction of shorting defects in OLED devices. *SID Int Symp Digest* 38:845–848
- Tyan YS, Rao Y, Ren X, Kesel R, Cushman T, Begley W, Bhandari N (2009) Tandem hybrid white OLED devices with improved light extraction. *SID Int Symp Digest* 40:895
- US Department of Energy Solid-State Lighting Research and Development Multi-Year Program Plan (2014) Prepared by Bardsley Consulting, SB Consulting, SSLS Inc, Navigant Consulting, Radcliffe Advisors, Washington, DC
- Wang Q, Oswald IWH, Perez MR, Jia H, Shahub AA, Qiao Q, Gnade BE, Omary MA (2014) Doping-free organic light-emitting diodes with very high power efficiency, simple device structure, and superior spectral performance. *Adv Funct Mater* 24:4746–4752
- Weaver M, Xu X, Pang H, Ma R, Brown J (2014) Color tunable phosphorescent white OLED lighting panel. *SID Int Symp Digest* 45:672
- Xu X, Weaver MS, Brown JJ (2013) Phosphorescent stacked OLEDs for warm white light applications. *SID Int Symp Digest* 44:845
- Yamae K, Tsuji H, Kittichungchit V, Ide N, Komoda T (2013) Highly efficient white OLEDs with over 100 lm/W for general lighting. *SID Int Symp Digest* 44:916
- Yersin H, Finkenzeller WJ (2008) Triplet emitters for organic light-emitting diodes: basic properties. In: Yersin H (ed) *Highly efficient OLEDs with phosphorescent materials*. Wiley-VCH, Weinheim, pp 1–97
- Yersin H, Rausch AF, Czerwieńiec R (2012) Organometallic emitters for OLEDs: triplet harvesting, singlet harvesting, case structures, and trend. In: Brutting W, Adachi C (eds) *Physics of organic semiconductors*. Wiley-VCH, Weinheim
- Yoyama H, Goushi K, Shizu K, Nomura H, Adachi C (2012) Highly efficient organic light-emitting diodes from delayed fluorescence. *Nature* 492:234
- Zhang Q, Li B, Huang S, Nomura H, Tanaka H, Adachi A (2014) Efficient blue organic light-emitting diodes employing thermally activated delayed fluorescence. *Nat Photonics* 8:326–332
- Zheng T, Choy WCH (2008) An effective intermediate Al/Au electrode for stacked color-tunable organic light emitting devices. *Proc SPIE* 69992:69992R
- Zhou X, He J, Liao LS, Lu M, Ding XM, Hou XY, Zhang M, He XQ, Lee ST (2000) Real-time observation of temperature rise and thermal breakdown processes in organic LEDs using an IR imaging and analysis system. *Adv Mater* 12(4):265–269

OLED Manufacturing Equipment and Methods

Jeffrey P. Spindler, John W. Hamer, and Marina E. Kondakova

Contents

Introduction	418
Vacuum Thermal Evaporation	418
Alternate Vapor Deposition Methods	421
Alternate Thermal Transfer Methods	424
VTE Equipment Configurations	426
Research Systems	426
Production Systems	427
Future Production Systems	429
Organic Source Configurations	429
Point Sources	429
Linear Sources	430
Metal Sources	433
VTE Equipment Productivity	435
VTE Equipment Maintenance	438
Future Directions	439
References	440

Abstract

OLED manufacturing can be classified into two categories: dry and wet methods. The dry manufacturing method refers to the conversion of raw organic materials from a solid powder form into a gas phase. In this method, the powder is heated to above its sublimation temperature to form a vapor which then condenses onto a substrate, all while in a high-vacuum environment. The wet method refers to the application of organic materials dissolved in a solution or a condensed liquid phase. In this wet form, the solution is applied to the substrate by spin coating, slit coating, inkjet printing, or other methods and then dried before applying the next layer. In

J.P. Spindler (✉) • J.W. Hamer • M.E. Kondakova
OLEDWorks LLC, Rochester, NY, USA
e-mail: jspindler@oledworks.com

both methods, the organic material is finally condensed onto a substrate in a solid-phase form. This chapter describes the methods and equipment typically used to manufacture OLED devices, with a focus on dry manufacturing methods such as vacuum thermal evaporation (VTE) and other vapor deposition techniques.

Introduction

Currently, the vast majority of commercial OLED displays and lighting panels are produced using vacuum thermal evaporation. The reason for this is simply that VTE produces OLEDs with the best performance in terms of efficiency and lifetime, in addition to being the most mature manufacturing method. Other methods such as solution processing hold the promise of lower manufacturing costs due to easier scalability and better material utilization; however, none have achieved commercial scale yet despite many years of development. Much progress has been made over the last 10 years to reduce the performance gap between VTE and non-VTE methods; however, non-VTE approaches typically lag behind in at least one or more performance attributes, such as lifetime. One simple explanation is that with VTE, the fabrication process completely occurs under a high-vacuum environment, typically in the range of 10^{-7} torr. With other methods, fabrication occurs under a lower-vacuum or non-vacuum environment, at least in part. Despite great efforts to maintain an inert environment such as a nitrogen-purged dry box, the non-VTE approaches provide a greater opportunity for water to enter into the organic thin films during formation. Incorporation of water and oxygen, even at the part-per-million level, is known to have detrimental effects on OLED device performance (Knox and Halls 2006). Several studies have shown that OLED device performance, particularly lifetime, is well correlated with the baseline pressure of the vacuum system in which the organic layers are deposited (Bohler and Dirr 1997; Ikeda et al. 2006). The emissive layers and their adjacent layers are particularly sensitive to moisture levels in the vacuum environment (Yamamoto et al. 2014).

In this chapter, we will introduce the basic principles behind VTE and then briefly touch upon a few alternate vapor deposition methods and thermal transfer methods that are under development. We will then describe the different VTE equipment configurations and their appropriate application. The most important component of any OLED deposition system is the material source design; therefore, the various source configurations will be discussed in further detail. We will also highlight the issues affecting productivity and maintenance of typical VTE systems and finally discuss some future trends and considerations for OLED manufacturing equipment and methods.

Vacuum Thermal Evaporation

Vacuum thermal evaporation is a well-known physical vapor deposition (PVD) technique, whereby a vaporized form of a material is generated from a heated source, transported through a vacuum, and finally allowed to condense onto the surface of a

substrate in order to grow or deposit a thin film of the material. The evaporation process has been used for deposition of a wide variety of materials, especially metals such as aluminum, and it is the most common method used for deposition of organic materials. It is based on the concept that a finite vapor pressure exists above any material. Vapor pressure is defined as the pressure exerted by a vapor on the solid or liquid phase with which it is in equilibrium at a given temperature in a closed system. At pressures lower than the vapor pressure, more atoms or molecules of the liquid or solid vaporize and escape from the surface than are absorbed from the vapor, resulting in evaporation. At the equilibrium vapor pressure, the exchange is equal and there is no net evaporation.

The vapor pressures of most common elements and gases were characterized in the 1960s by Dr. Richard Honig from RCA Laboratories, who produced a series of vapor pressure curves still used today (Honig and Kramer 1969). These curves show the vapor pressure of a material as a function of temperature, whereby the vapor pressure increases with higher temperature. A given material either sublimates (transitioning from a solid directly to a vapor) or evaporates (transitioning from a liquid to a vapor). In the case of most organic materials used in OLED devices, the dry powder material evaporates via sublimation given that a higher vapor pressure is achieved well before reaching the melting point. However, a few commonly used organic compounds are known to melt first, such as 4,4'-bis[*N*-(1-naphthyl)-*N*-phenylamino]biphenyl (NPB). Most organic compounds used in OLED devices have sufficiently high vapor pressures to be evaporated at temperatures below 400 °C.

From the kinetic gas theory, the evaporation rate into a vacuum is given by the Hertz-Knudsen equation:

$$J = \frac{\alpha \cdot M \cdot A (P' - P)}{\sqrt{2\pi \cdot M \cdot RT}}$$

where

J = evaporation rate (kg/s)

A = surface area

α = sticking coefficient for gas molecules onto the surface

P' = vapor pressure

P = partial pressure in gas mixture

M = molecular weight (kg/mol)

T = temperature

R = gas constant = 8.3143 (J/mol-K)

This expression describes the vapor flux. The vaporized molecules are transported to the substrate surface ideally in a line-of-sight fashion. A high-vacuum environment ensures a long mean free path and minimizes collisions of the gas molecules. The distribution of evaporant depends on the geometry of the source. Theory and experiments have shown that the distribution of vapor from a simple source is cosine (in a close analogy to Lambert's cosine law in optics), such that a mass of material evaporated from a small area (point source) at a given angle θ is

proportional to $\cos(\theta)$. For a point source, the distribution depends on r and θ as illustrated in Fig. 1.

The distribution of vapor onto the substrate can be expressed as

$$\frac{dM_s}{dA_s} = \frac{M_e \cos \theta}{4\pi r^2}$$

where

θ = tilt of dA_s from radial direction

Projection of dA_s onto sphere of radius $r = dA_s \cos \theta$

dM_s = mass hitting dA_s

M_e = total evaporated mass

The thickness uniformity of a film deposited from a point source onto a planar substrate can be depicted as in Fig. 2.

The maximum thickness occurs where $l = 0$. Applying the cosine law, the thickness d relative to the maximum thickness d_0 can be expressed as

$$\frac{d}{d_0} = \frac{1}{\left(1 + (l/h)^2\right)^2}$$

As an example, in order to achieve a film thickness uniformity of 95 % over a 400 mm substrate, a throw distance of over 1.2 m is calculated. This becomes impractical for large-area deposition systems. Therefore, point sources are used mainly in smaller research and development coaters with small-area substrates, typically less than 100 cm² in size. To overcome the limitations of point sources and enable uniform coatings over much larger areas, equipment manufacturers have developed large-area sources such as arrays of point sources, area sources such as circular manifolds or large-area showerheads, and linear sources which are essentially a series of point sources arranged in a linear fashion. These techniques allow for the source and substrate to be in much closer proximity, which improves the utilization of the expensive organic materials. The material utilization using a point source is typically less than 5 %, meaning that more than 95 % of the material is deposited onto surfaces other than the substrate, such as the chamber walls, shields, masks, and other surfaces.

Much is understood about the mechanisms of thin-film growth (Ohring 2001). The gas molecules impinging on the substrate surface may adsorb and stick immediately, but they are much more likely to diffuse around on the surface to find an appropriate site before sticking. Some molecules may immediately reflect off the surface, and some may desorb after some residence time. Since the incident molecules have a higher kinetic energy than the surface, they must decrease in energy and equilibrate with the substrate surface before sticking. This process is influenced by the substrate temperature, evaporation rate, vacuum pressure, throw distance, and other factors. There are several stages of film growth. Within the mean residence time, impinging molecules migrate along the surface and clusters start to form. Since the clusters have smaller surface-to-volume area than the molecules, the desorption

Fig. 1 Cosine distribution from a point source

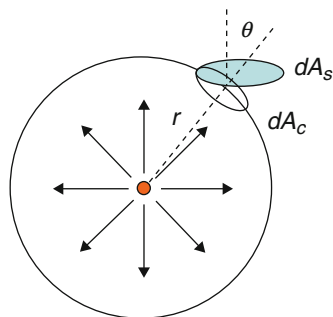
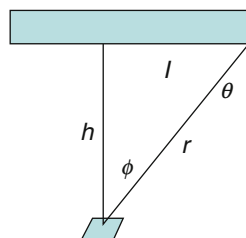


Fig. 2 Film thickness uniformity on a planar substrate



rate decreases, and nucleation can occur. Clusters can join and coalesce into new islands, exposing fresh surface areas where new molecules can adsorb. Larger islands join together leaving holes and channels, which then fill to form a continuous film. The film morphology and roughness can be varied by adjusting the deposition process parameters, particularly rate and temperature as well as vacuum pressure. Studies have shown that OLED device performance can be improved by controlling and stabilizing the film morphology through optimization of the deposition process conditions (Caria et al. 2006). In general, evaporated thin films of most organic materials are largely amorphous structures with randomly oriented interspersed crystallite regions, which results in films that are relatively smooth in nature.

Alternate Vapor Deposition Methods

Before getting into further detail about VTE equipment configurations and deposition sources, it is appropriate to discuss a few alternate vapor deposition methods that have been developed or are currently under development. Organic vapor-phase deposition (OVPD) was pioneered by Stephen Forrest at Princeton University, where the first heterojunction OLED devices fabricated using OVPD were demonstrated (Baldo et al. 1997). The technology is owned by Universal Display Corporation and is now exclusively licensed to Aixtron for equipment manufacturing. The basic principle of OVPD involves gas-phase transport, where vaporized materials are transported to the substrate via an inert carrier gas. The deposition rate is proportional to the molar flow of the source materials, and the rate is controlled by mass flow control of the carrier

gas, which is either nitrogen or argon. A thorough description of OVPD technology is given by Heuken and Meyer (Heuken and Meyer 2006).

The process takes place at a much higher pressure (~ 1 mbar) than VTE ($< 10^{-6}$ mbar). Although this avoids the need for high-vacuum systems, it is not clear whether or not state-of-the-art OLED devices with performance similar to VTE can be produced using the OVPD technique. While single-stack OLED devices have been reported with performance similar to VTE devices, stacked OLED devices fabricated by OVPD have yet to be demonstrated. OVPD allows for precise control of gas flows for multiple component layers, such that it is possible to grade the composition throughout the layer. This technique has been shown to produce devices superior to nongraded devices (Keiper et al. 2012). The main advantages of OVPD are the high deposition rates, high material utilization, and excellent film thickness uniformity. A high rate of flux at lower temperature is made possible by efficient vapor generation as dry powder particles are introduced into a hot gas stream. The particles are heated uniformly and can maintain wide separation, resulting in large effective surface area and low partial pressure around the particles. This is in contrast to a heated crucible as in the case of VTE, which requires a higher thermal load to effectively vaporize the material. High material utilization of $> 60\%$ is possible by the very close proximity of the showerhead and substrate. The substrate is actively cooled to allow for rapid and controlled condensation of the vapor as well as avoidance of a thermal overload to the substrate. The chamber walls are also heated to avoid condensation of vapor on chamber walls, which reduces cleaning requirements and yield loss associated with particle generation due to buildup of organic material within the chamber. The close coupled showerhead technology allows for highly uniform film thickness (1–2 %) over a very large area with homogeneous mixing of multiple components, due to uniform distribution of vapor within the heated showerhead. In contrast to VTE systems, the substrate is processed face up during the OVPD process. If particles are generated around shadow masks or other fixturing within the chamber, the substrate orientation could be a disadvantage.

Early versions of OVPD systems held the organic materials in heated sources for extended periods of time. Long-term exposure to elevated temperatures could potentially degrade the material, similar to VTE. Recently, the company has been developing a family of sources that reduce the thermal exposure of the organic material (Gersdorff et al. 2011). The concept is demonstrated in Fig. 3. The bulk material is kept at room temperature and fed into an aerosol generator with an inert carrier gas. The aerosol particles are delivered to a flash vaporizer as needed, so the thermal exposure is very short and the vaporization temperatures can be kept much lower than in the case of VTE. This technique allows for the possibility to use thermally sensitive organic materials which may be prone to decomposition.

Unlike VTE, OVPD has not yet achieved widespread adoption for OLED manufacturing. This may be due in part to the lack of availability of fully integrated production systems which not only deposit the organic layers but also deposit the cathode metal and in some cases thin-film encapsulation layers, as well as provide automated handling systems. Successful VTE equipment suppliers have offered full-scale automated manufacturing systems and not left the equipment integration up to

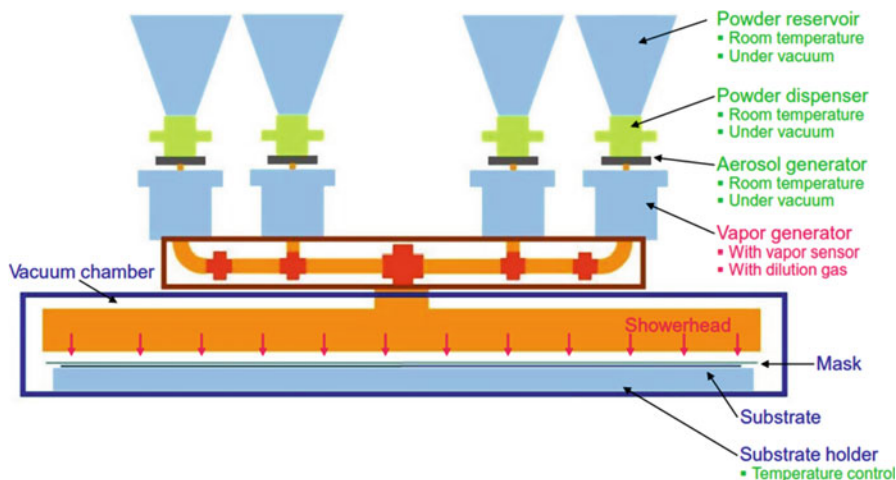


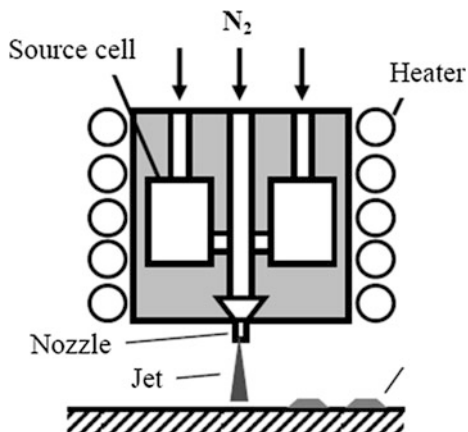
Fig. 3 OVPD short thermal exposure source design (Reproduced from Long et al. 2012)

the OLED manufacturer. To address this issue, Aixtron has recently partnered with Manx AG to develop and deliver full-scale automated Gen8 ($2.3 \text{ m} \times 2.5 \text{ m}$) production systems based on OVPD technology. This effort appears to be focused toward the AMOLED television market where the main benefits of OVPD technology, namely, high material utilization and excellent uniformity over a large area, may be realized. In OLED lighting deposition equipment systems, the most common system configuration for high-throughput uses substrates moving continuously past in-line linear sources mounted transversely to the substrate-motion direction. There have not been any demonstrations of OVPD in such machines, perhaps due to the added complexity of requiring active cooling of the moving substrates.

Organic vapor jet printing (OVJP) is another vapor deposition technique under development by the Forrest group at the University of Michigan (Shtein et al. 2003). The concept allows for precise patterning of organic materials and may be more applicable to displays, although the technique has been used to print RGB stripes in a color-changing OLED lighting panel demonstrated by UDC. A schematic of the OVJP system is shown in Fig. 4.

Unlike solution processing, the organic materials are not dispersed in a solvent. Evaporated material is volatilized and entrained in a heated carrier gas stream and then fed into heated channels and distributed into arrays of nozzles. The nozzle arrays are formed in silicon using MEMS fabrication techniques much like inkjet printheads are formed. Vapor from multiple sources can be mixed before being distributed into the nozzles. The vapor is finally collimated into fine microjets and allowed to condense on a cooled substrate in precise patterns. The carrier gas is allowed to escape through channels around the raised nozzles and is pumped away in the surrounding vacuum chamber. The gap between nozzle and substrate is very small, typically less than $100 \mu\text{m}$, to minimize shadowing and maintain high aspect ratio features. The nozzle shape is also an important consideration in the final deposition profile. The substrate is

Fig. 4 Organic vapor jet printer



moved on a precision stage relative to the nozzles to form the desired patterns on the substrate. OLED devices deposited by OVJP have been demonstrated with efficiency similar to that of OLEDs deposited by VTE, although the devices tend to show more droop in efficiency at higher current densities (McGraw and Forrest 2012). Scaling of the technology is possible by utilizing multiple nozzle arrays, with the potential for parallel deposition of RGB stripes simultaneously. Further research and development is required to improve device performance and optimize the process technology before OVJP can be considered for commercial applications. This technology is targeted for OLED devices with patterned RGB pixels or stripes, rather than displays or lighting based on unpatterned deposition of all organic layers.

Another vapor deposition method worth mentioning is vapor injection source technology (VIST), which was developed at Eastman Kodak Company (Long et al. 2009). VIST is essentially a flash vaporization technology, where organic powder held at room temperature is metered by an auger screw at a controlled rate toward a flash heater element. The vapor is then distributed uniformly into a heated manifold with apertures to allow the vapor to exit. A Gen5 linear manifold was demonstrated with the capability to deposit organic layers with 99 % uniformity over the substrate area. Because the vaporization is instantaneous, multiple component layers could be deposited from a single-powder mixture having a predetermined composition. Vaporization rate could be controlled through a low-temperature Pirani pressure gauge which provided feedback to drive power to the flash heater, so the rate response time was almost instantaneous. Despite promising advances with this technology, it was never commercialized.

Alternate Thermal Transfer Methods

Next, we discuss another class of flash vaporization technologies that involve thermal transfer of an organic material from a donor substrate by means of laser irradiation. These methods were developed to address the fine patterning issues

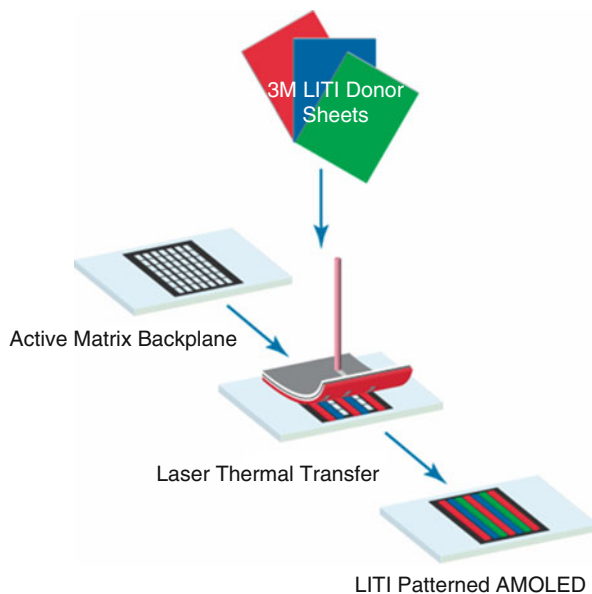
associated with OLED displays and provide a direct write method for color patterning. Although more useful for OLED display manufacturing, these techniques could be applied to patterning of RGB striped architectures for color-changeable OLED lighting.

Laser-induced thermal imaging (LITI) was developed by 3 M, in collaboration with Samsung (Lee et al. 2002). A Gen4 LITI pilot line was installed by Samsung to evaluate the technology on a commercial scale (Lee et al. 2007). The technique involves several process steps: deposition of the organic material on a specially designed donor film, precise alignment of a large format laser imaging system to a substrate, lamination of the donor onto the substrate, laser imaging of the donor film, and removal of the imaged film, as shown in Fig. 5.

The donor film is a transparent substrate with three layers: an absorption layer of carbon black-based ink that absorbs the laser energy and converts it to heat, an interlayer that forms a barrier between the absorbing layer and organic material, and the transfer layer, which is the organic medium that will be transferred to the substrate. The interlayer was found to be important to protect the sensitive organic transfer layer from chemical, mechanical, and thermal damage. Typically only the colored emitting layers (EMLs) are transferred by LITI. The adjacent layers are deposited by standard VTE approach. Significant lifetime improvement was demonstrated by multilayer transfer of both hole transport layers (HTLs) and EML, compared to just transfer of the EML (Wolk et al. 2008). Despite advances in device lifetime, development of simplified device architectures that avoid thermal transfer of the blue EML, and demonstration of many display prototypes fabricated by LITI, the technology is still under development and has yet to reach mass production status (Suh et al. 2009).

Radiation-induced sublimation transfer (RIST) is another laser thermal transfer technique developed by Kodak (Borson et al. 2005). The donor film is a polyimide-based substrate with Si- and Cr-absorbing layers, followed by the organic transfer layer. The main difference from LITI is that the donor is held in noncontact with the substrate by use of a clamshell mechanism to maintain a gap of 1–10 μm . Theoretically, this could reduce particle contamination resulting from contact transfer as in the case of LITI. Device performance approaching that of VTE was demonstrated, although lifetime was not quite comparable. A slight variation of RIST called laser-induced pattern-wise sublimation (LIPS) was developed by Sony (Hirano et al. 2007). The main difference was that the donor substrate was glass instead of polyimide. A rigid donor sheet had the potential benefits of simpler handling, maintaining dimensional stability over large areas, and cost savings due to reuse of the donor substrate. Again, devices made by this technique showed poorer lifetime as compared to VTE control devices, and little has been mentioned about this development effort since. Another similar laser transfer technique called laser-induced local transfer (LILT) was developed by Kroger and Kowalsky (Kroger et al. 2005). Although OLED display prototypes were successfully fabricated using most of these laser transfer methods, none have been commercialized to date, likely due to inadequate device lifetime. Fundamentally, the emitting layers in OLED devices and their interfaces with adjacent layers are very sensitive to

Fig. 5 Schematic of the LITI process (Reproduced from Wolk et al. 2008)



chemical, thermal, and morphological variations. Attempting to mass transfer a bulk film by flash vaporization using a laser may cause degradation of the organic material and make it difficult to control interfacial boundaries between layers. Additionally, there are cost concerns associated with fabrication and handling of donor substrates and materials, as well as waste of unused material on the donor substrate.

VTE Equipment Configurations

Research Systems

Vacuum thermal evaporation equipment is used in research mainly for formulation development and for material testing. Lower cost systems based on single chambers are usually employed (“bell-jar” systems) – and most must be vented to change either the substrates or to reload or change the materials. In order to efficiently screen new materials against formulations with reference materials, these systems can hold 5–20 materials and 6–12 substrates. The most common arrangement is to have a shutter system or rotating substrate platen in order to allow exposing one substrate at a time. More complex systems include substrate shutters as well. These systems typically do not contain enough material sources to deposit full multi-stack white formulations, so simpler model formulations are used for the material screening and layer optimization. Substrate size in this type of system is typically limited to 50–70 mm square shapes. Usually several small active areas (icons) are designed onto the substrates – from 0.05 cm² to 5 cm². The smaller icons are good for

accelerated lifetime testing allowing several current densities to be tested simultaneously. Larger icons are useful for light extraction development work.

These systems usually have two masks – an organic mask and a cathode mask – and a system for changing from one mask to another. Often, the masks are held in cavities in one circular platen that can be rotated separately or together with the circular platen where the substrates are being held. In order to increase productivity by reducing the pumpdown time, these bell-jar lab systems are typically installed with a dry N₂ glove box, enabling multiple runs per day. This also permits encapsulation of the OLED devices before exposure to oxygen and moisture.

Cluster-tool systems are used for large substrates (up to 200 mm by 200 mm) and direct simulation of production operations. These systems can cost several million dollars and are housed in clean room facilities with support operations such as substrate cleaning. Cluster-tool systems have one or more central robots in a 6–8 facet central chamber – where the robot can reach to individual chambers serving the function of deposition, load, unload, pass-through to another cluster, flip, pretreatment, etc. In more sophisticated systems, the masks can also be handled by the robots. Gate valves isolate the individual chambers so that they can be opened for material filling and cleaning. Typically each chamber is dedicated to a layer and holds less than six sources. The robot passes substrates one at a time into the peripheral chambers, and typically the substrates are held stationary on the mask, with a sufficient source to substrate distance such that the deposition from the point source was uniform over the active area of the substrates. If additional layers were required for more complex structures, additional clusters would be added. Usually an unload chamber is connected to an automated (robotic) encapsulation system in a dry box. Typical productivity is 8–12 substrates in a working day. These larger substrates can be used to make prototype devices of both displays and lighting panels.

Production Systems

Early production systems for both displays and lighting were based on larger version of the cluster tools. Today the most common size of lighting substrate is G2 (370 × 470 mm). In the deposition chambers, the vapor sources can be either point sources or linear sources. If point sources are used, the glass is far from the sources and the glass and mask are rotated during deposition to achieve uniform distribution across the active area. If linear sources are used, then usually the glass and mask are stationary and the cluster of linear sources (one for each material in the layer) is translated below the substrate. In both of these systems, the sources are typically maintained at rate even while the substrates are being moved in and out of the chamber, resulting in <10 % material use efficiency (MUE).

In modern in-line production systems, linear sources are used and the glass and masks are transported in carriers through a series of chambers where one layer is deposited per chamber. Usually the linear sources have shutters, and some have shutoff valves to conserve materials during idle periods.

Typical TAC times for these systems are 3–6 min, with a goal to reduce this to 1 min or below.

With in-line production systems, the glass must be transferred from carriers with organic-pattern masks to carriers with cathode-pattern masks before the cathode deposition. The carriers and masks must be recycled back for reuse within the machine. Two approaches are used: the carriers and masks can be returned along an upper level in the deposition chambers, or the in-line sections can be built in a loop providing for one-way carrier flow.

The Philips G2.5 (400 × 500 mm) production line is shown in Fig. 6. This machine uses the loop configuration for both the organic and cathode in-line sections, with turning chambers at the corners (Fig. 6).

In 2014, Konica Minolta announced that they were building a roll-to-roll OLED lighting facility that would be capable of producing one million 7 × 7 cm flexible panels per month on their proprietary barrier-coated substrate (Tsujiura et al. 2014). This line is expected to start production in 2015. The line was said to be capable of color-changing panels as well.

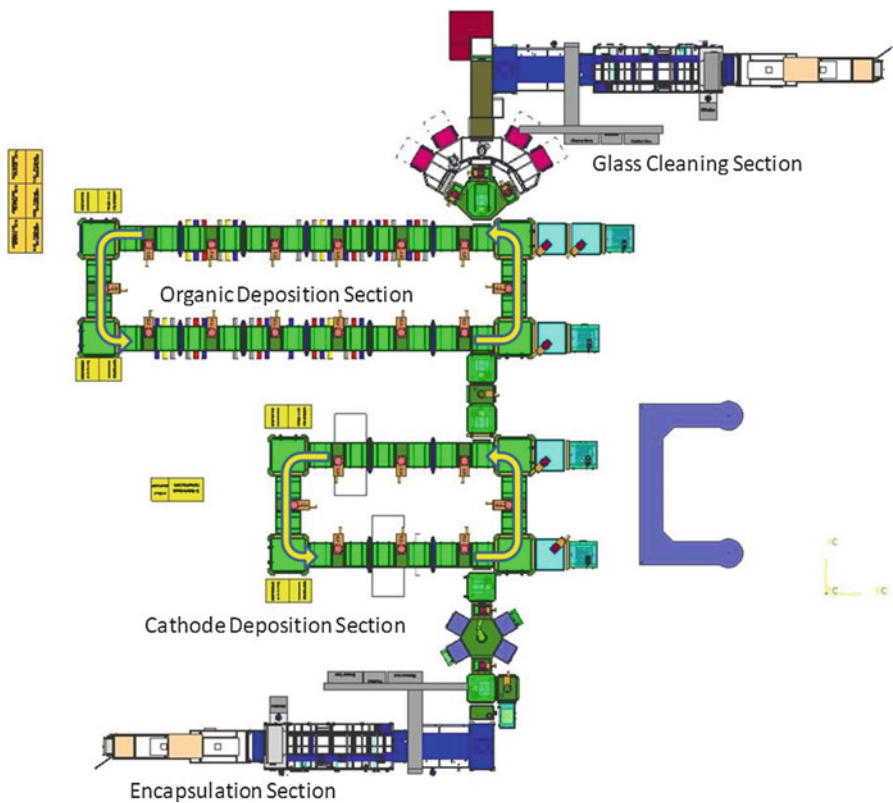


Fig. 6 Philips G2.5 OLED lighting production line (Reproduced from Hoffman 2015)

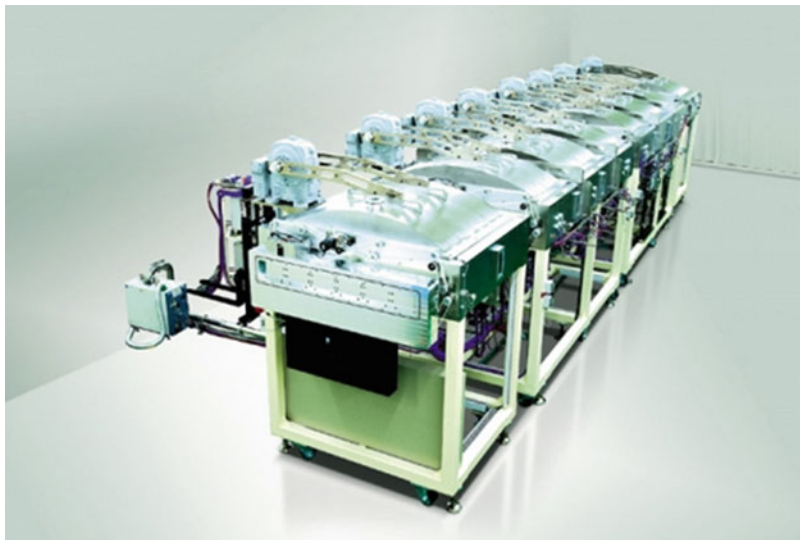


Fig. 7 Sunic G5 OLED lighting production system

Future Production Systems

Several designs have been proposed for higher-throughput production systems. Applied materials proposed a design for Gen4 glass which held the glass in a vertical orientation (Hoffman et al. 2010). A government-sponsored project in Korea (June 2010 to Aug 2012) developed the technology for a Gen5 horizontal glass system shown in Fig. 7 (Lim et al. 2013).

In February 2015, LG Chem announced their plans to purchase a G5 (1.25 × 1.1 m) OLED lighting system for \$184 million USD. This is expected to reduce the cost of a 10 × 10 cm OLED lighting panel to about \$5 (<http://asia.nikkei.com>). At 60 s TAC time and 80 % uptime, this machine will be able to process more than 500,000 m² of glass substrates per year.

Organic Source Configurations

Point Sources

The selection of materials for heating the organic materials depends mostly on chemical compatibility. The most common source material is metal – tantalum is popular in research coaters, whereas stainless steel is used in production machines. Alternatives include quartz glass and graphite.

In simple point-source systems, the vapor emits directly from the crucible into the chamber. The photo in Fig. 8 is from a G3.5 Tokki machine using an array of point sources to achieve uniformity. Each turret usually contains one material. There is a

Fig. 8 Turrets of point-source vaporizers at the *bottom* of a rotating substrate deposition chamber



heating position and a preheat position, and the other five positions hold additional quantities of material. This chamber is a spinning substrate chamber so the sources are arranged roughly under the edge of the rotating substrate active area in order to achieve good uniformity across the substrate.

Linear Sources

The earliest linear sources for organic material deposition were quartz boats with tantalum top heaters containing small precise orifices spaced to achieve a uniform deposition on the substrate (Van Slyke et al. 2002). The prototype source shown in Fig. 9 was made by Kodak and used at SKD in Japan (Hamer et al. 2005). The three sources deposited one layer of three components, where each source was monitored and controlled separately.

The three sources are tipped together so that the center of the plumes would overlap at the substrate to get the correct composition for the widest area as demonstrated in Fig. 10 below (Long et al. 2009).

The next generation of linear source for organic deposition used a side-heater design and graphite crucibles, as shown in Fig. 11. This design was developed by ULVAC in Japan and also used at SK Display.

The linear sources from a Sunic Gen 5 machine are shown in Fig. 12 (Murano et al. 2014). The orifices are angled such that the vapor plumes converge at the same position on the substrate.

Two innovations that were introduced to reduce the waste of organic materials and hence reduce the cost of lighting panels were the use of valves and heated shields. The use of valves between the vapor generation section and the nozzle in order to shut off the flow while the nozzles are idle was introduced by Veeco (Kim et al. 2009).

Fig. 9 Early linear source design

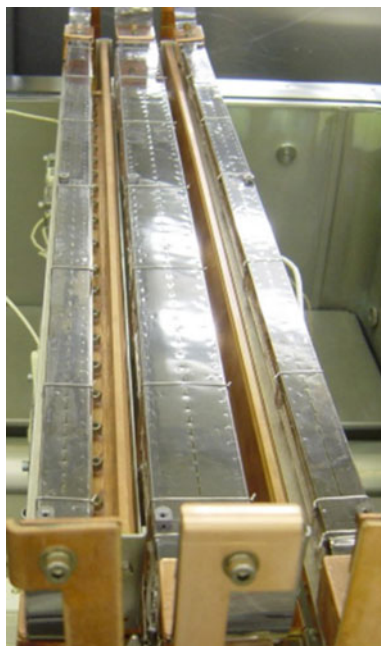


Fig. 10 Schematic of overlapping vapor plumes from a set of linear sources (Reproduced from Long et al. 2009)

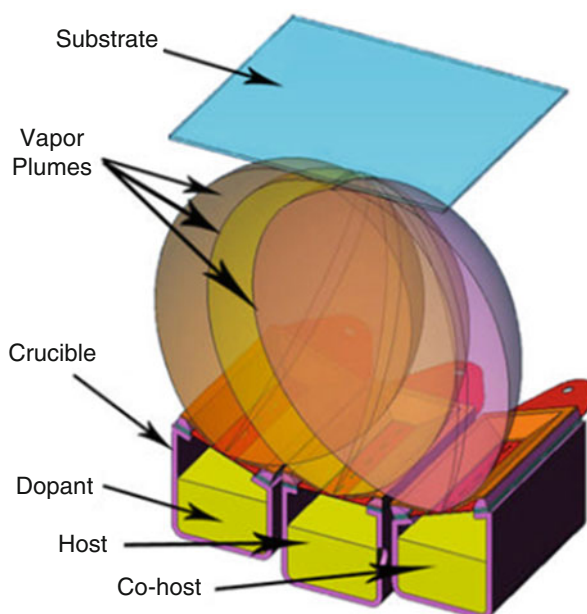


Fig. 11 Linear sources with side-heater design

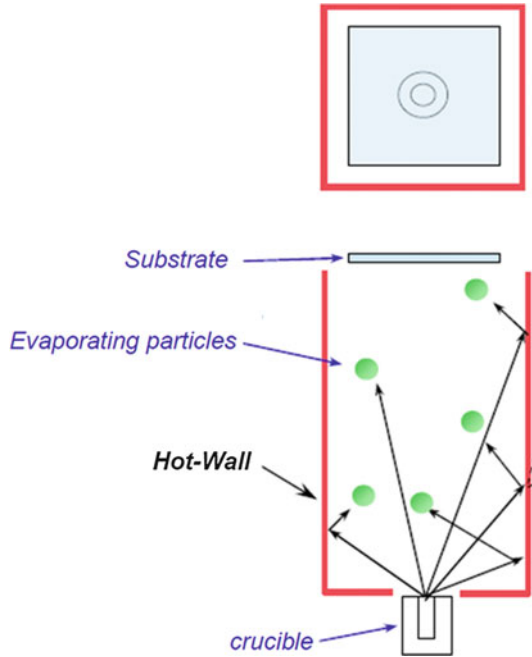


Fig. 12 Sunic G5 linear sources (Reproduced from Murano et al. 2014)



Tokki introduced “hot-wall” sources to reduce the waste due to material deposited on cold shields (Matsumoto et al. 2003). Recently, Panasonic has employed the hot-wall source technology to produce high-efficiency white OLED lighting prototypes as shown in Fig. 13 (Komoda 2011).

Fig. 13 Hot-wall OLED deposition source
 (Reproduced from Komoda 2011)



There are two challenges in linear source design:

1. Achieve the greatest possible evaporation rate without degrading the materials. This is limited by the compound in the formulation that evaporates at the desired rate nearest its thermal decomposition point. Improvements are possible by source designs that use large areas for sublimation, create the minimum pressure drop from evaporation to the substrate, and minimize the temperatures of the surfaces along the way.
2. Achieve the greatest possible MUE. This can be accomplished by using heated-wall sources or by moving the nozzles close to the substrate. In both cases, care to prevent overheating the substrate is important.

Metal Sources

Most OLED lighting panels are bottom emitter architecture with an opaque top cathode. The most common cathode metals are aluminum and silver. Silver has the advantage of higher reflectivity in the visible wavelengths, which is particularly helpful with scattering-type light extraction mechanisms because more of the photons exiting from these devices are reflected from the cathode than in a non-light extraction device.

The three most common ways for depositing the cathode are thermal evaporation, sputtering, and e-beam evaporation. Of these, thermal evaporation is the most

common because the particles (atoms or clusters of atoms) have very low kinetic energy compared to other methods. When the particles have larger kinetic energy, they impact the surface of the OLED and penetrate into the organic layers. This damage results in higher voltage and lower lifetime (Gil and May 2010). Barrier layers and thick doped electron injection layers can be used to reduce this damage at a small voltage penalty (Hung et al. 1999). With early e-beam evaporation sources, the generation of UV and x-rays also damaged the OLED organic layers.

For thermal evaporation, sources are usually either ceramic crucible type with external heaters, ceramic boat type, or metal boat type. The ceramic crucible boats typically hold more and must be heated and cooled more slowly. These type of sources are common in the semiconductor industry and they are particularly suitable for continuous operation where the large quantity must be kept liquid for an extended period of time. Figure 14 shows an example of a large-capacity ceramic crucible for holding liquid metal.

Each material has its own difficulty. Liquid aluminum is chemically very reactive so pyrolytic boronitride is a common choice. Liquid silver has a very high surface energy, causing it to wet up the sides of vessels which can cause problems.

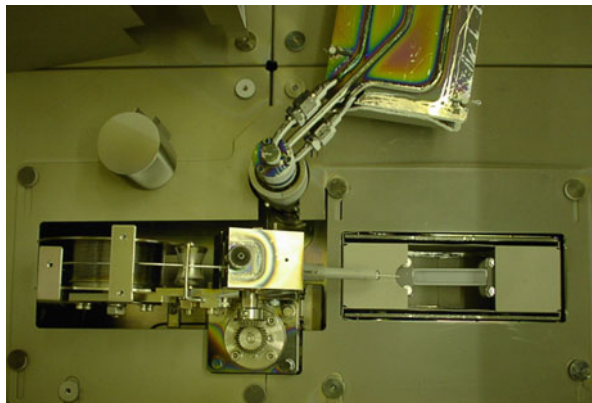
One solution which is used for aluminum is to “wire feed” a small diameter solid wire from a spool into an electrically heated shallow cavity in a pyrolytic boronitride block. The resistance of the liquid pool can be used to sense the liquid level and the wire feed rate can be adjusted to maintain the liquid level. Figure 15 shows an example of a production wire feed mechanism.

Work has been done to create “linear sources” for metal vapor. Due to the reactive nature of liquid aluminum with metal, these have proven expensive and difficult to

Fig. 14 Ceramic crucible for holding liquid metal for evaporation



Fig. 15 Wire feed Al evaporator system with shield over wire feeder removed



manufacture. For continuous in-line deposition machines, the more common approach is to use several “point sources” across the width of the glass. Figure 16 shows the Sunic design for a Gen 5 OLED lighting deposition machine which uses linear arrays of three metal evaporators for aluminum across the width of the chamber (Lim et al. 2013). One linear array is used while the other array is refilled by the feeders. Sunic demonstrated that the rate could be held within $\pm 5\%$ of the target rate of 10A/s over 325 h.

Sometimes higher temperature materials are used in OLEDs, such as LiF or metal oxides. These materials are typically evaporated from crucible-type point sources. The point sources can be metal or ceramic. A metal source with two small orifices is shown in the photo in Fig. 17.

VTE Equipment Productivity

VTE equipment is very expensive. Estimates for the capital cost of equipment used for OLED lighting panels made in the 2014 US DOE SSL Manufacturing Roadmap are shown in Table 1 (US DOE 2014, p. 81).

In a traditional manufacturing configuration, the cost of the OLED deposition equipment is about half of the total equipment cost. TAC time is defined as the period at which substrates flow in a continuous machine. Manufacturers typically claim their equipment is capable of depositing a layer with nominal thickness with good uniformity in both down longitudinal and transverse directions at TAC times of 1 min, as is the case with the Gen 5 Sunic tool (Lim et al. 2013). In practice, TAC times are initially quite a bit longer, often in the range of 3–6 min when the equipment is in its first few years of operation (Park 2012). Continuous in-line machines must run at the speed dictated by the layer that is the slowest to deposit. This limitation could be due to the thermal instability of the material with the shortest lifetime at the vaporizer operating temperatures. Typically it is necessary to raise the vaporization temperature about 10 °C in order to double the vapor generation rate,

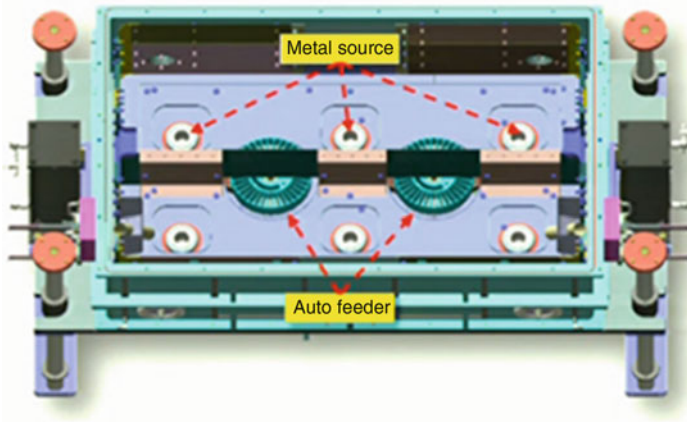
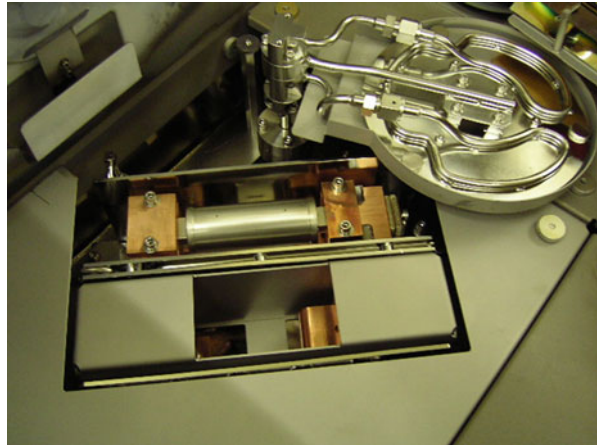


Fig. 16 Linear arrays of point sources for metal deposition in Gen 5 Sunic tool (Reproduced from Lim et al. 2013)

Fig. 17 High-temperature point sources with shielding over source removed



and TAC time is inversely proportional to the vapor generation rate. If thermal decomposition is the problem, possible solutions include:

1. Select an alternative material that vaporizes at the required rate at a temperature which is a safe margin below the decomposition temperature and reformulate the product to use the alternative material.
2. Select a different vaporizer design that can generate more vapor at a lower temperature – such as a crucible with a larger surface area for vapor generation.
3. Redesign the formulation to use a lower laydown (layer thickness or composition) of the problem material.

Table 1 Figures for 80 % yield, 80 % uptime, 80 % glass pattern efficiency, and 1 min TAC time

Gen size	Glass size (m)	Area (m ²)	Equipment cost estimate (USD)	Area product per year (m ²)	Depreciation per m ²
2	0.37 × 0.47	0.17	\$50–100 M	46,798	\$134–267
5	1.1 × 1.3	1.43	\$150–300 M	384,823	\$49–97
8	2.2 × 2.5	5.50	\$300–600 M	1,480,090	\$25–51

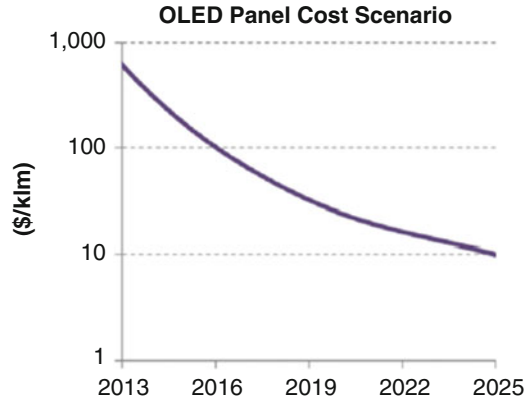
4. Use two or more deposition sections for the layer with the problem material in order to achieve the desired TAC time.
5. Reload the thermally sensitive material more frequently in order to prevent decomposition.

Achieving a low TAC time is very important, as the output of the machine is inversely proportional to the TAC time. A major part of the cost of goods sold (COGS) of the OLED lighting panels can be due to the equipment depreciation. At 1 min TAC times, with standard industry goals of 80 % uptime, 80 % yield, 80 % glass pattern usage efficiency, and 8-year depreciation lifetime for production equipment, the depreciation cost per square meter of yielded product is shown in Table 1. The DOE has published panel cost targets that it believes are required in order to realize a “viable OLED lighting market” (US DOE 2014, p. 5) – shown in Fig. 18.

At the common brightness for OLED lighting panels of 10 k lm/m², the path in Fig. 18 shows that the panel cost in 2025 must reach \$100/m². Only the Gen 8 machine in Table 1 has a hope of competing at that point, and even then it is not acceptable to have such a 25–50 % of the COGS allocated to depreciation. Possible solutions to this problem are to either achieve shorter TAC time or develop lower cost equipment or a combination of the two.

Organic materials used in OLED devices are expensive, so the MUE of the equipment and its operation are critical to achieving the cost reduction required in Fig. 18. Common estimates for MUE for organic deposition from a linear nozzle array in Gen2 equipment is 25 % and in Gen5 equipment is 50 % (Lim et al. 2013). If the layer is a single component, then the sources can be moved closer to the substrate to achieve a higher MUE; however, no machines are currently designed with variable substrate-to-source distances. Moving the nozzle array closer to the substrate may not be possible for multicomponent layers where the plumes from multiple linear nozzle arrays must overlap to get the desired deposition composition over much of the deposition area (i.e., the coating window in the shield through which the substrate is exposed to vapor). Since the nozzles in multicomponent layers all point at the substrate at different angles, it is inevitable that there are composition gradients across the longitudinal direction of the coating window and thus result in a composition gradient through the thickness of the deposited layer. There is a trade-off between reducing the composition gradient by reducing the coating window and reducing the MUE (thereby raising the cost of the organic material). Designing products that are robust to this gradient or take advantage of this gradient (Zhang et al. 2014) creates a win-win situation.

Fig. 18 OLED panel cost projection (Reproduced from US DOE SSL Manufacturing Roadmap 2014)



While moving the linear nozzle arrays closer to the substrate increases the MUE, and hence for the same vapor generation rate, enables a lower TAC time, there is the disadvantage that heat transfer to the substrate from the nozzle arrays becomes larger. Thus, heat shielding of hot parts becomes more important as the source to substrate distance becomes less. In VTE deposition, the substrate acts as a thermal integrator, and its temperature increases as each subsequent layer is deposited. There can be some subtle effect on device performance with substrate temperature that must be kept in mind. The most dramatic effect is where a thick layer has enough thermal energy to undergo a morphology rearrangement, such as the growth of large crystals – which is facilitated by high deposition rates of organic vapor onto the surface. This morphology change is generally very undesirable, but in some cases, it can result in a desired morphology; for example, in the case of Novaled’s nanocrystallizing electron transport material that can produce a layer that can extract more light from the device (Murano et al. 2012).

VTE Equipment Maintenance

The organic material and metal that does not deposit onto the substrate in the active area deposit onto the masks, frames, shields, shutters, and other parts of the machine. When the buildup of organic inside the machine becomes thick enough, the coating will begin to flake off resulting in particles inside the vacuum chambers. Coatings that deposit on moving parts will also generate particles when the parts move – such as when rollers contact the coated surfaces. For these reasons it is necessary to periodically stop the production operation, cool down the sources, vent the vacuum chamber, and clean the surfaces that have organic or metal deposited on them. The frequency of this operation depends on the material usage efficiency and the design of the machine. At the same time the organic and metal crucibles are refilled and the QCM heads have new quartz crystals loaded. The target limit for this scheduled maintenance for most machines is less than 10 %. Usually it takes 2–3 days for these

maintenance operations, including time for the chamber to regain base pressure so OLED deposition operations can resume.

Several machine design features allow for a longer operating time between maintenance shutdowns. These include higher material usage efficiency by having the linear nozzle array closer to the substrate, larger crucibles for material, loadlock arrangements to allow refilling of crucibles without venting the chamber, and the use of QCM heads that hold large numbers of spare crystals. If possible, as many moving parts should be below the substrate and the substrates should be electrically neutral before coating in order to reduce the attraction of particles to the substrate. QCM designs exist which allow heating the crystal in place to evaporate the deposited organic coatings, thus allowing crystals to be used in service for longer periods of time.

In order to reduce the downtime for maintenance, most machines have removable shields that can be quickly swapped for clean ones. A major factor in the downtime is the time required for pumping back to high vacuum afterwards. For large chambers, more than a day can be required. Larger pumps can help reduce this time, and ways to reduce the exposure of the chamber parts to water vapor such as keeping the chamber closed and under nitrogen also help. The metal finish of the inside of the chamber, shields, and other parts also affects the amount of water absorbed; however, highly finished chambers and parts can increase the cost of the equipment considerably. The goal is to design a machine so that it uses the materials with high efficiency and thus permitting less frequent shutdown for maintenance.

It is common that the first OLEDs made with a freshly cleaned chamber will have performance lower than those made during normal production – often including lower lifetime. This could be due to residual water vapor in the chamber or the presence of residual cleaning solvents. Some “seasoning” of the chamber by evaporating organics or other materials is often helpful to reduce both the magnitude and duration of the off-spec production period.

Future Directions

For the foreseeable future, the majority of high-performance OLED products will continue to be produced using VTE. It is expected that larger generation VTE production lines will be established, particularly in Asia where the display market infrastructure is centered, in order to meet the growing demand for OLED lighting products. How fast and how large the market grows depend strongly on how quickly the costs can be reduced to reasonable levels. VTE systems with larger area capacity, higher throughput, and greater organic material utilization will go a long way toward reducing costs and fulfilling the demand. There is also room for novel production methods and systems that have much lower capital costs and can be scaled appropriately as the market grows while still producing reasonably priced products even at low or moderate volumes. Such novel methods may allow more manufacturers to participate and help to grow the market faster.

References

- Baldo MA, Kozlov VG, Burrows PE, Forrest SR, Ban VS, Koene B, Thompson ME (1997) Low pressure organic vapor phase deposition of small molecular weight organic light emitting device structures. *Appl Phys Lett* 71:3033
- Bohler A, Dirr S (1997) Influence of the process vacuum on the performance of organic light-emitting diodes. *Synth Met* 91:95–97
- Borson M, Tutt L, Nguyen K, Preuss D, Culver M, Phelan G (2005) Non-contact OLED color patterning by radiation-induced sublimation transfer (RIST). *SID Int Symp Dig* 36:972–976
- Caria S, Da Como E, Murgia M, Zamboni R, Melpignano P, Biondo V (2006) Enhanced light emission efficiency and current stability by morphology control and thermal annealing of organic light emitting diode devices. *J Phys Condens Matter* 18:S2139–S2147
- Gersdorff M, Long M, Keiper D, Kunat M, Gopi B, Cremer C, Beccard B, Schwambers M (2011) Enabling high throughput OLED manufacturing by carrier gas enhanced Organic Vapor Deposition (OVPD). *SID Int Symp Digest* 42:516–519
- Gil TH, May C (2010) Origin of damages in OLED from Al top electrode deposition by DC magnetron sputtering. *Org Electron* 11:322–331
- Hamer J, Yamamoto A, Rajeswaran G, Van Slyke SA (2005) Mass production of full-color AMOLED displays. *SID Int Symp Digest* 36:1902–1907
- Heuken M, Meyer N (2006) Organic vapor phase deposition. In: Klauk H (ed) *Organic electronics: materials, manufacturing, and application*. Wiley VCH, Weinheim, pp 203–232
- Hirano T, Matsuo K, Kohinata K, Hanawa K, Matsumi T, Matsuda E, Matsuura R, Ishibashi T, Yoshida A, Sasaoka T (2007) Novel laser transfer technology for manufacturing large-sized OLED displays. *SID Int Symp Dig* 38:1592–1595
- Hoffman U (2015) Manufacturing of high quality OLED. Presentation at 4th annual China international OLEDs summit 2015, Shanghai
- Hoffman U, Landgraf H, Campo M, Bruch J, Keller S, Koenig M (2010) New concept for large area white OLED production for lighting. *SID Int Symp Digest* 41:925–928
- Honig RE, Kramer DA (1969) Vapor pressure data for the solid and liquid elements. *RCA Rev* 30:285–305
- <http://asia.nikkei.com/Business/Companies/LG-Chem-to-make-OLED-panels-very-affordable>. Accessed March 2015
- Hung LS, Liao LS, Lee CS, Lee ST (1999) Sputter deposition of cathodes in organic light emitting diodes. *J Appl Phys* 86:4607–4612
- Ikeda T, Murata Y, Kinoshita Y, Shike J (2006) Enhanced stability of organic light-emitting devices fabricated under ultra-high vacuum condition. *Chem Phys Lett* 426:111
- Keiper D, Meyer N, Heuken M (2012) Introduction to Organic Vapor Phase Deposition (OVPD) for organic opto-electronics. In: Logothetidis S (ed) *Nanostructured materials and their applications*. Springer, Berlin/Heidelberg, pp 155–170
- Kim WK, Han SY, Choi JO, Patrin J, Bresnahan R (2009) Development of large-sized AMOLED manufacturing system. *SID Int Symp Digest* 40:1359–1362
- Knox J, Halls M (2006) Chemical failure modes of AlQ3-based OLEDs: AlQ3 hydrolysis. *Phys Chem Chem Phys* 8:1371–1377
- Komoda T (2011) High quality white OLEDs and resource saving fabrication: processes for lighting application. Presentation at 2011 printed electronics and photovoltaics, Dusseldorf
- Kroger M, Huske M, Dobbertin T, Meyer J, Krautwald H, Riedl T, Johannes HH, Kowalsky W (2005) A novel patterning technique for high-resolution RGB-OLED-displays: Laser Induced Local Transfer (LILT). *MRS Proc* 870:H3.4
- Lee ST, Lee JY, Kim MH, Suh MC, Kang TM, Choi YJ, Park JY, Kwon JH, Chung HK (2002) A new patterning method for full-color light-emitting devices: Laser Induced Thermal Imaging (LITI). *SID Int Symp Digest* 33:784–787

- Lee ST, Suh MC, Kang TM, Kwon YG, Lee JH, Kim HD, Chung HK (2007) LITI (Laser Induced Thermal Imaging) technology for high-resolution and large-sized OLED. *SID Int Symp Digest* 38:1588–1591
- Lim GM, Lee JH, Kim JJ, Song KM, Jang TH, Hwang IH, Choi JS, Oh YM, Lim DC, Choi CS, Im Y, Lee YJ (2013) Development of highly productive In-line vacuum evaporation system for OLED lighting. *SID Int Symp Digest* 44:767–770
- Long M, Koppe B, Redden N, Boroson M (2009) Responsive vacuum deposition technology for cost-effective OLED manufacturing. *SID Int Symp Digest* 40:943–946
- Long M, Gersdorff M, Keiper D, Dauelsberg M, Beccard B, Cremer C, Trimborn KH, Poque A (2012) OLED manufacturing; Scaling from pilot production to economical manufacturing. Presentation at 2012 plastic electronics conference, Dresden
- Matsumoto E, Maki S, Yanagi Y, Nishimori T, Kondo Y, Kishi Y, Kido J (2003) The high deposition rate and high material yield evaporation method for OLED layers. *SID Int Symp Digest* 34:1423–1425
- McGraw G, Forrest SR (2012) Fluid dynamics and mass transport in organic vapor jet printing. *J Appl Phys* 111:043501
- Murano S, Pavicic D, Furno M, Rothe C, Canzler TW, Haldi A, Loser F, Fadhel O, Cardinali F, Langguth O (2012) Outcoupling enhancement mechanism investigation on highly efficient PIN OLEDs using crystallizing evaporation processed organic outcoupling layers. *SID Int Symp Digest* 43:687–690
- Murano S, Gilge K, Ammann M, Werner A (2014) AMOLED manufacturing – challenges and solutions from a material makers perspective. *SID Int Symp Digest* 45:403–406
- Ohring M (2001) *Materials science of thin films*, 2nd edn. Academic, London
- Park J (2012) A marketing perspective, OLED lighting. Presentation at 2012 OLED world summit, San Francisco
- Shtein M, Peumans P, Benziger JB, Forrest SR (2003) Micropatterning of organic thin films for device applications using organic vapor phase deposition. *J Appl Phys* 93:4005
- Suh MC, Kang TM, Cho SW, Kwon YG, Kim HD, Chung HK (2009) Large-area color patterning technology for AMOLED. *SID Int Symp Digest* 40:794–797
- Tsujimura T, Fukawa J, Endoh K, Suzuki Y, Hirabayashi K, Mori T (2014) Flexible OLED using plastic barrier film and its roll-to-roll manufacturing. *SID Int Symp Digest* 45:104–107
- U.S. Department of Energy (2014) Solid-state lighting research and development: manufacturing roadmap. Prepared by Bardsley Consulting, SB Consulting, Navigant Consulting, Inc, SSSL, Inc, and Radcliffe Advisors, Inc, Washington, DC
- Van Slyke S, Pignata A, Freeman D, Redden N (2002) Linear source deposition of organic layers for full-color OLED. *SID Int Symp Digest* 33:886–889
- Wolk M, Lamansky S, Tolbert W (2008) Progress in laser induced thermal imaging of OLEDs. *SID Int Symp Digest* 39:511–514
- Yamamoto H, Weaver M, Murata H, Adachi C (2014) Understanding extrinsic degradation in phosphorescent OLEDs. *SID Int Symp Digest* 45:758–761
- Zhang Y, Lee J, Forrest SR (2014) Tenfold increase in the lifetime of blue phosphorescent organic light-emitting diodes. *Nat Commun* 5:5008. doi:10.1038/ncomms6008

Part IV

Intelligent Lighting System Integration

Dimming

Joseph Denicholas

Contents

Introduction and the Importance of Dimming	447
Fundamentals of LED Dimming	447
Efficacy Improvement Under Analog _O and PWM _O Dimming	451
Color Fidelity Under Analog _O and PWM _O Dimming	452
Implementing Switched PWM _O Dimming	452
Dimming Considerations for the System Integrator	454
Introduction to Command and Control Systems for Dimming	455
Power Line Communications (PLC) and Phase-Based Dimming	455
Low-Voltage AC Systems	458
0–10 V Dimming	459
Intelligent Lighting Systems	460
System Architectures for Intelligent Lighting	460
Communications Infrastructures for Intelligent Lighting	463
Conclusions	464
References	464

Abstract

Dimming is critical to delivering the promise of solid-state lighting, but it needs to be implemented properly per the requirements of the application. Basic types of dimming, both analog and PWM, and their effect on LED characteristics are covered. Various methods of implementing fast and accurate PWM dimming are described. Phase-based dimming and the inherent complications associated with it are given some treatment, given its prominence in the market today, as is the 0–10 V dimming standard. The chapter then looks to the future and proposes various system-level architectures that support optimal intelligent lighting

J. Denicholas (✉)
Texas Instruments, Dallas, TX, USA
e-mail: joseph.denicholas@ti.com

solutions. Trade-offs and considerations when selecting a communications infrastructure, including the physical layer, protocol, and application layer, are detailed.

Glossary

Analog dimming	The act of dimming via reduction in average current level in the absence of PWM.
Contrast ratio	The minimum light intensity achievable by a dimming method with respect to the maximum or undimmed value, expressed as a ratio. For example, if a dimming method provides a way to achieve 0.1 % of full intensity, the contrast ratio would be 1,000:1.
EMI	Electromagnetic interference
Flicker	A catchall term for all light modulation, consistent or inconsistent in time, of any amplitude, visible or invisible. For further information, refer to Miller (2013).
Flicker index and flicker percentage	Terms used to express the extent and nature of light modulation. For more information, refer to Poplawski and Miller (2011).
IGBT	Insulated gate bipolar transistor
LED driver	The portion of system circuitry that processes power between the input source and the LEDs. LED drivers are also typically where a dimming command is translated to current, and hence light, modulation.
MOSFET	Metal oxide semiconductor field effect transistor
PLC	Power line communications
PWM	Pulse width modulation, a signal modulation technique that conforms the width of the signal pulse according to a pulsed input. It serves to decrease the average value of a signal proportional to the respective on and off times.
Ripple	Fluctuations in light that are (a) consistent over time and (b) less than 100 % in amplitude.
RMS	Root mean square

SELV	Safety extra low voltage
Strobing	Fluctuations in light that are (a) consistent over time and (b) typically 100 % in amplitude.
TRIAC	Triode for alternating current

Introduction and the Importance of Dimming

It would not be an exaggeration to say that dimming is critical to the mission of delivering the promise of solid-state lighting. With the exception of LED, all high-efficiency lighting technologies, be they fluorescent, high-intensity discharge (HID), or even more recent plasma-based technologies, suffer from significant challenges when attempting to dim. Challenges include difficulty of control, slow response time of the light source, or reliability and lifetime reduction under dimming. LEDs suffer none of these disadvantages in that dimming is generally straightforward to perform and control, the response times are on the order of nanoseconds or less, and dimming only increases lamp reliability and lifetime by decreasing the overall thermal load on the system. While improvements in system efficacy and reliability are clear benefits of solid-state lighting, they do not exploit what is arguably the key advantage of the LED – it is extremely easy to dim and doing so decreases energy usage, increases reliability, and greatly improves user experience. It also enables new technologies such as color mixing and demand response load shedding.

While more straightforward when compared against other high-efficiency lighting technologies, LED dimming is nontrivial. Poor selection of LED driver components and dimming methods can easily result in poorly performing and unreliable systems. Perhaps the most offensive result of a poorly architected dimming system is flicker. At a minimum, flicker is annoying and distracting. At its worst, it can cause headaches, eyestrain, and neurological problems. Consequently, flicker must be avoided at all costs.

This chapter aims to provide a comprehensive review of dimming fundamentals, methods, and applications. While a general foundation in power electronics may be helpful, it is not required to understand the fundamental concepts in this chapter.

Fundamentals of LED Dimming

Emitted light from LEDs is linearly proportional to the level of current being driven, excluding variations in efficacy inherent to LEDs. There are generally two methods to reduce the average current level, namely “analog” and pulse width modulation or PWM. An immediate point of confusion that is common when discussing dimming concerns input command/stimulus versus output LED current modulation. One should

always be clear whether or not it is the input control signal or the actual output current from the LED driver that is being described. On the input side, some LED driver systems accept no dimming command, some only PWM (e.g., from a microcontroller), some only analog (e.g., from a thermistor), and some both PWM and analog. But, this speaks only to how the driver itself is being communicated with and nothing about how the current, and hence the light, is actually being modulated at the output of the driver. In order to avoid confusion, for the purposes in this chapter, an “I” subscript character will be used when referring to the input control method, such as analog_I or PWM_I, while an “O” will refer to the output modulation method, such as analog_O or PWM_O. No subscript will be used when both input and output are being referred to.

Figure 1a shows LED current as reduced via analog_O dimming. In this example, though the current contains a ripple component from the switching action of the driver, its average, or DC, value is reduced, thus dimming the light. No large signal pulsing of the waveform is present, and therefore no filtering is typically required to avoid visible flicker.

With PWM dimming, the LED current is actually being engaged and disengaged at a rapid rate. If done quickly enough, the integration time constant of your visual system will low-pass filter the light, and it will simply appear to be dimmed. Equation 1 gives the ideal control-to-output transfer function:

$$I_{\text{LED}} = I_{\text{F(MAX)}} \cdot \text{DC}_{\text{PWM}} \quad (1)$$

where I_{LED} is the average LED current, $I_{\text{F(MAX)}}$ is the average value of the undimmed current, and DC_{PWM} is the duty cycle of the PWM signal. Depending on how the PWM is implemented, there are several factors that can cause the actual transfer function to deviate from this ideal value.

LED drivers mainly use two techniques to implement PWM_O dimming, namely “enable” and “switched” dimming. In the case of enable dimming, the entire driver is turned on and off in order to modulate its output, as shown in Fig. 1b. The current within each pulse ramps slowly to and from zero as the driver is enabled and disabled. The transient speed of the driver will determine its ability to resolve the PWM_I command. From this curve, it is straightforward to see how the control-to-output function will become increasingly nonlinear as the PWM_O duty cycle approaches zero. First, as the ramp time becomes an ever more significant portion of the total on time, the error induced in the average current signal will likewise become more pronounced. Secondly, it is even possible that at low duty cycles the current is not provided enough time to reach its proper peak value; instead, it is ramped up and then immediately backed down resulting in triangular pulses of current.

If these distortions need to be avoided per the requirements of the application, *switched* dimming is suggested, as shown in Fig. 1c. In the case of switched dimming, current flow to the LEDs is abruptly interrupted with the use of a series or parallel switch, likely in the form of a metal oxide semiconductor field effect transistor (MOSFET). Switched dimming results in extremely sharp edges with very short rise and fall times. The control-to-output function in this case will be as close as possible to the ideal theoretical value given by Eq. 1. A detailed discussion on

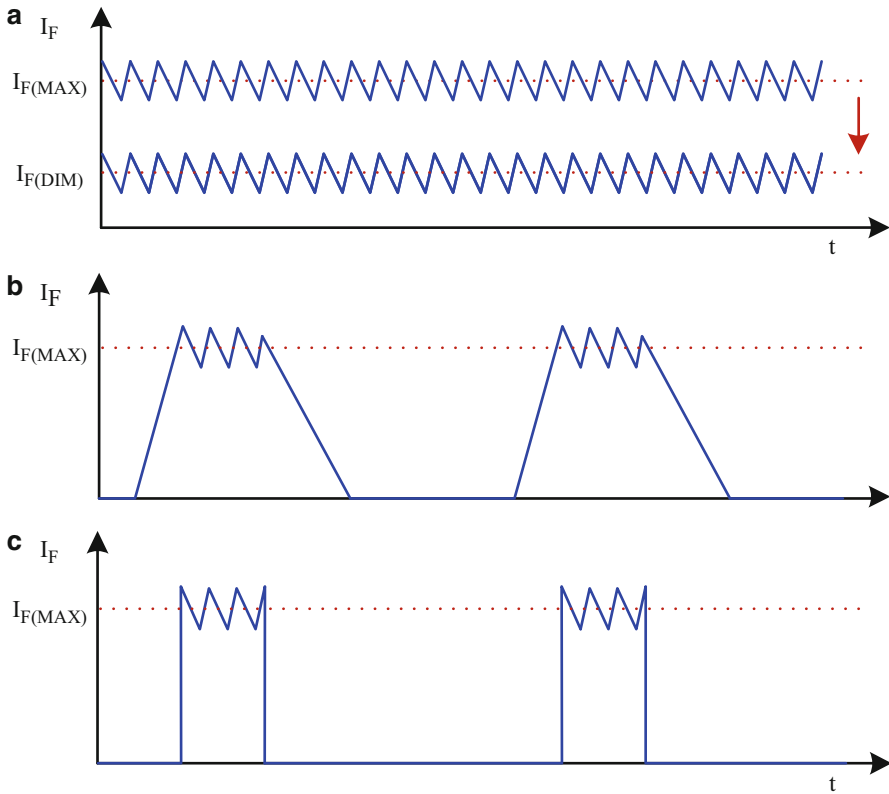


Fig. 1 Output LED current dimming examples. (a) Analog_O, (b) enable PWM_O, and (c) switched PWM_O

implementation of switched PWM_O dimming is provided in the section titled, “Implementing Switched PWM_O Dimming.”

There are many trade-offs to consider when determining the proper way to modulate the light, and care should be taken to determine the best method. Analog_O dimming has the advantages of simplicity, zero flicker, improved LED efficacy when dimming, and low electromagnetic interference (EMI) and noise generation. It has the disadvantages of slow modulation speed, limited contrast ratio (typically 100:1 or less), and control-to-output nonlinearity (especially when deep dimming). In addition, matching of light output between various LEDs in the same string will degrade with dimming. The effect can be pronounced to the point where at very low levels, some LEDs will actually appear off, while others are still generating perceptible light. Finally, if color fidelity is extremely important, analog_O dimming is probably not the best choice, as the dominant wavelength being produced and the level of phosphor excitation (in the case of white LEDs) will vary with dimming level. The effect can even be so pronounced that it makes white LEDs appear more yellow.

PWM_O dimming has the advantages of high contrast ratio (5,000:1 or greater is readily achievable), modulation speed, and consistent spectrum. It has the disadvantages of increased EMI/noise generation and can be more costly to implement, proportional to the performance level desired. The frequency at which modulation needs to occur to avoid visible flicker is debatable and largely dependent on the application. It is difficult for the human visual system to detect any flicker over 200 Hz when no motion is present. When motion is present or digital photography is likely to occur in the space, higher frequencies (in excess of 5 kHz) are desired in order to avoid the stroboscopic effect. As an interesting experiment, to observe the effect directly, simply PWM_O an LED at 200 Hz and physically move it quickly in a circle, and it will become immediately obvious that the light is not on 100 % of the time.

It should also be noted that contrast ratio and PWM frequency (at both the input and output) are related. For example, if a 10,000:1 contrast ratio is desired at a PWM frequency of 1 kHz, both the PWM_I and PWM_O signals will need to be capable of resolving the resulting 100 ns signals without significant distortion. This requirement will have significant consequences for the PWM_I signal generator, likely in the form of microcontroller clock resolution. Even more challenging, it will also have consequences for the current generator as determined by the LED driver's ability to both accept and resolve the PWM_I signal and drive the PWM_O current signal to the LEDs without distortion. As previously mentioned, switched dimming should be used if high fidelity is required. See the section titled "Implementing Switched PWM_O Dimming" for further information.

A summary of the comparison metrics between analog_O and PWM_O dimming can be found in Table 1.

If extremely high contrast ratios (in excess of 5,000:1) are desired, a combination of analog_O and PWM_O dimming can be used. The resulting contrast ratio can be calculated according to Eq. 2:

$$\frac{1}{DC_{PWM} \cdot P_{ADIM}} : 1 \quad (2)$$

where DC_{PWM} is the duty cycle of the PWM and P_{ADIM} is the percentage of analog dimming applied. For example, if a 0.1 % PWM is superimposed on a signal reduced

Table 1 Comparison of analog_O and PWM_O dimming

Metric	Analog _O dimming	PWM _O dimming
Speed of response	>10 us	<10 ns
Contrast ratio	<100:1	>5,000:1
Matching under deep dimming		✓
EMI/noise	✓	
Light spectrum fidelity		✓
High-speed motion/photography/video	✓	
Control-to-output linearity		✓

to 20 % via analog dimming, the actual light output will be approximately 0.02 % of its maximum level for a resulting contrast ratio of 5,000:1.

Some LED drivers accept only analog_I dimming inputs, which can be challenging if they are being produced by a PWM_I generator like a microcontroller. This is typically not a problem as the PWM_I signal can typically be filtered using common and inexpensive techniques. This filtering will, however, introduce time delay, or phase shift, into the control-to-output transient relationship.

Efficacy Improvement Under Analog_O and PWM_O Dimming

While the luminous flux per forward current relationship is very linear, it contains nonlinearities due to an effect called LED “droop.” Droop is the decrease in relative luminous efficacy observed as driving current increases, even for LEDs maintained at the same junction temperature. The effect is shown in Fig. 2.

The figure shows that when this particular LED is operated at 25 % of its maximum current level, it is actually 50 % more efficient at converting current to light. This is the main reason that it is typically unwise to drive LEDs at their maximum rated current if efficacy is a primary goal of the end system. Regardless, the end result of the droop effect is simply that analog_O dimming can be used to provide higher overall system efficacy than PWM_O dimming. As an example, an LED dimmed to 1 A_{DC} will be brighter than the same LED running 2 A at a PWM duty cycle of 50 %, even though their average current and thermal load are equivalent.

It should also be noted that LED efficacy will also improve somewhat due to the decrease in temperature that is a likely result of dimming, regardless of modulation method. Efficacy improvement of approximately up to 10 % is possible.

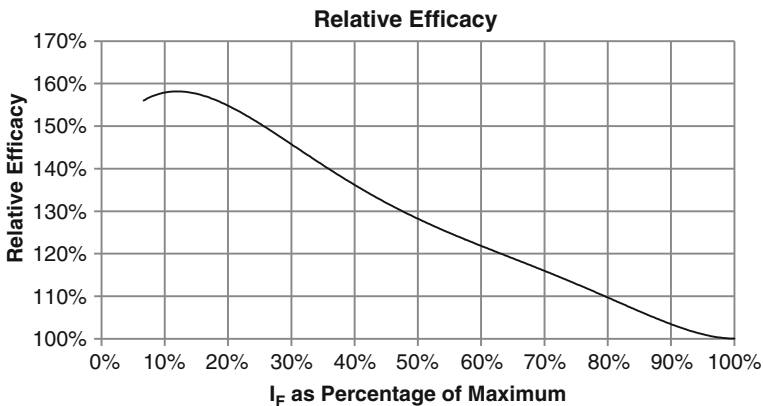


Fig. 2 The effect of droop on luminous efficacy

Color Fidelity Under Analog_O and PWM_O Dimming

Analog_O dimming causes shifts in the spectrum of both white and color LEDs. Many applications require high levels of color fidelity, including even some general lighting applications. Having some idea as to the level of shift that is likely to occur is an important factor for the lighting designer. Note that the figures quoted here are for reference only; the technical documentation for the particular LEDs under consideration should be consulted for more accurate and detailed information. For phosphor-converted white LEDs, the shift can be as high as 0.02 in “x” and 0.05 in “y” on the 1931 CIE chromaticity diagram when varying the current over a range of 100:1. The spectral shift of white LEDs is described in detail in Dyble and Narendran (2005). It is interesting to note that this shift occurs due to a shift in the blue LED’s dominant wavelength of just a few nanometers!

For color LEDs, a shift in the dominant wavelength of 5–10 nm should be expected for red, green, and blue. Spectral shifts of color LEDs are outlined very clearly in Gu and Narendran (2006). Again, the technical documentation for the respective LEDs should always be consulted.

PWM_O dimming solves many of the color shift issues associated with analog_O dimming, but some shift is likely regardless of the type of dimming implemented. Given that the LEDs are likely the most significant contributor to heat, significant temperature variation is likely to occur under all dimming conditions. For white LEDs, expect a shift of up to 0.005 in both “x” and “y” over a reasonable temperature range of up to 100 °C.

Implementing Switched PWM_O Dimming

High-speed, switched PWM dimming is implemented with one of two methods, series and shunt, as shown in Fig. 3. As will be described, the proper choice depends on the type of LED driver and its respective output configuration.

In the case of series switched dimming, the current flow is directly interrupted with the series switch. This arrangement should be used if the output of the LED driver is low impedance, likely due to significant amount of output capacitance in parallel with the LED load. Driver topologies that are able to generate output voltages greater than their input voltage typically contain low impedance outputs. If the reader is familiar with switch-mode power supplies, boost, buck-boost, and flyback topologies will all have low impedance outputs. One note of caution when attempting to implement series dimming concerns the behavior of the driver when the switch is open. If the driver has not been commanded to be off, it will sense that its output current is zero, far below the regulation set point, and the driver will therefore be doing everything it can to raise it. At a minimum, this behavior will cause the LED current to be incorrect when the switch is closed. In the worst case, the behavior will cause driver components to become damaged. As such, enable command and series switched dimming are often combined.

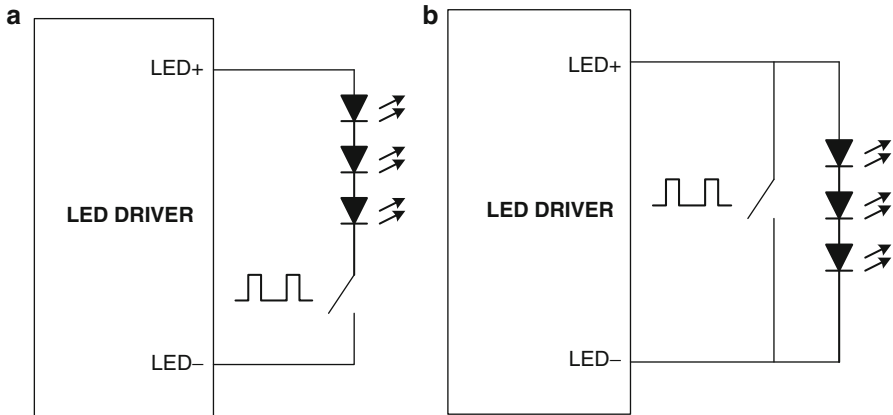


Fig. 3 Switched dimming, (a) series and (b) shunt

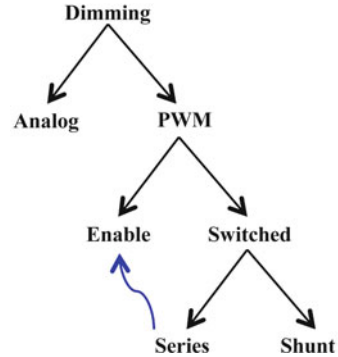
In the case of shunt switched dimming, the output of the driver is actually shorted out via the switch. While this may seem strange at first, if the driver is a relatively adequate representation of a current source, the output current is independent of the output voltage and should not be greatly affected. Said another way, if the driver's output is high impedance, it will likely be able to support a short across its output. The buck regulator is the most common high-output-impedance current source, which is one reason it is commonly referred to as the ideal LED driver topology. While it may seem so at first, in the case where the driver contains a switch-mode power supply, the result is not all that inefficient as output power is essentially zero during the shorted condition. However, if the driver contains a linear-type regulator, efficiency will drop significantly and care must be taken so as not to overstress the driver as the excess power will be dissipated in the regulator's pass element. Care must also be taken when routing the current-carrying lines, be they wires or circuit board traces, between the switch and LEDs. Any inductance will result in delays between the switching action and actual LED current modulation and will likely cause overshoots and undershoots in both the current and voltage signals present in the system. It should be noted that all circuit board traces and wires are inductive, though printed circuit board layout and wiring techniques to minimize the amount of inductance are beyond the scope of this handbook.

As a note of caution, both linear regulators and buck switching regulators may have a significant amount of capacitance present at their output. The presence of this capacitance transforms a high impedance output into a low impedance output. As with all low impedance output drivers, the proper switched dimming solution will be series, not shunt.

In summary, a simple flow chart of decisions to be made when dimming can now be formed, shown in Fig. 4.

The curved arrow from series to enable is shown to remind the reader that enable dimming is often combined with series switched dimming in order to ensure the

Fig. 4 Dimming decision flow chart



driver does not continue attempting to deliver power to an open load when the series switch is disengaged.

Dimming Considerations for the System Integrator

Obviously, comprehensive knowledge of the concepts in this chapter will be important when designing an LED driver system from the ground up, but surprisingly they are equally critical when specifying an off-the-shelf driver as well. Challenging as it may be, it is worth the time and effort to determine how an LED driver actually reduces the current to ensure the method and performance are appropriate for the application. If the dimming method is not specified in any of the available technical documentation, one may still be able to make a determination by observing the characteristics of the LED current and comparing them against the examples shown in Fig. 1.

In some cases, one may desire to use a standard, off-the-shelf constant current driver but develop customized switched dimming circuitry. In order to do so, intimate knowledge of the driver's internal architecture may be required, specifically the configuration of the output circuitry. As previously discussed, a driver with a low impedance (aka capacitive) output should typically be dimmed with switched series dimming, often in conjunction with enable dimming. Conversely, a driver with a high impedance output should be dimmed with switched shunt dimming. As an additional layer of complication, it is even possible that a driver that is able to withstand a short across its output cannot handle a short to the ground or return path. It is also important to differentiate between survival and operating capabilities. While most LED drivers are designed to survive a shorted output condition, many are not designed to operate properly under such a condition. One key indicator to look for is a relatively consistent characteristic in the output current waveform when shorted. While its shape is likely to change somewhat given the significant difference in operating conditions between the on and off states, its overall shape and average value should not be severely altered. Also, it is likely good practice to measure the driver's temperature and input power during the shorted condition to ensure long-

term reliability and safety of the drive system. If the input power does not decrease very significantly, it will be important to understand where in the system that power is now being dissipated, as it is no longer in the LEDs. If high power is being dissipated in the shunt, those components will likely be damaged.

Before experimenting with drivers in these ways, be sure to have obtained a number of them, as damage is likely to occur. This is not something to be discouraged or intimidated by; working with power-processing electronics almost always involves breaking the circuitry in order to better understand the driver and its limitations. After all, one cannot know how reliable a system truly is unless the limits are well understood and verified by experimentation. Preparation is key to fruitful and enjoyable experimentation. Be sure to use proper safety procedures and equipment, especially eye protection. A handheld thermal imaging camera can be an absolutely invaluable tool, as it clearly highlights both the locations and amount of power dissipation. Again, have multiple devices ready, as debugging and fixing a damaged circuit can literally require days of work that could have been avoided. Finally, thoroughly comprehending all the technical documentation provided with a driver is the right first step when investigating its capabilities.

Introduction to Command and Control Systems for Dimming

To this point, this chapter has mostly concentrated on how LED driver currents are actuated in order to effect a dimming command. This section focuses more on how LED drive systems are commanded and controlled. A system-level perspective will be taken with less emphasis on low-level circuitry and algorithms. The section starts with a discussion of power line communications methods, focusing on high-voltage, offline, phase-based dimming, as this is currently a major topic and one of the more challenging aspects of solid-state lighting. It will then move on to discuss other major communications methods and the current and future system architectures that support them.

Power Line Communications (PLC) and Phase-Based Dimming

Communicating over AC and DC power lines has long been a holy grail due to the facts that wired communications tend to be more secure and reliable than their wireless counterparts and the power lines are already present in the system, which optimizes complexity and cost. That said, with the exception of phase-based dimming, PLC has thus far failed to garner significant market acceptance due to the following three complicating factors:

- (1) The structure of the power lines is not often known or is known to be complicated in that there may be switches, transformers, or other hardware in between two devices that are trying to communicate. Further, the owners of the power line hardware may not be amenable to the deployment of repeaters or

communication transformers on their assets. Some companies that deploy power line communications also offer RF-based bridges to alleviate some of these issues.

- (2) The power signals require modulation, often at high frequencies that require fast current and voltage transitions that can create EMI.
- (3) Because they are carrying power, the lines are often extremely noisy, lowering signal fidelity and hindering clear communications.

Phase-based dimming suffers greatly from factors (2) and (3) above, as well as a host of other issues. Note that phase-based dimming is also commonly referred to as triode for alternating current or “TRIAC” dimming, even though the main power-processing device in the dimmer may not actually be a TRIAC. The solid-state lighting industry has been forced to deal with phase-based dimming due to the massive installed base (multi-billions of units) and the simple market reality that consumers will usually not be willing to pay more for something that is perceived to work less well than the technology being replaced.

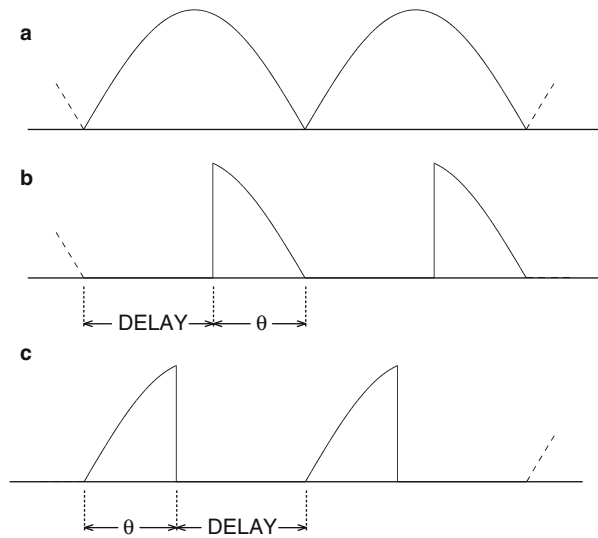
While all phase-based dimmers chop the incoming AC line, they may do so at either the beginning or end of the cycle. Dimmers that chop the leading edge are called “forward phase” dimmers, while those that chop that trailing edge are called “reverse phase” dimmers, as shown in Fig. 5.

The “on time” of the AC waveform is commonly referred to as the “conduction angle” and is described as an angle between zero and 180 degrees regardless if the dimmer is forward or reverse phase. Forward phase dimmers are based on a device called the TRIAC, a silicon-controlled rectifier-type of device that is capable of handling significant currents in a latched mode. “Latched” means that once the device has been triggered, it remains on until the current level approaches zero. Reverse phase dimmers typically use a MOSFET or insulated gate bipolar transistor (IGBT) as their main power-processing device and contain a variety of implementations in order to perform the timing functionality. Some higher end dimmers that have a ramp up and down function, or a touch pad interface, include a microcontroller.

Though this handbook focuses mainly on intelligent lighting, the problem of interfacing LED lamps to phase-based dimmers is so pervasive. The author believes that any chapter on dimming must include a discussion on the most common challenges. The majority of the challenges with TRIAC dimming are based in the nature of the TRIAC device itself, the components typically included in dimmer hardware, and the differences between incandescent and LED loads.

The timing circuitry of all dimmers needs current in order to function, often referred to as “bleeding” current. This current must be passed through the lamp even when the lamp is off and is usually on the order of 5–30 mA. This is very straightforward when dealing with resistive incandescent loads but far less trivial when implementing more complicated LED driver circuitry. Therefore, the LED driver must allow this current to pass relatively unhindered or the timing of the dimmer will be corrupted.

Fig. 5 (a) AC mains, (b) forward and (c) reverse phase-based dimming



The incandescent load is “heavy” in that it is inefficient and requires large currents to produce appreciable light. By contrast, typical LED-based loads require significantly less power than their incandescent counterparts, and the load is further reduced under dimming scenarios. The end result is that solid-state lamps are severely challenged in their ability to maintain “holding” current levels required for proper TRIAC operation. Also called “latching” current, this is the level of current that must be sustained in order for the TRIAC to remain in its latched or conducting state. If the latching current is not sustained to a high enough level, flicker occurs when the TRIAC then misfires, as it prematurely perceives a low level of current to imply the end of the AC line cycle.

Incorrectly perceiving the end of the line cycle can also be caused by the interaction between the dimmer and LED driver filter components. Unlike the resistive incandescent load, the overwhelming majority of LED drivers will contain a reactive input filter in order to achieve regional EMC regulatory compliance. The dimmer also contains such filtering circuitry for the same reason, and the two filters together form a resonant tank circuit that is excited by the dimmer’s firing action. If the resulting current waveform rings close to zero, it will cause the TRIAC to again incorrectly interpret the situation as the end of the AC line cycle, and conduction will terminate. Therefore, the driver’s input filter must be carefully tuned in order to dampen this behavior. This is also the main challenge when designing TRIAC dimmable drivers that are intended to operate on a “universal” input line, irrespective of global region. Different regions of the world maintain different EMC requirements and therefore the dimmers present very different filter characteristics. This makes the task of tuning the driver’s input filter very difficult without additional complexity and cost.

In order to achieve appreciable contrast ratios, TRIAC dimmable LED drivers typically perform an interpretation of the incoming waveform. There are mainly two types of interpretation, root mean square or “RMS” and “angle” sensing. With RMS interpretation, the RMS value of the incoming waveform is used to inform the driver as to the appropriate level of light. The challenge is that any phenomenon that disturbs the RMS value, be it normal line variation or noise, will also drive a change in light output. Incandescent sources also suffer from this issue, but the long thermal time constant of the filament typically filters out line noise. With angle interpretation, the driver actually attempts to measure the conduction angle and determine the appropriate amount of drive current according to a mapping function that describes the light versus conduction angle relationship. The shape and limits of the mapping function are dictated by NEMA standard SSL 7A-2013, which outlines the requirements for basic compatibility. The main issue with this method is that any line disturbance that affects the conduction angle, especially a TRIAC misfire, will likely cause conduction angle misinterpretation and, in many cases, flicker.

As in most cases with engineering for the mass market, the challenge lies not in simply accomplishing a task but accomplishing it at minimal cost while maintaining adequate levels of performance and reliability. Phase-dimmable LED drivers that perform power conversion in two stages fare far better than those that perform the entire conversion in a single stage. The first stage of a two-stage system can act as a buffer to condition the line while also providing a more constant, consistent power source for the second stage that is directly driving the LEDs. But, two-stage architectures are inherently more costly and often less efficient.

Outside of the phase-dimmable arena, semiconductor vendors of both analog and digital technologies have focused on supporting PLC, which has led to significant advances in fidelity and robustness. Several protocols based in power line communications are making progress in market adoption. Some of them include X10, Universal Powerline Bus by Powerline Control Systems, Inc., Insteon, LonWorks by Echelon, HomePlug, Intellon, CEBus, Yitran, Netricity, Corinex, and Watteco, among others. Companies that focus on building automation and control networks have also developed proprietary solutions, as have some of the larger lighting companies. The IEEE Power Line Communication Standards Committee, formed in late 2011, guides industry activities, as does the European Committee for Electrotechnical Standardization or CENELEC.

Low-Voltage AC Systems

Low-voltage AC systems typically operate from 12 V_{AC}, which implies a step-down transformer is implemented in between AC mains and these systems. Regardless of the type of transformer, it is placed after the dimmer so as to be in between the dimmer and lamp load. Step-down transformers are of two types, magnetic and electronic. Magnetic transformers are standard transformers in every way; there are no electronics but instead contain only multiple windings wrapped around a core of ferromagnetic material. Though much larger in size and weight when compared

against their electronic counterparts, they are extremely simple and reliable. Electronic transformers are much smaller and lighter but contain switching electronics in order to perform the step-down conversion, which makes them inherently less reliable.

The most common type of electronic transformer is based on a self-oscillating circuit architecture referred to as the “ringing choke” converter. As it pertains to driving an LED load, the issue with the ringing choke converter is that it requires a minimum load in order to switch properly. Further complicating the situation is the fact that by far the most common low-voltage AC lamp is the small MR-16 form factor. An additional or “dummy” load to the lamp is likely impossible given the already challenging mechanical and thermal constraints associated with this form factor.

As a result of the aforementioned complicating factors, retrofitting phase-dimmable, low-voltage AC systems with LED lamps is difficult at best. All of the previously discussed issues associated with phase dimming remain, along with the new challenge of attempting to maintain the electronic transformer’s minimum load requirements under dimming conditions. As of this writing, it is the opinion of the author that there is currently no high-performing, robust, and inexpensive solution on the market today, as phase dimming of low-voltage AC solid-state lamps remains one of the industry’s most significant retrofit challenges.

0–10 V Dimming

Also referred to as “4-wire low-voltage” dimming, this protocol uses dedicated wires to communicate a 0–10 V_{DC} signal from the dimmer to a plurality of lamps. The protocol is used heavily in fluorescent lamp dimming, as specified by the NEMA ANSI C82.11 standard. But, it has been increasingly adopted in solid-state lighting applications given the lack of other available and widely accepted dimming standards.

The LED driver must be designed for the 0–10 V standard by providing a current between 10 μ A and 2 mA to the dimmer. The dimmer, in turn, sinks the current from one or more drivers and presents a voltage proportional to the slider position, but relatively independent of the number of connected drivers. According to the NEMA standard, the lamp scales its output so that at 10 V the controlled light should be at 100 %, and below 1 V the light should be at its minimum value. Note that the minimum value should not be zero light output or “off,” as the standard clearly states that there shall be stable light output for all voltages between zero and 11 V. The question of galvanic isolation of the communication lines can be a confusing one, and it is best to work closely with safety regulatory agencies such as ETL and UL to ensure the final product will be compliant. The communication lines are likely to be routed in the same conduit as AC mains, which require the driver to meet safety compliance for high-voltage, offline systems. The determining factor is likely the classification of the driver and the resulting classification of the luminaire. A class 2 driver requires a low-voltage output protected from high voltage, whereas as a

class 1 driver does not. The driver must be able to withstand polarity reversal and should output a minimum level of light under such a fault condition.

One common question is how the 0–10 V_{DC} signal should be translated to something the internal driver circuitry is able to interpret. The incoming signal is clearly an analog_I-type signal. The translation required is determined by (a) what the desired output current modulation method is, be it analog_O or PWM_O, and (b) how the internal circuitry receives dimming commands in order to effect that outcome. If the internal circuitry requires an analog_I-type input command, a relatively straight-forward scaling may be possible. But, if the input command is the more common PWM_I type, the design is more challenging in that a ramp and comparator are required to perform the translation.

Intelligent Lighting Systems

Given the aforementioned complications when retrofitting solid-state lamps into conventional systems, the industry is eager to move to more modern, capable systems. Solid-state lamps can readily be interfaced to a microcontroller via PWM_I, a filtered version of PWM_I that generates an analog_I control signal, or both. Much higher performance can also be achieved not only in terms of contrast ratio and flicker but also overall capabilities as well. The range of applications is limitless, from simple dimming to scheduled and sensor controlled lighting, color mixing, utility peak load shedding, gaming, plant growth, and medical applications. The list is literally endless.

For the purposes of this chapter, any system with capabilities beyond basic dimming shall be referred to as “intelligent.”

System Architectures for Intelligent Lighting

As long as the industry is stuck in retrofit mode, we will continue to deploy systems that are far from optimal in terms of both performance and cost. One of the most basic advantages of LED lighting is that most light engines are low voltage, often safety extra low-voltage (SELV) systems that operate with relatively low current levels. This implies that, unlike all other high-efficiency lighting technologies, LED loads are very easy to power remotely. This simple fact offers the opportunity to completely reengineer lighting systems as we know them.

Basic system architectures can be thought of as existing on a spectrum pertaining to their level of integration. Today, most intelligent solid-state lighting systems follow the discrete approach shown in Fig. 6, wherein each lamp is a wholly independent entity.

In the example shown in the figure, each lamp contains its own 25 W LED driver, likely driven directly from AC mains. Each lamp also contains its own hardware to support intelligent functionality, represented in Figs. 6, 7, and 8 by a radio tower symbol. Such hardware will likely include a communications interface and system

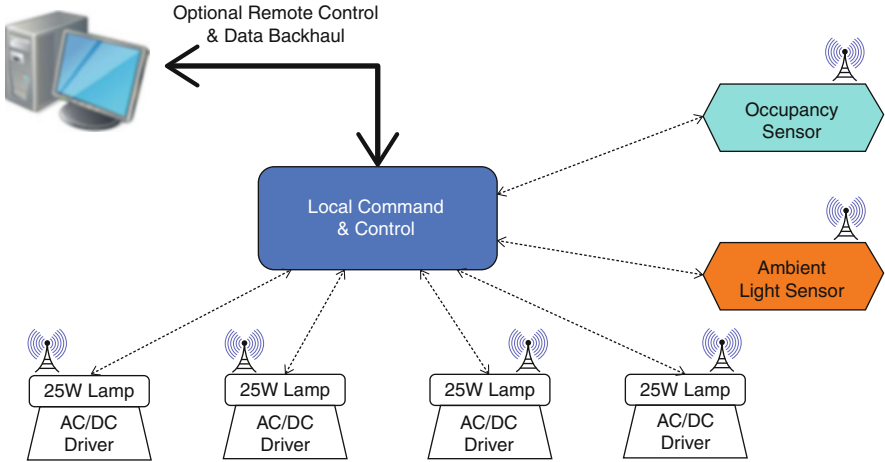


Fig. 6 Discrete approach to intelligent lighting

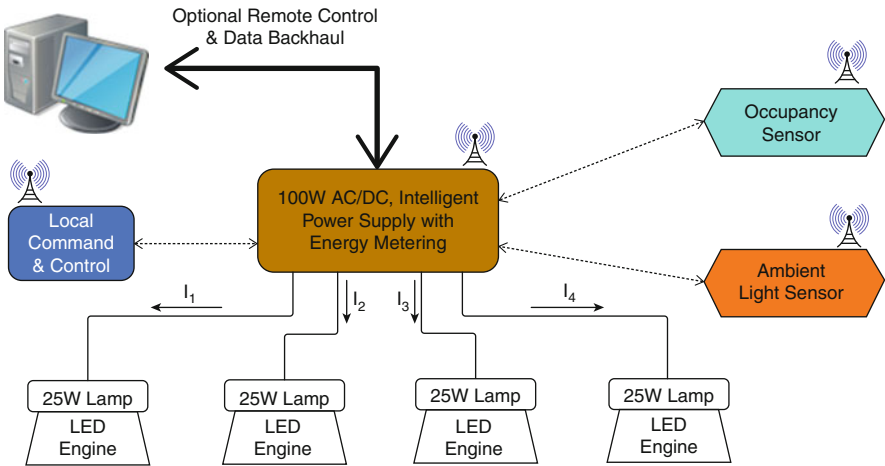


Fig. 7 Balanced approach to intelligent lighting

microcontroller and may include an energy meter, temperature sensor, and other functions depending on the application.

On the other end of the spectrum is a fully integrated approach wherein each lamp is not much more than an LED bank, heat sink, and optics, and the majority of power and control is contained in and driven from a central location, such as an IT closet. Again, this is possible because LEDs are low-voltage, low-current devices.

A more balanced, hybrid approach, similar to what several companies such as Redwood Systems are currently deploying, is likely a more optimal solution. One such example is shown in Fig. 7.

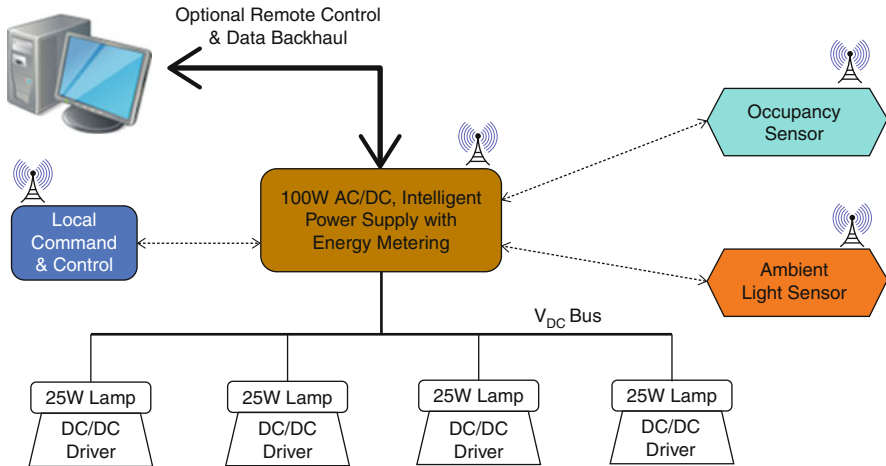


Fig. 8 DC bus approach to intelligent lighting

In the balanced approach, both power conversion and intelligent functionality are locally centralized. In the example shown in Fig. 7, a 100 W AC/DC converter drives four SELV, low-current, class 2 output channels. In contrast to the discrete approach, a 100 W power supply is both more efficient and significantly cheaper than four 25 W power supplies. With respect to intelligent functionality, this system architecture is also significantly more cost optimized as the number of microcontrollers, radios, and sensors is greatly reduced without sacrificing performance or capabilities. Communications are still performed between the 100 W system basis power supply and both remote sensors and localized command and control points.

Another variant of the balanced approach would be to implement a DC power distribution grid, as shown in Fig. 8.

In the example shown in Fig. 8, a DC voltage is routed to each lamp in contrast to the example in Fig. 7 wherein a DC current is routed to each lamp. A DC/DC LED driver is therefore required to convert the DC bus voltage to the constant current required by the LED load. An obvious question is how the driver is informed of the desired light level, and herein lies an opportunity for innovation. The 100 W AC/DC converter could communicate over the DC bus via a form of DC power line communications. PLC is much more straightforward in this case because the power line is localized, DC, and completely under the control of the system basis power supply. The EMerge Alliance (<http://www.emergealliance.org>) aims to drive the concept of DC power distribution into commercial building installations. Compliant products supporting DC distribution are already available. DC distribution is ideal for solid-state lighting because the majority of the cost, complexity, and difficulty in developing LED drivers involve the high-voltage AC/DC conversion and the myriad regulatory compliance issues (mostly safety and EMC) that have to be meticulously engineered for.

Communications Infrastructures for Intelligent Lighting

Selecting the appropriate communications infrastructure to support a particular application is no easy task given that literally hundreds of options are available. What is most important is that the designer comprehends a set of basic considerations that need to be addressed. Below is such a set of considerations but is by no means exhaustive.

- Wired or wireless?
- Transmission distance between nodes
- Unidirectional or bidirectional communications?
- Open standard or proprietary?
- Interoperability (with other communications standards)
- Number of nodes supported
- Mesh, tree, star, or hybrid network configuration
- Commissioning requirements
- Robustness/redundancy: Is the network self-forming? Self-healing?
- Frequency of operation (per region)
- Data rate and latency
- Security level and procedures
- Cost of hardware, software, and installation
- Maintenance needs and costs
- Support for data backhaul
- EMC considerations (interference and susceptibility)

The selection process involves three main domains: the physical layer, protocol, and application. Each element is not entirely independent as many infrastructures dictate all elements. For example, the ZigBee personal area network protocol is a wireless, IEEE 802.15 physical layer organized in a mesh network. The ZigBee Home Automation Profile provides an application layer for lighting controls. In this respect, ZigBee is an example of a complete solution.

The selection of the communication infrastructure is relatively independent of dimming considerations. The majority of these systems will require a microcontroller in order to manage the communications, sensor interfaces, and other system management and diagnostic functionality. Consequently, it is very likely that the control input will be a PWM_I output from a microcontroller. As discussed previously, the PWM_I signal can either be used directly or filtered to create an analog_I signal if analog_O dimming is required and the LED driver does not contain a PWM_I to analog_O conversion function. One other element that needs to be considered is latency. Users will likely have a very poor experience if there is significant latency between the control input they provide and the actual light modulation. While allowable latency will depend heavily on the application, more than about 0.5–1.0 s of delay is about the maximum that is typically accepted by users.

Conclusions

Simply put, the LED is the most efficient, reliable, and controllable light source ever created. While increasingly economical due to gains in LED lumen-per-watt efficacy, implementing solid-state lighting without dimming is akin to slowly driving a Formula 1 race car in a straight line. The majority of applications, be they municipal, commercial, residential, biomedical, entertainment, or industrial, could benefit from dimming or the additional functionality enabled by dimming, such as color control.

The intelligent lighting space, especially as it pertains to communications infrastructures, is still in its infancy in terms of what the final system architectures will be. In time, a limited set of combinations of capability, complexity, cost, and security will emerge to fit applications where it makes economical sense for adoption to occur. The market winners will be those companies that can most quickly determine, develop, and deploy the optimal solutions for the right applications in the right global regions.

A simple Internet search will reveal volumes of additional information far in excess of that contained in this chapter. The objective herein was not to cover every topic in minute detail but instead to inform the reader as to the most common trade-offs to consider and complications to avoid when developing dimmable solid-state lighting systems.

References

- Dyble M, Narendran N (2005) Impact of dimming white LEDs: chromaticity shifts due to different dimming methods. SPIE Proceedings 5941
- Gu Y, Narendran N (2006) Spectral and luminous efficacy change of high-power LEDs under different dimming methods. In: Sixth international conference on solid state lighting, proceedings of SPIE 6337, 63370J
- Miller Naomi (2013, Nov) Managing risks: Flicker. http://apps1.eere.energy.gov/buildings/publications/pdfs/ssl/miller_flicker_portland2013.pdf. Retrieved 5 Dec 2013
- Poplawski M, Miller N (2011, Nov) Exploring flicker in SSL: what you might find, and how to deal with it. In: Paper presented at the ArchLED conference, Chicago. <http://www.architecturalssl.com/sslinteractive/media/293/ArchLED%20flicker%20presentation.pdf>. Retrieved 5 Dec 2013

Conventional IR and Ultrasonic Sensor Systems

J. P. Steiner

Contents

Introduction	466
History of Passive Infrared Sensing	469
History of Ultrasonic Sensing	470
Black Body Radiation	470
Material Behavior	471
Lens	474
Detecting Infrared Radiation of Objects	478
Responsivity	480
Signal Chain	481
Noise	483
Temperature Effects	489
System Response	491
Testing	495
Applications	497
Ultrasonic Sensing	501
Ultrasonic Field Patterns	502
Doppler Shift	504
Propagation Parameters	506
Diffraction	507
Reflections	509
References	512

Abstract

Sensors have been used for decades for security, control of lighting, personal convenience and the control of heating, ventilation and air conditioning. For lighting control, sensors are widely used to meet energy codes in buildings and save significant energy. This chapter will discuss the fundamental science used in

J.P. Steiner (✉)
Lutron Electronics Inc., Coopersburg, PA, USA
e-mail: jpsteiner@lutron.com

conventional passive infrared and ultrasonic sensors. Passive infrared sensors detect the motion of warm bodies by detecting the heat they radiate. Ultrasonic sensors detect the motion of a body by measuring the Doppler shift of reflected ultrasonic waves transmitted by the sensor.

Introduction

Sensors have been used for many decades to reduce the amount of energy consumed in the commercial and residential spaces. The typical application uses the sensors to control the lighting in the space, but some are used to reduce the heating and cooling load. Referring to Fig. 1, lighting is the largest contributor to the electric power consumption in commercial buildings. The efficacy of the sensor depends strongly on the application. For example, the performance of a sensor used in a private office depends heavily on the occupancy pattern of the office and its occupant's behavior. One occupant may be very energy conscious and turn off the lights whenever she leaves, while another occupant may never turn off the lights when he leaves. Some private offices may be occupied continuously during the day, while other offices may have occupants that frequently attend meetings away from their office.

The problems that sensors solve are related to occupant behavior. Often, the occupant is not educated about the societal benefits of saving energy and its relationship to minimizing the use of lighting. Alternatively, occupants may not feel that they actually have control of lighting because it is not "theirs" to control. Sometimes they might just forget to turn off the lights. In any case, having controls that automatically turn the lights off will save energy.

Often designers retrofitting a building for energy savings will instrument various area types with data loggers to determine occupancy patterns and hence the potential savings afforded by an occupancy sensor. The typical logger will measure both the occupancy status of the area and the state of the lighting (*ON/OFF*). Loggers will

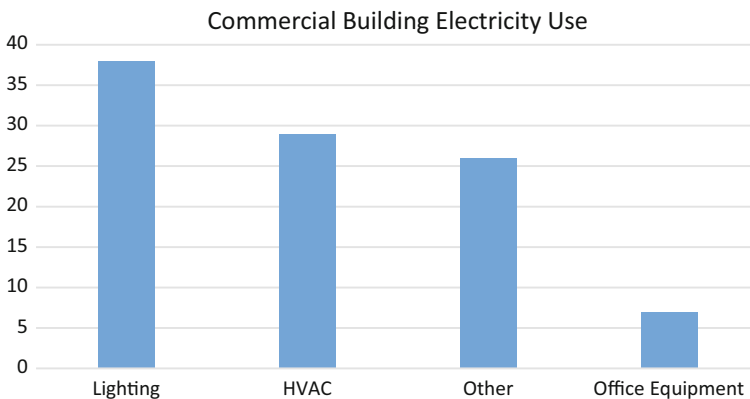


Fig. 1 Percentage of electricity use in a commercial building

collect data over a week or two along with time stamps. Data will then be compiled into a report indicative of the savings potential for the associated areas.

There has been some disagreement about how much energy is actually saved by occupancy sensors. In the past couple of decades, results have been published stating the potential saving. However, because occupant behavior varies, these results can only be statistically applied. Two of the more complete studies (Maniccia et al. 2000; VonNeida et al. 2000) examined 158 rooms with the following breakdown:

- 42 restrooms
- 37 private offices
- 35 classrooms
- 11 break rooms
- 33 conference rooms

These rooms were studied using a data logger of the type previously mentioned. The study looked at:

- Day shift versus night shift, 12 h each
- Weekday versus weekend
- Time-out setting of the sensor

The results in Table 4 of VonNeida et al. (2000) describe average saving during the day and night for various sensor time-out settings. Table 2 of VonNeida et al. (2000) describes the average occupancy rate for these spaces. Table 1, columns 2 and 3, shows the averaged savings for all the time-outs tested. Table 1, column 4, shows the combined day and night occupancy rate for the spaces. This data confirms that using sensors will save a considerable amount of energy.

Sensors have a parameter called a time-out, which is the time period the sensor takes to go from occupied to vacant. When occupancy is detected, the sensor enters the occupied state and starts a countdown timer. If no motion is detected and the countdown timer reaches zero, the sensor enters the vacant state. If the sensor detects motion during the countdown time, then the countdown timer resets to its starting value. Sensors usually have time-outs that are adjustable with values ranging from a few seconds to 30 min. The shortest time-outs, of just a few seconds, are reserved for

Table 1 Room occupancy rates and average savings from using an occupancy sensor

Space type (%)	Average daytime savings (%)	Average nighttime savings (%)	Average daily occupancy rate (%)
Break room	45	22	24
Classroom	82	55	16
Conference room	67	44	11
Private office	68	33	18
Restroom	78	53	20

testing the sensor. Some spaces have large-scale motion and are only sporadically occupied, for example, a copy or print room. In this case, the appropriate time-out would be short. In other cases, such as a private office, the occupant may sit still for extended periods of time, perhaps quietly reading. In this case, the time-out should be longer because the probability of sitting completely still diminishes as time passes. Time-outs that are too short and turn lights out accidentally are a major source of annoyance with sensors. As mentioned, Table 4 of VonNeida et al. (2000) describes savings for various sensor time-out settings. While it is tempting to maximize savings, sacrificing occupants comfort and reduced productivity may not be worth the additional savings.

Caution should also be exercised on reducing the time-out depending on the lighting type. Fluorescent lamps can have their life reduced significantly if switched too often (see Narendran et al. 2000; <http://hewilliams.com>, 2014). Other types of lighting such as LED, incandescent, and halogen do not share this characteristic and can be switched as often as needed. If the lighting is HID, it is critically important to not have too short of a time-out since these lights often need many minutes to turn back on, if accidentally shut off.

The energy savings associated with using sensors is the reason sensors are required by energy codes. The three prominent codes exist in the United States: ASHREA 90.1 2013, Title 20/24 2012, and IECC 2012. The last one, IECC 2012, is an international energy code. In general, these codes require either a vacancy sensor, a partial on occupancy sensor, or an occupancy sensor depending upon the space. A partial on occupancy sensor turns the lights on no more than 50 %. The requirements for sensors in IECC 2012 are shown in Table 2.

Sensors are categorized into two varieties: occupancy and vacancy sensors. Their principles of operation are identical in that the fundamental sensing technology is the same. For an occupancy sensor, the load will automatically turn on and after some time-out period will automatically turn off if occupancy is no longer detected, whereas for a vacancy sensor, the person entering the space must manually turn the load on and once the sensor no longer detects occupancy, the load will automatically turn off when the time-out period expires.

Passive infrared (PIR) is the most widely used sensing technology because it is the simplest and least expensive to manufacture. Furthermore, the circuitry can be implemented using low-power techniques, which lends itself to battery operation or

Table 2 IECC 2012 sensor requirements

Sensor type	Space type							
	Classroom, lecture hall, training room	Conference, meeting, multipurpose room	Private office < = 250 sq. ft.	Open office, > = 250 sq. ft.	Corridor	Restroom	Stairwell	Storage room
Occupancy				X	X	X	X	
Vacancy or partial on occupancy	X	X	X					X

energy-harvesting schemes. The next most prevalent technology is ultrasonic (US). Ultrasonic technology uses significantly more power and consequently is not used in low-power applications. It is also more expensive to implement and is only used when the benefits of the technology are needed. Both these technologies have their strengths and weaknesses and are often combined to create a hybrid sensor, called a dual technology sensor (DT), that outperforms sensors with just a single technology.

Both the PIR and US sensors require motion to detect a person. People standing or sitting motionless will not be detected. The fundamental physics behind the PIR sensor is that it detects heat in motion and uses the difference between the ambient background temperature and the person or object being detected. The fundamental physics behind the US sensor is detection of the Doppler shift in the ultrasound caused by a moving object. The US sensor transmits a continuous wave signal which is reflected from the surroundings. If an object is moving, the transmitted signal reflecting from the object undergoes a frequency shift (Doppler shift). One key difference between these two technologies is that one is passive and the other active. The PIR sensor does not transmit any signal; it just passively watches the temperatures in the space. In contrast, the US sensor actively transmits a signal.

History of Passive Infrared Sensing

Keller (2000) provides a brief history of the development of passive infrared (PIR) motion. The very first PIR detectors were designed for intrusion detection. The original concept was developed by Herbert Berman (1972) in which he used a segmented spherical mirror to create discrete detection zones. The use of discrete detection zones allowed for modulating the infrared onto the detection element. Berman also used a small thermistor to detect infrared energy and germanium substrate with dielectric coatings to create an optical filter with an optical band pass from 4.5 to 20 μm . The spatial modulation caused modulation of the electrical signal produced by the thermistor. The electrical signal was then filtered by a band-pass filter tuned to respond to signals produced by a human walking through the detection zones. These developments made it practical to produce low-cost intrusion detectors and to some extent are found in modern PIR sensors.

Unfortunately, early sensors were subject to false alarms from a variety of sources. Early electronics had a phenomenon called “popcorn” noise which is a large, short-duration pulse that resembles a detected PIR signal. Further sources of interference are caused by environmental factors such as air flow, mechanical shock, and vibration. These problems persist today in modern PIR sensors, and Berman’s second patent (Sprout and Berman 1975) addressed some of these problems by creating an electronic circuit that combated these noise problems. In their circuit configuration, they use two infrared sensitive elements of equal size in two electronic channels such that they generate equal but opposite signals. An event is detected only if two signals of opposite polarity are generated simultaneously. An analogous principle is used in modern sensors.

During the 1970s, the use of pyroelectric materials in intrusion detection systems was becoming common (Rossin 1973). In this patent, it is taught that performance of the detection system can be improved by placing two elements in series opposition. Any interference that causes a signal to develop on both active detection surfaces generates an equal but opposite signal and thus cancels out. Also, around this time ceramic-based pyroelectric materials were developed (Liu 1974), which are low cost and require simple manufacturing steps to create detectors. Eltec took a step forward by creating a differential detector that minimizes the buildup of static charge. This static buildup created “popcorn” noise when it discharged, causing false detection. In the early 1980s, pyroelectric occupancy sensors began to be used to control lighting and other power devices to save energy (Blissett and Dunbar 1982). By the mid-1980s, pyroelectric detector controlled by light became common (Philips 1985).

History of Ultrasonic Sensing

Ultrasonic occupancy sensing has its roots in SONAR. The earliest acoustic method traces back to the turn of the twentieth century when a patent was filed for an echo location system used to detect icebergs shortly after the Titanic sunk. During World War I, the technology advanced because it was needed to detect submarines. Near the end of World War I, the British developed sensor using quartz piezoelectric transducers. Doppler RADARs were developed during World War II and are the basis for the Doppler in SONAR. The earliest occupancy control for lighting first patented was filed by Ravas of Westinghouse in 1967 (Ravas 1969). This patent used the Doppler effect to detect motion and control the lights in response to motion.

Black Body Radiation

As mentioned PIR detectors respond to changes in temperature. In the following, details about the physics of the PIR detectors and sensors will be examined. From a high level, it is necessary to understand the source of the infrared radiation, optics necessary to focus the energy onto the detector, and the operation of the pyroelectric detector itself. Understanding these basic principles will lead to a better understanding of the detection system itself.

All objects having a temperature greater than 0°K (−273.15 °C) emit radiation. A perfect black body radiator emits electromagnetic radiation according to Planck’s law (Everett 1998). The radiation emitted through the hemisphere in front of the black body is given by

$$W_{\lambda} = 2\pi hc^2 \frac{1}{\lambda^5 \left(e^{\frac{hc}{\lambda T}} - 1 \right)} \quad (1)$$

In Eq. 1, the variables are:

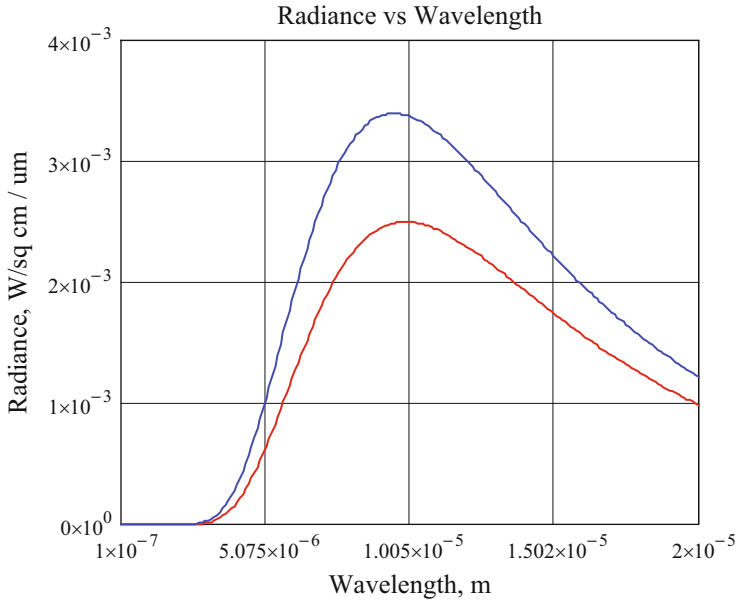


Fig. 2 The spectral radiant emittance of a human (blue) and a room at 72 °F (red)

W_λ is the spectral radiant emittance in W/cm²/micron.

h is Planck’s constant.

k is Boltzmann’s constant.

c is the speed of light.

λ is the wavelength in microns.

T is the absolute temperature in K°.

This equation is plotted in Fig. 2, the background having a temperature of $T = 295^\circ\text{K}$ (72 °C) and a person having a temperature of $T = 306^\circ\text{K}$ (91 °F). These are the typical temperature of a room and the temperature of a human’s skin (Elert 2014), respectively. Examining these curves shows that they both have their peaks in the region of the electromagnetic spectrum from 8 to 14 μm . The region from 8 to 14 μm is referred to as long wavelength infrared radiation (LWIR). This result suggests that using this portion of the spectrum is best for detecting human movement and is the basis for PIR occupancy sensing. For the above reasons, occupancy sensors using PIR detectors are designed to only respond to differences in optical power in the range of wavelengths from 8 to 14 μm .

Material Behavior

A fictional perfect black body emits all of the optical power it absorbs, but this is not true for normal objects. A normal object, of the type of interest, only partially absorbs and emits the radiation. The ratio of the amount of the emitted radiation to

that emitted by a perfect black body at temperature T is a dimensionless quantity referred to as the emissivity and is denoted by ε .

Furthermore, when an object is at thermal equilibrium, the amount of energy it is absorbing is equal to the amount it is emitting. This is expressed as Kirchhoff's law for thermal radiation:

$$\varepsilon = \alpha \quad (2)$$

where ε is the emissivity and α is the absorptivity. In general, these quantities are functions of the radiation wavelength, $\alpha(\lambda)$ and $\varepsilon(\lambda)$, but for the purposes of this explanation, they can be considered constants as the range of wavelengths under consideration is narrow (8–14 μm). Basically, an object that is a good absorber is also a good emitter. So, a good reflector is a poor emitter. A black body absorbs all radiation that is incident so $\alpha = 1$ which also means $\varepsilon = 1$. A gray body has an α and ε that are less than one. Good reflectors have an α and ε that are close to zero.

In the 8–14 μm range, some approximate emissivity values are given in Table 3. The important facts that should be noticed from this table is that most objects have a high emissivity and can be readily detected using an infrared detector. The emissivity values in Table 3 are a collection of values from tables readily available on the Internet (<http://www.thermoworks.com>, 2014). The ones chosen for Table 3 are common materials used in rooms where occupancy sensors might be used. Of particular interest are polished brass and aluminum because these materials are commonly used in rooms for decorative purposes along with similar materials such as polished chrome. Materials like these, with a low emissivity, are great reflectors. When there are good reflectors in a room, this could spell trouble for a PIR sensor. PIR sensors are usually installed into the space in such a manner as to not "see" outside the space. These good reflectors act as mirrors and can cause the sensor to "see" activity outside its intended field of view.

A simple procedure can approximately measure emissivity with an infrared thermal camera. The procedure involves connecting a reference-measuring device such as a thermocouple to the material to be evaluated. The material's temperature is

Table 3 Approximate emissivities of common materials found in rooms in the range of 8–14 μm

Material	Emissivity, %
Natural wood	90–95
Concrete	71–88
Cloth	95
Shrubs	90
Gypsum	80–95
Skin	95–98
Brick	81–94
Glass	92–94
Polished brass	0.5–5
Polished aluminum	1–4

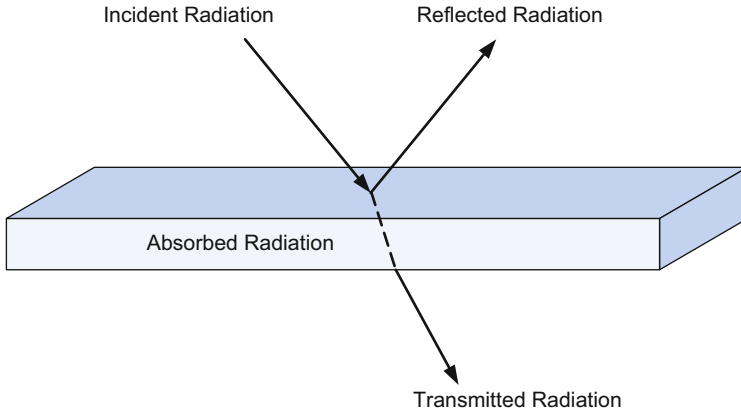


Fig. 3 Behavior of a material with electromagnetic radiation incident upon it

measured with the thermocouple and the emissivity setting on the camera is adjusted until the temperature on the camera matches. The material's emissivity is the setting on the camera.

When radiation is incident onto a material, the radiation is either absorbed by the material, reflected from it, or transmitted through the material (see Fig. 3). The amount of reflected radiation is described by the reflectance ρ . The amount of radiation absorbed is described by the absorptivity α and the amount transmitted through it is described by the transmissivity τ . For a material, these quantities sum to one (scattering is ignored):

$$\alpha + \tau + \rho = 1 \quad (3)$$

Most of the objects in view of the sensor are opaque so τ is approximately zero. So when a sensor views an object, the radiation received by the detector comprises both the reflected radiation and the emitted radiation (recall $\varepsilon = \alpha$). If one assumes that τ is zero, then the reflectance can be easily calculated from Table 3 as $\rho = 1 - \varepsilon$. One interesting case is glass, particularly the common question as to whether an infrared detector can see through glass. The transmissivity of glass in the 8–14 μm range is near zero (Berkley Lab 2014). This means that glass will either absorb (mostly) or reflect the infrared. In the case of a glass mirror, the reflection is off the surface of the mirror not the reflector behind the glass. So, no, an infrared detector cannot see through glass.

Other important materials are those used in the construction of the sensor itself. The first important material is the lens material. This, by necessity, has to have a high transmissivity. The lenses in PIR sensors are made from variants of polyethylene, so they have reasonably high transmittance around 0.6 (<http://www.fresneltech.com>, 2014). The material of the window in the detector itself is also important. Ideally, these windows should be constructed as optical filters that reduce the amount of radiation outside the 8–14 μm range. In practice, they are constructed as long pass

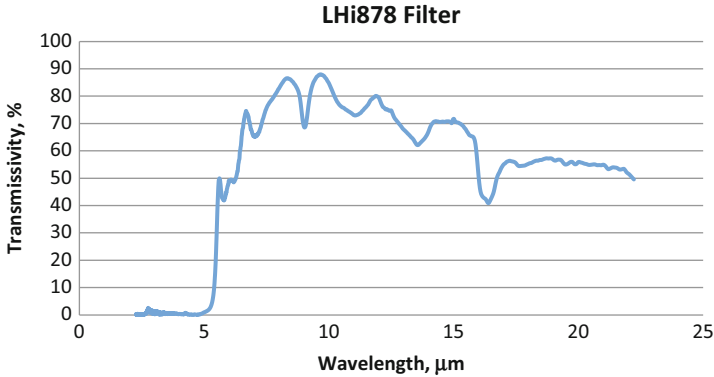


Fig. 4 Transmissivity of the window material in an LHi878 infrared detector

filters with a cut-on wavelength of 8 μm. These windows have an average transmissivity of approximately 0.5 (see Fig. 4).

Lens

As mentioned, the lenses are made of polyethylene, chosen because it is optically transparent to infrared, it is inexpensive, and is easy to mold. Typically, these lenses have an average transmittance a little greater than 60 % which must be taken into account when designing the sensor since it affects the overall system performance. The incident radiation attenuates as it passes through the material and the attenuation is a function of its thickness

$$I(z) = I_0 e^{-\alpha z} \quad (4)$$

where I_0 is the incident radiation, z is the thickness of the material, and α is the absorption coefficient. It is necessary to make the lens as thin as possible to preserve the signal level so as to maximize the detection of the object. This creates the problem that sensors can be very fragile, if not designed properly. If not designed appropriately, it is easy to crush the lens and render the sensor nonfunctional. Manufacturers have figured out how to construct the sensors so that they are vandal proof.

As mentioned earlier, PIR detectors react to changes in temperature. This requirement, that the infrared signal must change, drives design choices in the lens. The infrared energy must be modulated (turned on/off) onto the detector's active surface to maximize detection. This is accomplished by arranging the sensor's lens into an array of lenslets to help modulate the infrared energy reaching the detector surface. Each lenslet "sees" an area out in the field of view of the sensor and in between lenslets there are null spaces. As a person walks across the field of view, the infrared signal turns on (seen by lenslet) and off (in the null). Preferably, the individual lenslet

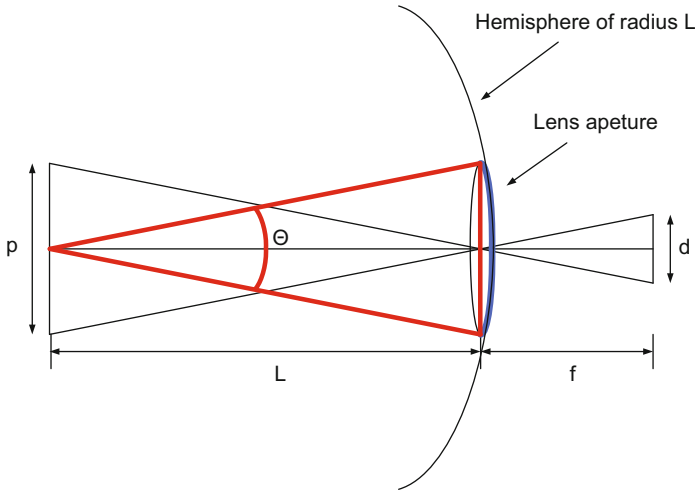


Fig. 5 Diagram showing the angle subtended by the lens aperture

is made as large as practical to capture as much energy as possible. However, there are practical limitations (limits on sensor size) that usually keep the lenslets small and this can significantly impact the sensor performance.

To understand the impact, it is necessary to determine the amount of the radiation captured by the lenslet. Recall that the infrared radiation from an object radiates out through the hemisphere in front of the object. To determine how much is captured by the lens, the portion of the energy radiating into the lens must be calculated. Referring to Fig. 5, the solid angle subtended by the lens in the diagram is approximately

$$\theta = \frac{A_{\text{lens}}}{L^2} \tag{5}$$

where A_{lens} is the area of the lens and L is the distance to the object. The total solid angle subtended by the hemisphere is 2π . This means that the total amount of radiation reaching the detector through the lens will be reduced by the fraction:

$$q = \frac{A_{\text{lens}}}{2\pi L^2} \tag{6}$$

Sensors use different types of lenses depending on the choice of the designer. The most popular lens is the Fresnel lens in a plano-convex configuration (see Fig. 6). The main advantage of this lens is its material thickness. As can be seen in Fig. 6, the thickness of the Fresnel lens is reduced significantly as compared to a standard convex lens (see dashed line in Fig. 6). Recalling that the attenuation (Eq. 4) through the material is a function of its thickness, the Fresnel lens will maximize the signal.

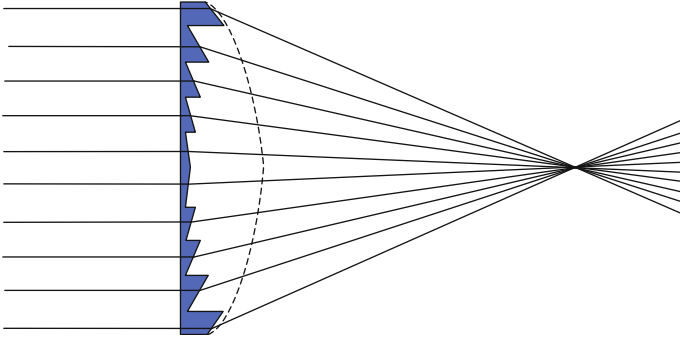


Fig. 6 A cross section of a typical Fresnel lens.

There is always a trade-off in the design between increasing the modulation (more lenses) and increasing the signal (larger lenses). Depending upon the application, the lenses can be designed to provide a large number of detection zones that “see” close up or a small number of lenses that “see” far. Sensors designed for private offices typically have a larger number of lenslets since the dimension of a typical private office is typically less than 300 sq. ft, while a sensor designed for a corridor application will have a few large lenslets but will be to “see” 150 ft. or more.

Another type of lens sometimes used for modulating the infrared is the spot lens. This lens array comprises a collection of convex–convex-type lenses. Usually, these lenses are used when the lenslet can be small and the thickness is also small to avoid excessive attenuation.

Lens arrays for sensors are normally depicted in a couple of different ways. One way of depicting a lens is to show the field of view zone emanating from the sensor. This method is typically used for wall-mounted sensors. An example lens is shown in Fig. 7a in which there are two tiers of lenslets: a top and bottom tier. The top tier has 9 lenslets, and the bottom tier has 7 lenslets. An example of the zones is shown in Fig. 7b, c. Figure 7c shows a side view that depicts the two zone tiers. Figure 7b is a top view that shows the zones from each lenslet.

The other way of depicting the sensors coverage is by showing the projection of the detector elements onto the floor. This is done for sensors that are mounted on the ceiling. An example of this is shown in Fig. 8. This type of drawing is key to understanding how the sensor operates. It shows the sensor’s actual field of view by showing the spots that illuminate the active area of the detector and the null spaces between the detection spots.

The field of view of an individual lenslet is the projection of the detecting elements out into space at a given distance. Referring to Fig. 9, the detector element is on the right side of the diagram. Using geometry, it can be seen that the detection element, having a vertical dimension d , projects out into space and at distance L it

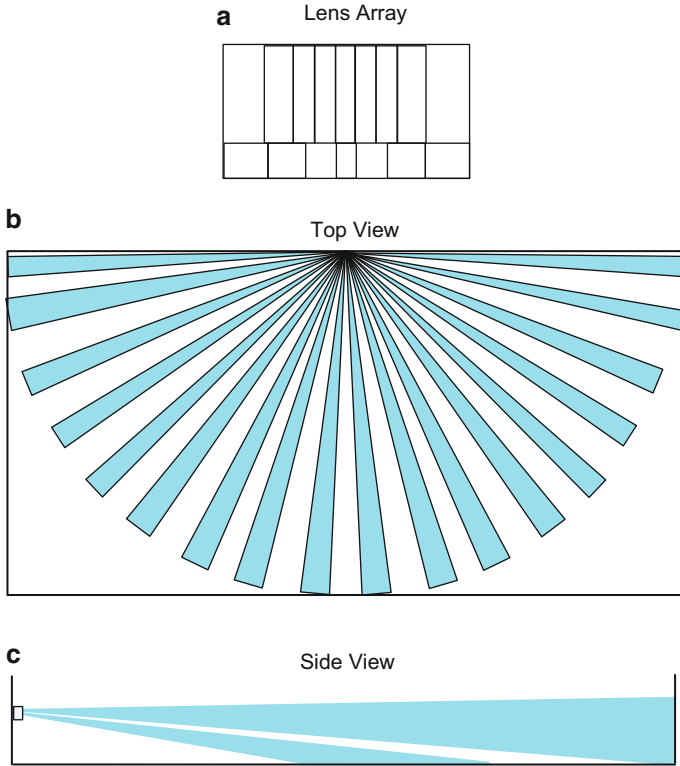


Fig. 7 Diaphragm lens array of a (a). Top view of the detection zones of a wall lens (b). Side view of the detection zones of a wall lens (c)

has a vertical dimension (Everett 1998). Using the definition of the $\tan(\theta)$, the size of the projection p is given by

$$p = \frac{dL}{f} \tag{7}$$

This relationship holds in both the x and y dimensions, so the two-dimensional projection out into space of the detector element at a distance L has an area given by

$$A_p = \frac{A_d L^2}{f^2} \tag{8}$$

where A_d is the area of the detection element and A_p is the area of its projection out in space. Most occupancy detectors have either two or four detection elements. Each lenslet has a projection of these two or four elements out into space. In the following, the signal level will be calculating and knowing the dimension of the projection will be important.

Fig. 8 Pattern of a ceiling sensor projected onto the floor showing the detection zones

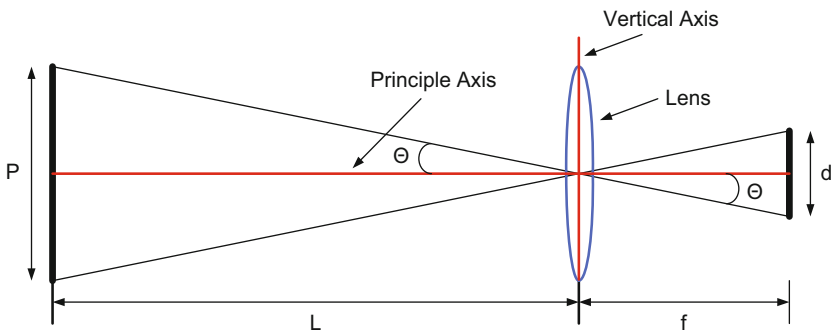
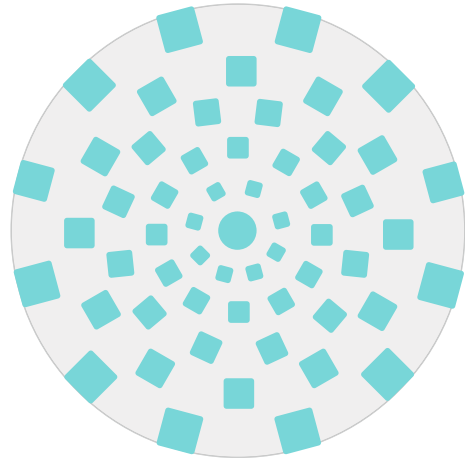


Fig. 9 Diagram of a single lenslet and the projection of the detector footprint out into space

Detecting Infrared Radiation of Objects

As previously discussed, the objects being detected are usually opaque in the sensors optical range. This means that the objects either reflect the incident radiation or absorb it and hence also emit the radiation. There are two terms arriving at the detector: the reflected term, ρ , and the emitted term, ε . Recalling that the transmission, reflection, and absorption all sum to one (Eq. 3) and that the transmission term is zero gives

$$\rho = 1 - \varepsilon \tag{9}$$

Having already discussed the projected footprint of the detection element and the amount captured by the lens, the optical power incident on the detector surface can now be calculated. Assuming the object being detected is smaller than the detector

footprint at distance L , the optical power, W_o , from its emission and reflection incident on the detector is (Everett 1998)

$$W_o = \frac{A_o A_{\text{lens}}}{2\pi L^2} \left[\varepsilon 2\pi h c^2 \int_{8\mu}^{14\mu} \Gamma \frac{1}{\lambda^5 \left(e^{\frac{ch}{\lambda k T_o}} - 1 \right)} d\lambda + (1 - \varepsilon) 2\pi h c^2 \int_{8\mu}^{14\mu} \Gamma \frac{1}{\left(\lambda^5 \left(e^{\frac{ch}{\lambda k T_b}} - 1 \right) \right)} d\lambda \right] \tag{10}$$

where the variables in the equation are:

A_o = the cross-sectional area of the object being detected that is in the field of view of the sensor

A_{lens} = the area of the optical aperture (lenslet)

L = the distance from the PIR detector lens to the object

Γ = transmission constant (function) of the optical path in the sensor (8 – 14 μm)

T_o = temperature of the object

T_b = temperature of the background (ambient)

λ = wavelength of the emitted radiation from the object and background

ε = emissivity of the target

h = Planck's constant

c = speed of light

Note that the first part of the equation has the emissive term emitting at the temperature, T_o , of the object. The second part of the equation is the reflected term that is the reflection of the background radiation at temperature T_b . The transmission function, Γ , is the combined transmission function of both the lens and the detector window. In general, Γ is a function of wavelength ($\Gamma(\lambda)$), but for simplicity, it is being approximated in the above equation as a constant in the range [8 μm , 14 μm] and zero elsewhere.

PIR detectors are made from pyroelectric materials that respond to changes in optical power incident on them. To detect an object, it must either move into or out of the field of view of the detector. If it is assumed that the object moves into the field of view, then the change in optical power is due to the change from the background temperature to the temperature of the object. The equation describing the difference of the optical power, W_b , incident on the PIR detector's surface is

$$\begin{aligned} \Delta W &= W_o - W_b \\ &= \varepsilon \frac{A_o A_{\text{lens}}}{2\pi L^2} \left[2\pi h c^2 \int_{8\mu}^{14\mu} \Gamma \frac{1}{\lambda^5 \left(e^{\frac{ch}{\lambda k T_o}} - 1 \right)} d\lambda - 2\pi h c^2 \int_{8\mu}^{14\mu} \Gamma \frac{1}{\left(\lambda^5 \left(e^{\frac{ch}{\lambda k T_b}} - 1 \right) \right)} d\lambda \right] \end{aligned} \tag{11}$$

The output of the detector is given by

$$V_{\text{det}}(t) = R_v(t) \Delta W$$

where $R_v(t)$ is the voltage responsivity of the detector in volts/watt. These integrals are easy to numerically calculate using a program such as Mathcad[®].

Responsivity

Early occupancy sensors used thermopiles or thermistors, but soon, the preferred choice became pyroelectric sensors. Pyroelectric sensors are made from materials that spontaneously polarize when heated by small amounts of thermal energy. The materials are either crystalline, ceramic, or polymer. Most modern-day occupancy sensors use pyroelectric detectors that are made from ceramic materials, such as PZT, because they are inexpensive to manufacture and durable in the intended applications. When the lattice temperature is increased in a pyroelectric material, the average energy level changes causing a change in the dipole moment in the lattice. This so-called spontaneous polarization gives rise to a change in the charge distribution in the material. The rate of change of the spontaneous polarization with temperature is described by the parameter called the pyroelectric coefficient p . This change in charge distribution causes a surface current on the electrodes on the pyroelectric material:

$$I_p(t) = pA \frac{dT}{dt} \quad (12)$$

where A is the area of the pyroelectric material, T is the temperature, and $\frac{dT}{dt}$ is the rate of change of the temperature. The Laplace transform of $I_p(t)$ is

$$I_p(s) = pAs \quad (13)$$

In a sensor, the pyroelectric element is mounted in such a way as to minimize the thermal conductance of the sensitive material to the case of the detector. The pyroelectric element has several parameters that are useful to describe how it works. A diagram of a pyroelectric detector is shown in Fig. 10. The thermal parameters are its thermal capacity C_T and its thermal conductance G_T with a thermal time constant of $\tau_T = C_T/G_T$. The current responsivity, R_i , of the material is the amount of current generated per watt of incident radiation and is given by

$$R_i = \frac{I_p(s)}{W(s)} = \frac{\epsilon p A}{G_T} \frac{s}{1 + \frac{C_T}{G_T} s} \quad (14)$$

To make the sensor useful, the current needs to be converted to a voltage. The usual means for affecting this conversion is to use the circuit shown in Fig. 10. To calculate the voltage responsivity transfer function, use the relationship $V(s) = I(s)/Y(s)$ where Y is the admittance of the network excited by the current (Whatmore 1986). The admittance is

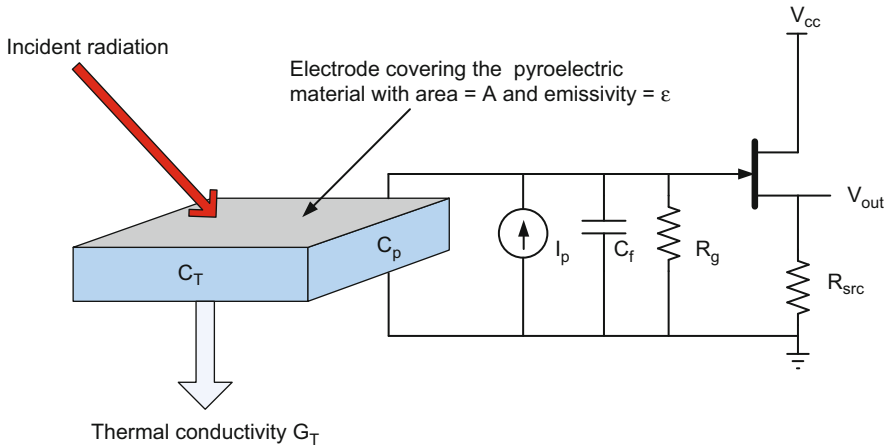


Fig. 10 Diagram of a pyroelectric detector. Conversion to a voltage is accomplished using R_g and the JFET

$$Y(s) = \frac{1}{R_g} + (C_p + C_f)s \tag{15}$$

where C_p is the capacitance of the pyroelectric material, C_f is the gate capacitance of the JFET transistor, and R_g is the detection resistor, connected to the pyroelectric material. Making the voltage conversion gives the voltage responsivity, R_v ,

$$R_v(s) = \frac{I_p(s)}{Y(s)W(s)} = \frac{cpA}{G_T} \frac{s}{1 + \frac{C_T}{G_T}s} \frac{R_g}{1 + R_g(C_p + C_f)s} \tag{16}$$

A typical responsivity is shown in Fig. 11 with the peak value being 72 dB or 4,000 V/W.

Signal Chain

The next part of the sensor system is the amplification stages. Since the typical pyroelectric detector for occupancy sensing has a gain of about 4,000 V/W (72 dB), the detected signals will be on the order of 1 μ V to about 10 mV. To be useful in the detection system, these signals need to be amplified. Furthermore, the band-pass transfer function formed by the pyroelectric detector may need further shaping to fine-tune the system for detection of human motion. A typical first amplification stage is shown in Fig. 12 (Colliard-Piraud 2013). The transfer function of this non-inverting stage is

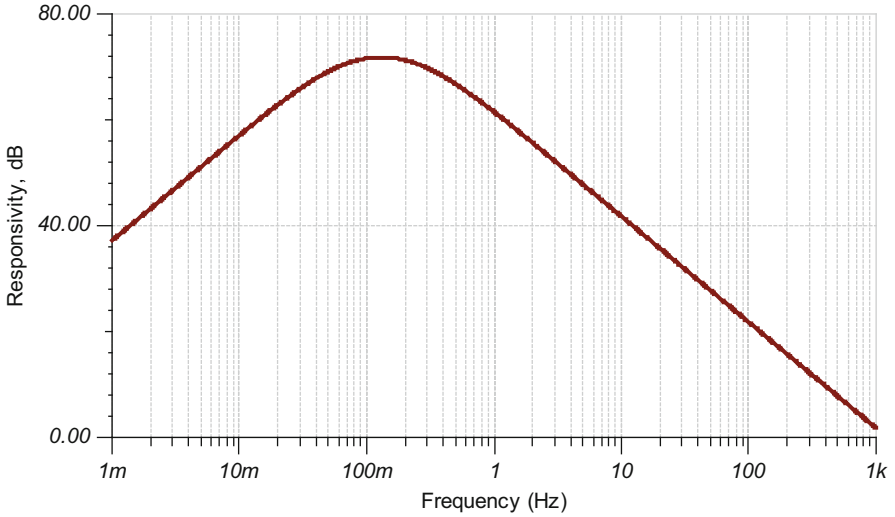


Fig. 11 Voltage responsivity versus frequency for an LHi878 detector from Excelitas

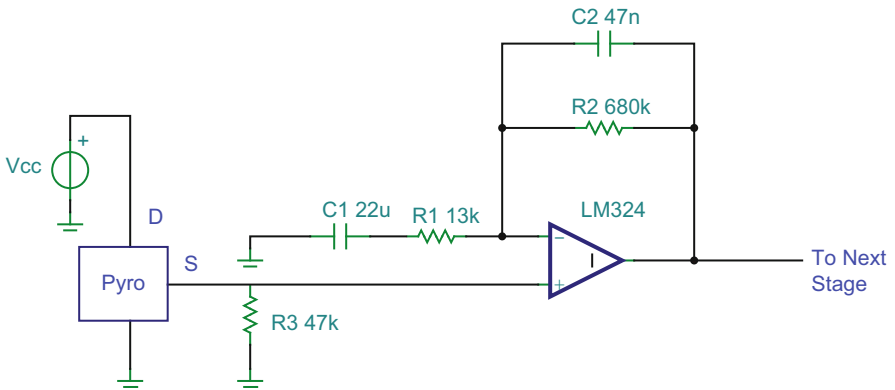


Fig. 12 Circuit diagram of the pyroelectric detector and the first amplification stage

$$H_{A1}(s) = \frac{1 + (C_2R_2 + C_1R_1 + C_1R_2)s + C_2R_2C_1R_1s^2}{1 + (C_2R_2 + C_1R_1)s + C_2R_2C_1R_1s^2} \tag{17}$$

This stage is a band-pass amplifier with a gain of 33.5 dB, a lower frequency cutoff of 0.6 Hz, and a high-frequency cutoff of 5 Hz. The second stage, shown in Fig. 13, is an inverting amplifier with a transfer function of

$$H_{A2}(s) = \frac{-C_3R_4s}{1 + (C_3R_3 + C_4R_4)s + C_4R_4C_3R_3s^2} \tag{18}$$

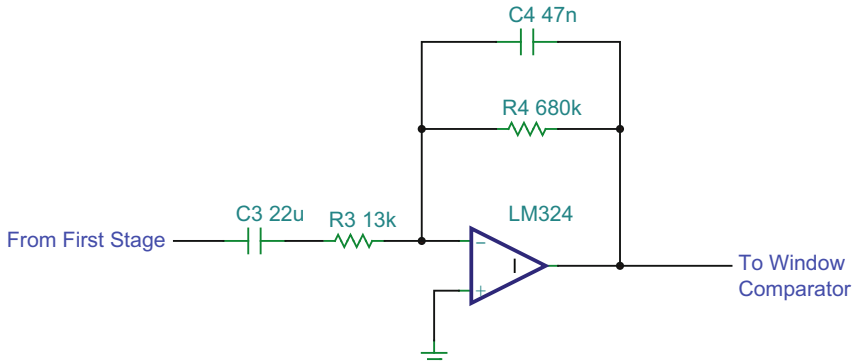


Fig. 13 Circuit diagram of the second amplification stage

It has a similar parametric behavior as the first stage other than the inversion.

Combining the pyroelectric detector’s response (Eq. 17) with the amplifier response (Eqs. 18 and 19) gives the overall transfer function of the sensor system, $H_s(s)$:

$$H_s(s) = R_v(s)H_{A1}(s)H_{A2}(s) \tag{19}$$

This response function is in V/W with a combined maximum of approximately 3,000,000 as shown in Fig. 15. This is reduced from the expected maximum of 9,000,000 (72 dB + 67 dB) because the choice of the center of the band pass for the amplifier does not directly coincide with that of the pyroelectric detector.

Having conditioned the signal, the signal needs to be detected to indicate the presence of an occupant. One of the most prevalent ways to do this is by using a window comparator as shown in Fig. 14. Since the motion signals are in effect AC coupled, the signals into the window comparator are bipolar and need both an upper and lower detection threshold. The window comparator is often implemented by using an analog to digital converter (ADC) in a microcontroller. The signal is offset by half of the ADC full-scale range to accommodate the bipolar nature of the signal and window comparator is then implemented digitally.

Noise

The one aspect of the system that has not been discussed is the noise in the system. The noise in the system is responsible for limiting the sensors ultimate sensitivity. There are various noise sources in the sensor and they need to be examined from a systems viewpoint since there are noise contributions all along the signal chain (Whatmore 1986; Aggarwal et al. 2007). Figure 16 depicts the noise sources. Starting from the left in Fig. 16, the first contributor is the thermal noise, V_{therm} ,

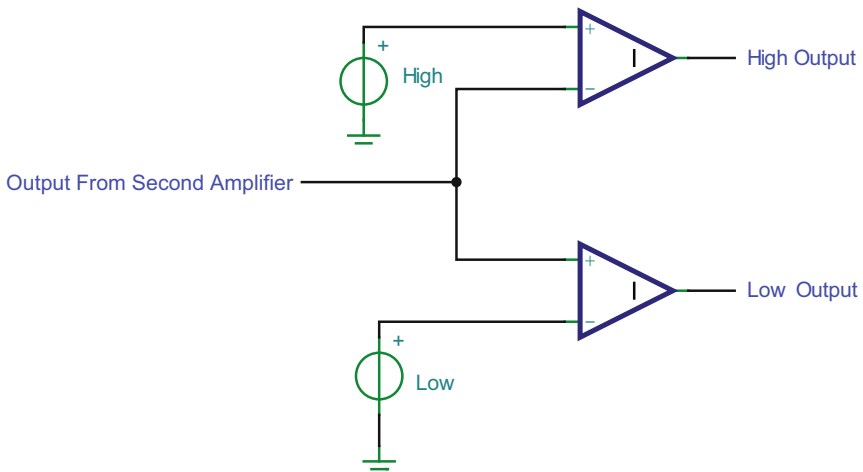


Fig. 14 Circuit diagram of the window comparator

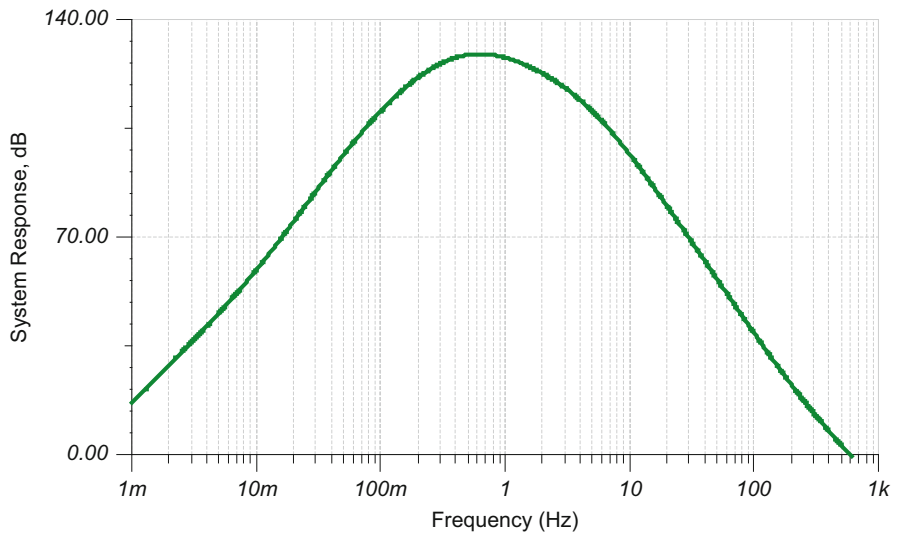


Fig. 15 Combined system response of the pyroelectric detector and the amplifier stages

contribution from the pyroelectric element. This term is due to random fluctuations of heat exchange and photons with the environment. This term is given by

$$V_{\text{thrm}} = \frac{R_v}{\varepsilon} \sqrt{4kGT^2} \tag{20}$$

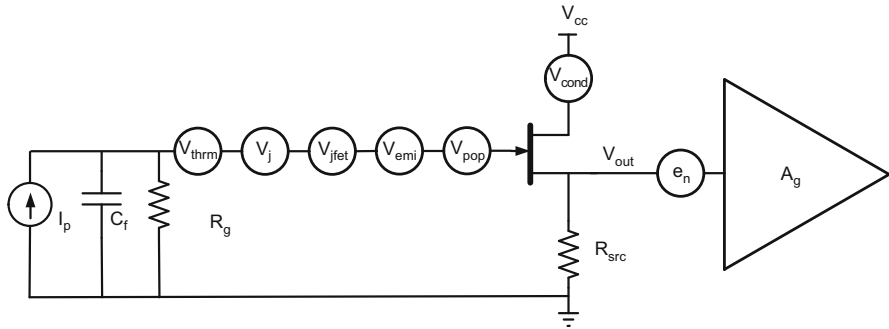


Fig. 16 System noise model

where $G = G_r + G_t$, G_r is the thermal radiation and G_t is the thermal conductance.

The next noise contributor is the Johnson noise, V_j , from the detection resistor on the gate of the JFET, R_g , and the losses in the dielectric of the pyroelectric material. This term is approximately given by (Whatmore 1986; Aggarwal et al. 2007)

$$V_j = \sqrt{\frac{4kTR_g}{1 + \frac{1}{\text{Tan}(\delta)^2}}} \tag{21}$$

The assumptions that are made about V_j are that the frequency is low and that the capacitance of the pyroelectric element is much greater than the JFET gate capacitance. The typical operating frequencies of occupancy sensors are much lower than $[R_g C_d \text{Tan}(\delta)]^{-1}$ which meets the low-frequency assumption. The next contributor is the noise in the JFET, V_{fet} , and an example of the noise in a 2N4118 is shown in Fig. 17. There is both a constant noise floor and a frequency variable term referred to as $1/f$ noise that contributes mostly at lower frequencies.

The next contributor is the noise from radiated electromagnetic interference. This noise can be generated in a variety of ways. Some typical sources are:

- Ignition noise from vehicles
- Motor brush noise
- Microwave ovens
- Cell phones and portable phones
- Wireless routers

This noise is transient in nature and shows up sporadically. Some pyroelectric detectors have an EMI capacitor built in to reduce the effects of this noise. Usually, it is up to the sensor designer to take steps in the circuit design and layout to minimize the effects of EMI. When placing the sensors into an application, it is important to avoid situations where the sensor is too close to these sources; otherwise, false tripping will result.

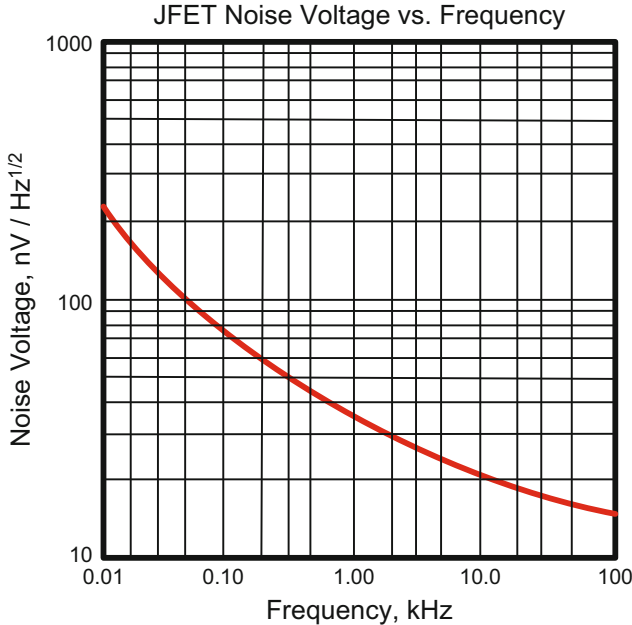


Fig. 17 Voltage noise of a 2N4118 JFET

The next source of noise is from a phenomenon referred to as “popcorn” noise. In early electronics, contaminants in the semiconductor (JFET and operational amplifiers) caused a trapping of electrons at the contamination site. When these electrons released, there is a sudden “pop,” a high-amplitude, narrow pulse. The name came from audio applications. In audio systems, when the electrons released, the speakers made a popping sound like popcorn popping. Modern electronics have become advanced enough that this phenomenon has all but disappeared from electronics. There is another source, the pyroelectric element itself. If there is a sudden change in the mean value of the signal on the pyroelectric detector’s active area, then an impulse occurs that can be several milliseconds in duration. This effect is rare and may only be observed after collecting weeks of data. However, the impulse can be large enough to cause a sensor to detect occupancy. An example of a “popcorn” noise pulse is shown in Fig. 18.

The next contributor is from conducted noise. Many sensors are ultimately powered from the AC mains and as a result are susceptible to conducted EMI. Sensors may also be powered from a 24 V dc bus, and in this case, radiated noise may couple to the dc bus and be conducted into the sensor circuit. Some typical sources are:

- Transients due to lightning strikes in the power system
- Motor switching
- Electrostatic discharge

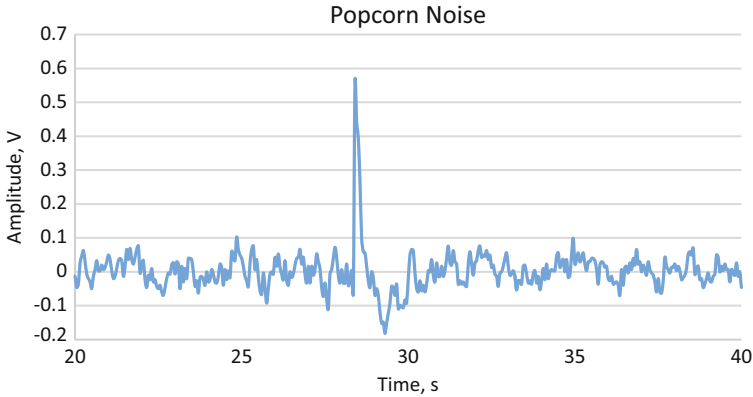


Fig. 18 Example of popcorn noise from a pyroelectric detector

- Relay switching
- Electrically fast transients

These sources of conducted EMI are generally dealt with by proper power supply design. Unfortunately, if the noise conducts through the power supply, then the noise will be passed from the drain of the JFET to the rest of the system with little attenuation. This can be a major source of false tripping if not properly dealt with.

The next contributor is the noise from the amplification stages using the operational amplifiers. It has several noise sources (see Fig. 19). Each of the resistors has a noise component but contributes very little (Carter 2008). The input current noise is also a minor contributor, so the equivalent noise model is given in Fig. 20 where the operational amplifier’s noise voltage is the main contributor so the total output noise, V_n , is given by

$$V_n = A_g \sqrt{(E_n)^2 + (e_n)^2} \tag{22}$$

The term e_n is the intrinsic noise of the operational amplifier and A_g is the gain of the amplifier stages. An example of the intrinsic noise of an operational amplifier is shown in Fig. 21. The term E_n is the noise from the rest of the system comprising V_{therm} , V_J and V_{fet} and is given by

$$E_n = \sqrt{V_{\text{therm}}^2 + V_J^2 + V_{\text{fet}}^2} \tag{23}$$

In practice, if a low-noise operational amplifier is used, then the e_n is only a minor contributor. It should be noted that the only the noise terms that contribute to a continuous noise floor are included. The terms like the EMI and popcorn noise are typically only transient in nature. An example of the power spectral density of the noise from an entire sensor system is shown in Fig. 22.

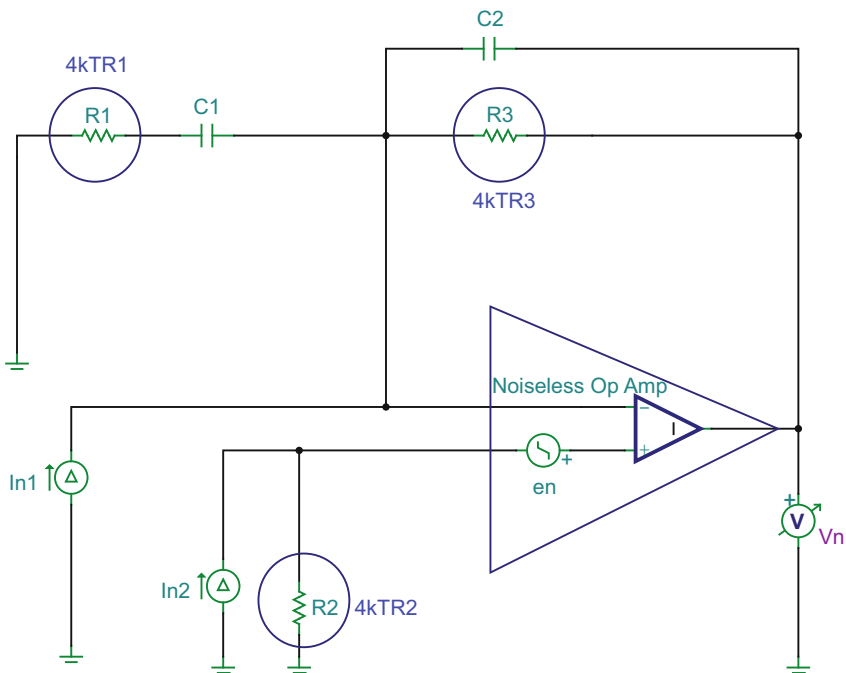


Fig. 19 Equivalent noise model of an operational amplifier showing all sources of noise

Since the detector is the major noise contributor, one has to be concerned with the noise characteristics of the detector. The detectors are often characterized by what is termed as the noise equivalent power (NEP). The NEP is the optical power incident on the active area of the detector necessary to give a signal to noise ratio of one in a signal bandwidth of one Hz. This is given by

$$NEP = \frac{E_n}{R_v} \tag{24}$$

Another term used is the detectivity which is $D = 1/NEP$, but it is difficult to compare detectors using these quantities because the responsivity is a function of the detector area, A . Another term was created that normalizes the detectivity to a unit area so detectors with different areas can be compared. This term is named D^* and is given by

$$D^* = \frac{\sqrt{A}}{NEP} \tag{25}$$

and should be used when comparing detectors.

Other noise sources that should be considered are environmental. All pyroelectric materials are also piezoelectric with the consequence that these detectors are

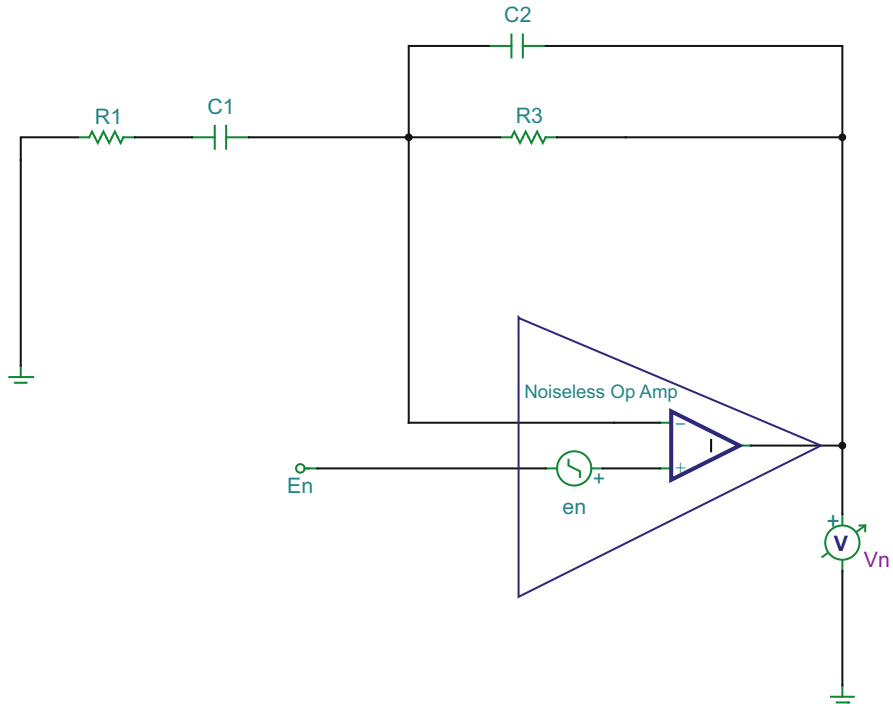


Fig. 20 Equivalent noise model of an operational amplifier showing the main noise contributors

susceptible to shock and vibration. Noise from this type of source is called microphonic noise. In the beginning of this chapter, it was mentioned that placing two pyroelectric elements in series opposition had the effect of canceling interference. A diagram of this arrangement is shown in Fig. 23. Manufacturers of detectors are careful to make parts that closely match the response of the two elements, typically to better than 10 %. The effect is a cancelation of common mode infrared signals and those caused by events inducing a piezoelectric response. These detectors are referred to as dual-element detectors and are widely used in wall mount sensors. Another type of series opposition sensor that is widely used is the quad-element sensor. The quad-element detectors have the same canceling effect as the dual-element detectors and are used in ceiling sensor applications (more on this later).

Temperature Effects

Some applications of sensors require that they operate in conditions other than room temperature. These applications include outdoor climates where the sensor can experience temperature extremes that range from $-25\text{ }^{\circ}\text{C}$ to $65\text{ }^{\circ}\text{C}$. Other applications combine detectors with other components that can reach high temperatures and

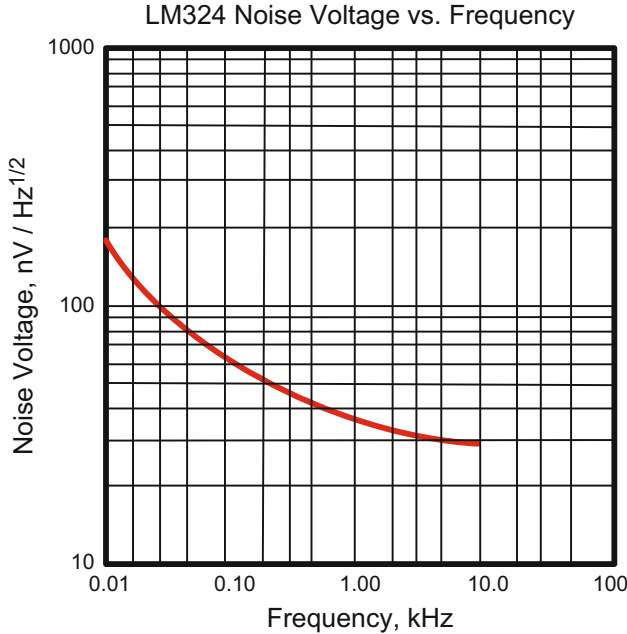


Fig. 21 Voltage noise of an LM324 operational amplifier

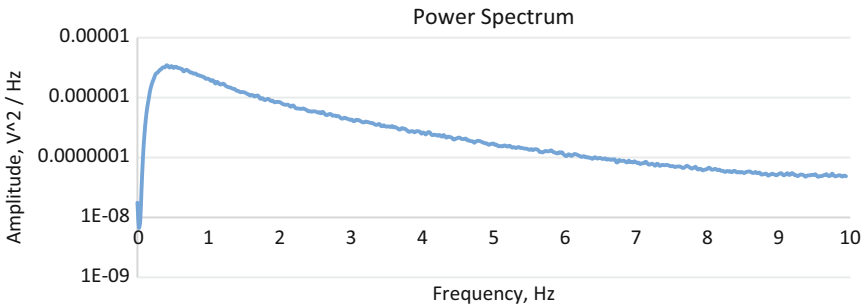


Fig. 22 An example of the power spectral density of the noise in a Lutron sensor

cause heating of the detector. A typical application of this type is a sensor dimmer switch where a detector is incorporated into the product along with power electronics, such as a TRIAC, that dissipates a significant amount of heat. For these reasons, it is important to understand the performance of sensor under varying temperature conditions.

For ceramic detectors, the responsivity has a negative temperature coefficient of approximately 0.1 %/C°, so changing the temperature from 25 °C to 65 °C only causes a slight degradation of 4 %. The largest impact is from the change in noise in

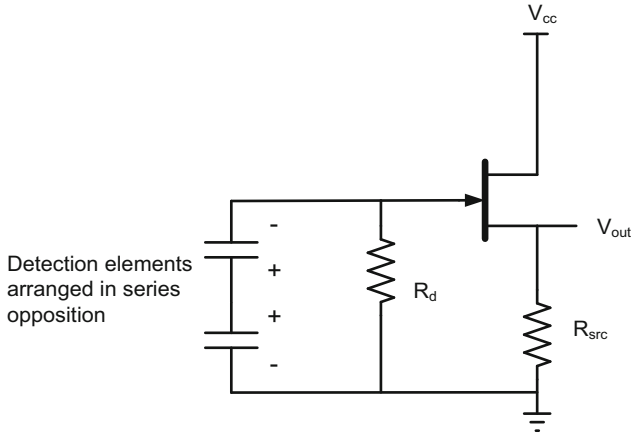


Fig. 23 Circuit diagram showing the pyroelectric elements arranged in series opposition

the system. As mentioned, there are many contributors to the noise in the detector and each of these aforementioned noise sources is a function of temperature. As temperature increases, the noise in the JFET increases as shown in the reference (Heinze, et al. 2004). Noise in the JFET only increases slightly from room temperature to 50 °C, but at 85 °C, the noise increases by almost a factor of ten.

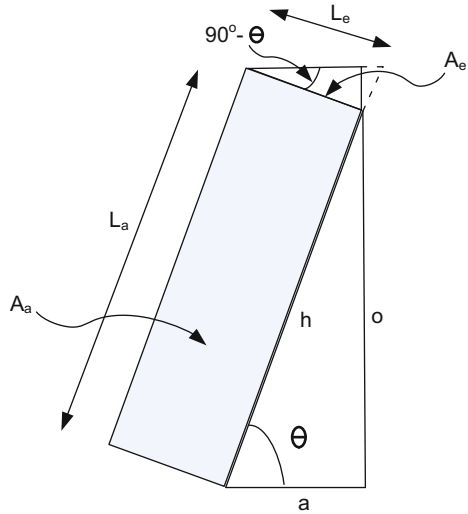
System Response

So far, the sensor system has been described piece by piece giving the characteristic of each section. The only part missing is the description of the object in motion. In a later section, test procedures will be described that standardize the way sensor coverage is reported. A robotic arm will be used in the test section. In Fig. 24, simple diagram of the arm is shown. The following parameters described it:

- L_a is the length of the arm which could also be the length of a hand.
- L_e is the length of the end of the arm which is square in dimension.
- A_a is the area of the arm which is $L_e * L_a$.
- A_e is the area of the end of the arm which is given by $L_e * L_e$.
- Θ is the angle through which the arm moves.
- Θ_0 rate at which the angle changes.

Assume that the arm moves parallel to the principle axis of the lens and that the arm is completely encompassed within the footprint of the detector projection. Using the definition of $\text{Sin}(\theta)$, the area of the exposed surface of the object being detected, $A_o(t)$, as viewed by the sensor is given by

Fig. 24 Diagram of an arm used to test sensor response



$$A_o(t) = A_a \sin(\theta_0 t) - A_e \sin\left(\frac{\pi}{2} - \theta_0 t\right) \tag{26}$$

This is sinusoidal excitation and is simply modeled using system theory. In practice, if a human is taking a drink from a cup, they start with their arm at rest, raise it to their mouth, drink, and then replace the cup on the surface from which it was taken. The act of raising and lowering the arm is then modeled by

$$A_o(t) = \text{Rect}\left(\frac{t}{2\tau_0}\right) \left[A_a \sin(\omega_0 t) - A_e \sin\left(\frac{\pi}{2} - \omega_0 t\right) \right] \tag{27}$$

where Rect() is the rectangle function of width $2\tau_0$ and ω_0 is the frequency at which the arm travels and is equal to one quarter of θ_0 . The parameter τ_0 is the reciprocal of ω_0 . If the rate of change of the angle is 90° per second, then the frequency is 0.25 Hz and the width of the rectangle function is 0.5 s. This amounts to a half sine pulse as shown in Fig. 25.

Using the previous definition of the change in power incident on the detector surface, ΔW , it is now possible to determine the theoretical response of the system to an arm moving. Assuming that the arm is completely within the footprint of the detector, the half sine pulse excitation can be applied to the system. Calculating the peak change in the optical power incident onto the detector at a distance of three meters gives a power of $0.1 \mu\text{W}$ (Eq. 11). The temporal response of the system using a LHi878 detector is shown in Fig. 26. After the detector pulse passes through the circuitry previously described, it is modified as shown in Fig. 27.

The Fourier transform of the half sine pulse is a Sinc() function centered at the frequency of excitation, ω_0 . Referring back to Fig. 15, it can be seen that the system

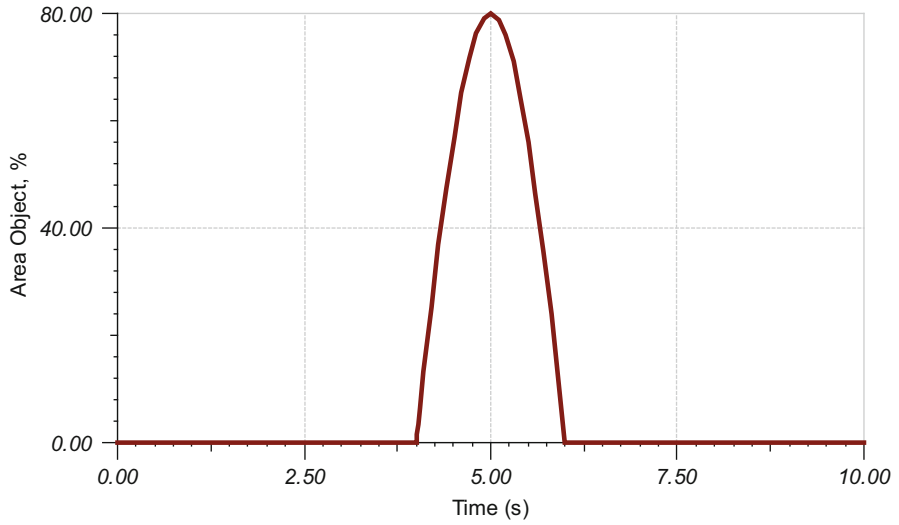


Fig. 25 Half sine pulse model of a human arm motion

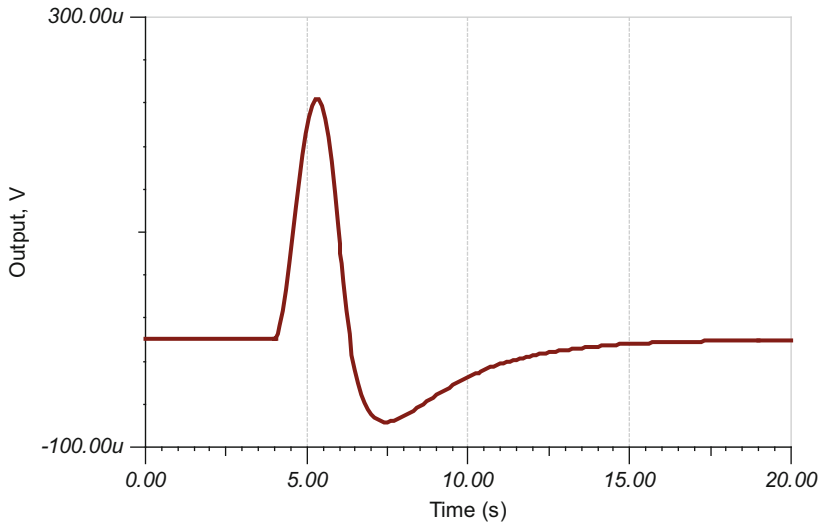


Fig. 26 Output of the LHi878 detector for a 100 nW half sine pulse

response achieves a peak near 0.75 Hz. The frequency of excitation for the example was at 0.25 Hz. As the speed of movement increases (or decreases), the spectral peak of the excitation will move. This change in the position of the spectral peak results in

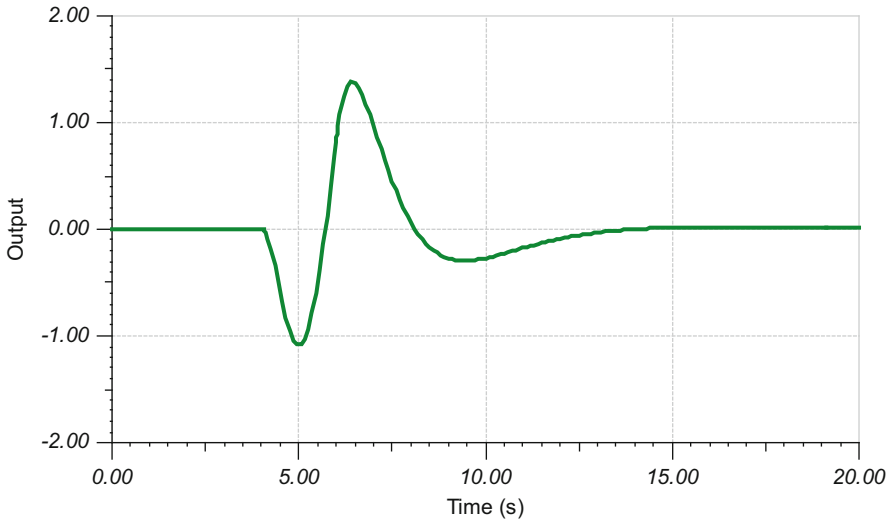


Fig. 27 Output of a Lutron sensor for a 100 nW half sine pulse

a reduced response, so moving very slowly or very quick will reduce the sensitivity of the sensor.

The previous model can be used to model a hand, an arm, or even a head movement. To model a body, however, a slightly different approach must be taken. If it is assumed that the detector footprint is rectangular in shape and that a body passes in front of the footprint perpendicular to the principle axis of the lens, this can be modeled as the convolution of two rectangle functions. This convolution results in a triangle function if the body has the same width as the footprint. While this is possible, it is not likely. Either the footprint is smaller or possibly larger. The result is a triangle with a flat top as shown in Fig. 28. The width of the optical power waveform will depend on the distance of the object to the detector and the speed at which the object passes across the footprint. At a fixed distance, the optical pulse will become narrower as the velocity of the object increases. The Fourier transform of a triangle (not flat topped) is a $\text{Sinc}^2()$ and as the pulse becomes narrower, the energy in the pulse spreads to higher frequencies, thus again reducing the amplitude of the response. This means that it will be more difficult to detect a person that is running than if they were walking. A good rule of thumb (Schmidt 2002) at which frequency of the energy is concentrated is given by

$$f = \frac{vf}{2\pi sL} \quad (28)$$

where f is the focal length (mm), v is the velocity of the object being detected (m/s), f is the frequency where the energy will be concentrated, s is the size of the detector, and L is the distance to the object being detected.

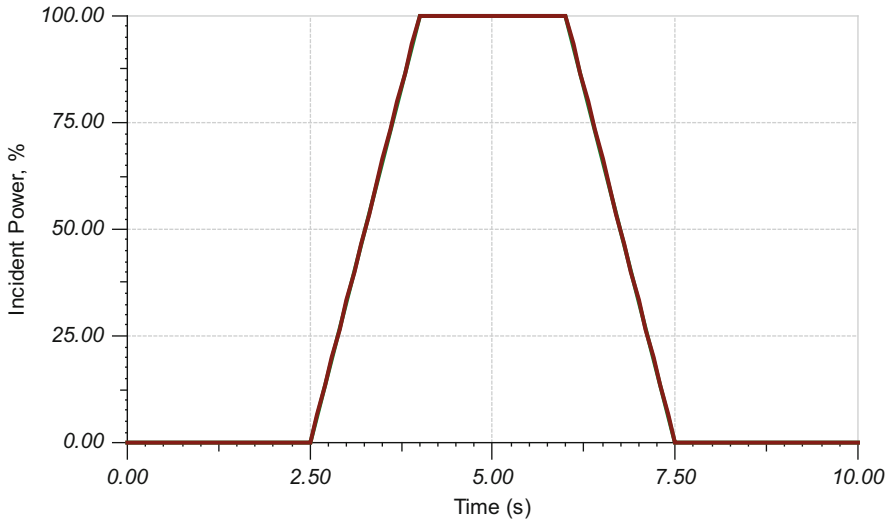


Fig. 28 Percent incident power of a large object passing through a projected footprint

Testing

Early on in the development of sensors, there was no methodology to compare sensors and this led to misunderstanding of sensor coverage claims. Eventually, a NEMA guide was developed that provided testing procedures, standardizing methods to measure and report sensor coverage. In 2011, this guide was reissued as the standard NEMA WD7-2011 (NEMA 2011). The standard contains two tests: major motion and minor motion. The major motion test simulates the large motions of a person walking in a room. The minor motion test simulates motions of a person sitting at a desk and taking a drink of a beverage from a cup or answering the phone.

These tests are performed on a grid laid out with 3 ft. by 3 ft. cells in a large room capable of testing the entire range of the product. One of the important aspects of the test grid is the floor covering. If the test area has a shiny floor surface, then the test results can be skewed since the floor will act like a mirror and potentially falsely improve the test results. The test space importantly needs to be larger than the coverage area of the sensor or reflections from the walls in the space could impact the measurements. The space needs to be environmentally controlled to 70°F and 30–70 % humidity.

The major motion test uses a person of average height and weight (5'7" and 170 lbs.). The person starts in the center of a cell and walks to the adjacent cell at a 4 ft./s velocity. The sensor is usually set to test mode that has a few second time-out and a means to indicate it detected an occupancy event. If an occupancy event is detected, then a positive detection is recorded for that cell.

Fig. 29 Diagram showing the motions that the heated robotic arm moves through in accordance with NEMA WD7-2011

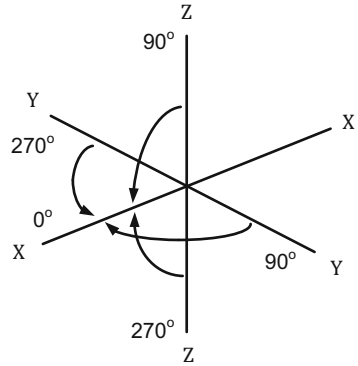
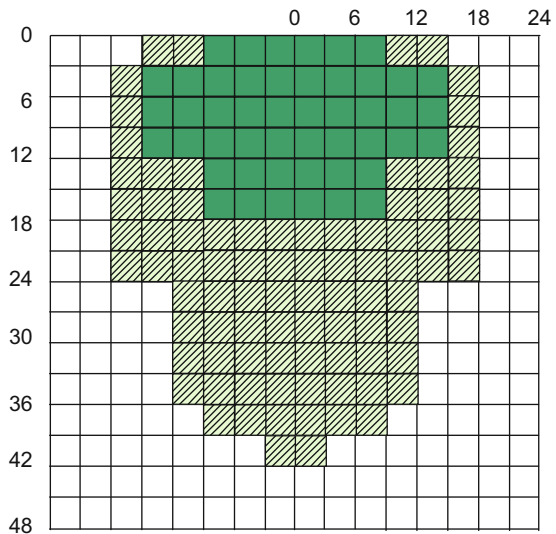


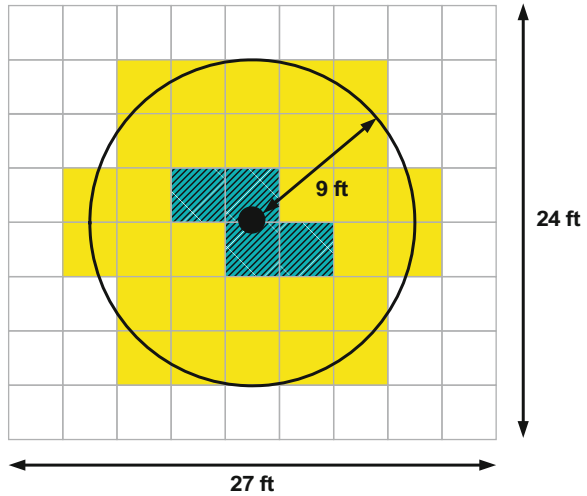
Fig. 30 Example of a coverage diagram used to show the coverage of a wall box sensor in accordance with WD7-2011



The minor motion test uses a robotic arm having a length of 15" and a cross-section of 3" x 3". The arm is heated to the same temperature as the surface temperature of human skin (95 °F). The arm is remote controlled to swing through a 90° motion at a velocity of 90°/s. This is similar to the arm previously described as a model for a human arm. The arm is placed in the center of the cell and operated through four separate motions. Fig. 29 shows the four motions that are similar to throwing a ball at the sensor or throwing a Frisbee to the sensor. The motion is always parallel to the principle axis of the sensor. If any one of the four motions causes an occupancy event to be detected, then a positive detection for that cell is recorded.

The advantage of using NEMA WD 7 -2011 is that coverage results provide customers with a standardized way in which to compare sensor coverage. A typical example is shown in Fig. 30 where the coverage pattern of a PIR wall box sensor is illustrated.

Fig. 31 Fine motion coverage for a Lutron sensor (yellow) and for a non-Lutron sensor (blue/crosshatched)



The National Lighting Product Information Program, NLPIP, used a robotic arm to test sensors in 1997 (D. Maniccia et al. 1997). This robot was similar to the one specified by NEMA, but had an extra dimension. The robot was constructed to also perform hand motions. Lutron adapted this test to simulate the fine motion of a hand moving. This motion simulates desk activities, such as turning the pages of a book. A robot was constructed that has a heated block the size of a typical hand (4" × 6" × 1"). The hand rotates about the wrist with a 90° motion at an angular velocity of 90°/s. The hand is heated to 90 °F which is typical for a human extremity. A typical fine motion result is shown in Fig. 31 for a PIR ceiling sensor.

When PIR detector manufacturers characterize their detectors, one of the parameters that they are interested in testing is the responsivity. This parameter is measured using a black body radiator in conjunction with an optical chopper. The black body radiator is set to a known temperature to produce a known optical power and the chopper essentially turns it on and off typically at a 1 Hz frequency. The optical radiation is viewed through a slit. Given the dimensions of the setup and the signal response of the device under test, the responsivity can be determined. An example of the response of a detector to an optical chopper is shown in Fig. 32.

Applications

Sensors are used in a number of ways within buildings. They can be installed in a wall box opening, and in that case, they usually have a combination of a sensor along with the power electronics to control the load. These sensors are powered from the ac mains. In North America, the mounting height of an electrical switch is not specified in the electrical codes; however, they are usually installed at a height of 48" from the floor.

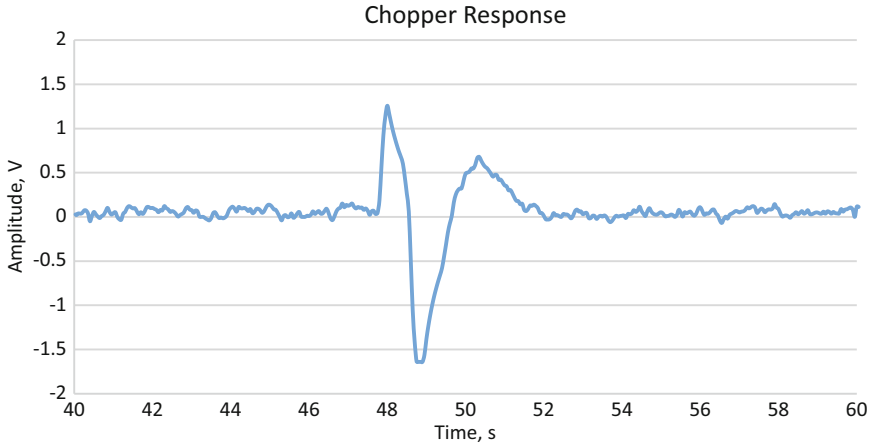


Fig. 32 Response of a Lutron sensor to an excitation by an optically chopper heat source

The pyroelectric detectors used in this application usually have dual elements and are most sensitive when the person walks perpendicular to the principle axis of the sensor. Recall that pyroelectric detectors only respond to changes in optical power. When objects pass perpendicular to the principle axis, the infrared energy is modulated because the objects pass from the field of view of one lenslet, into a null zone, and then into the field of view of the next lenslet. Furthermore, the projection of the detector's active area for each lenslet contains the two elements with a null space between them. This provides additional modulation of the optical energy and is illustrated in Fig. 33.

Since the detectors are sensing optical energy, they must have a line of sight with the objects being detected. Any obstruction between the sensor and the object that must be detected will render the sensor nonfunctional. The lens array for the sensor is constructed in a manner so that its field of view covers the entire room in which it is installed. For example, if the room is square and the sensor is mounted in the center of one of the walls, then it is desirable to have a 180° field of view. Designers have taken different approaches to achieve a 180° field of view. Some designers have used two detectors arranged at an angle to one another so that one detector "looks" left and the other "looks" right. The other approach is to properly design the lens and detection electronics so they are sensitive enough to "see" at angles of $\pm 90^\circ$.

The other aspects of the lens design that are important are to cover the vertical plane and to have enough range for the coverage. To cover the vertical plane, the lens is designed with two or more tiers of lens arrays. The top tier usually "looks" straight out. The lower tiers look down at the floor on an angle. This was illustrated previously in Fig. 7 that shows the coverage pattern of a wall box sensor. The other aspect of the design is to ensure enough range is possible. As objects are further out in the space, the amount of infrared energy captured by the sensor is reduced. A typical wall box sensor has a coverage pattern of approximately 1000 ft^2 , and assuming that pattern is square, this means that the range is a little more than 30 ft.

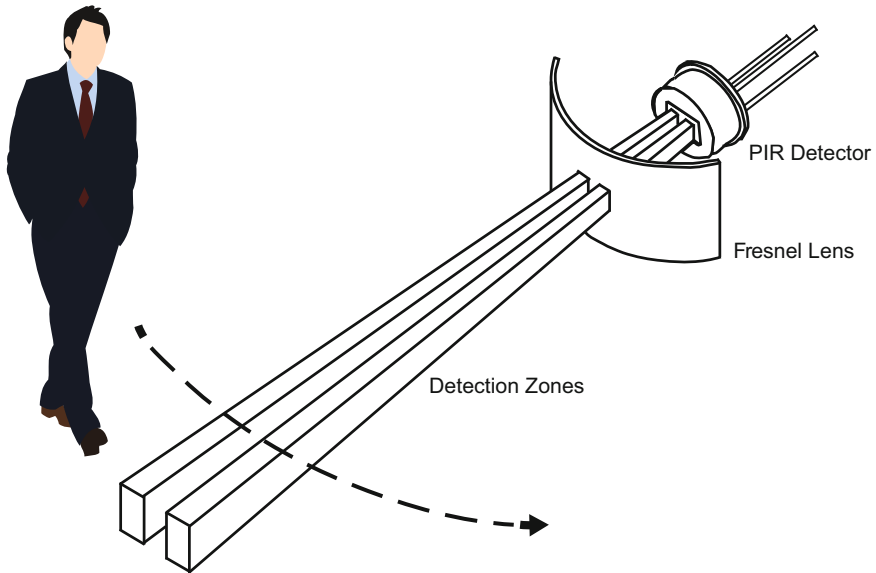


Fig. 33 Illustration of the modulation of the optical energy by a human walking past the dual-element projection from a single lenslet

straight ahead. It should be noted that typical patterns are not quite square but are somewhat elongated straight out in front of the sensor, but assuming it is square is a simple rule of thumb.

When placing a sensor in a space, it is important to know where an occupant is located. For instance, if the sensor is being used in a private office, it is best that the sensor has the best possible view of the occupant for optimal performance. Quite often, the doorway will be close to a corner and the electrical switch is by the doorway. Using a wall box sensor in this instance essentially wastes half of the sensor's usable coverage. If there is a small hallway into the office with the switch in the hallway, then the sensor will not have line of sight coverage of the office and will not work at all.

The office application for a sensor is the most demanding because the occupant will become annoyed if the sensor keeps turning the lights off while they are in the office. To reduce the probability that the light will falsely turn off, a few of things can be done. First, the gain of the sensor system should be set as high as possible to capture fine motion. However, setting the gain too high can cause a false turn on or never allow detection of the vacant state. Second, the occupant should ideally face the sensor since having one's back to the sensor obstructs its view of human fine motion. Third, the occupant should be in the coverage pattern associated with fine motion or minor motion. Being too far from the sensor reduces its effectiveness; closer is better. Fourth, the time-out of the sensor should be set as long as practical. It is very easy for a human to sit very still for one minute, so having a one minute time-out will cause the lights to regularly turn off and will be a great annoyance.

The probability that a person can sit very still for 15 min is very low and they will undoubtedly move enough to be detected by the sensor if its time-out is set to 15 min or longer.

If the electrical switch in the room is not located in a good location for the sensor to see the room, the alternatives are to use either a ceiling sensor or a wall mount sensor. A wall mount sensor is similar to a wall box sensor with a few exceptions. First, the wall mount sensors can be located on any wall at any location. They are usually mounted about 7' from the floor. These detectors also have dual-element detectors. Second, the wall mount sensors typically do not have power electronics and operate off a low-voltage 24 V bus supply. The power to the load is controlled by a device often referred to as a power pack. The power pack connects to the AC mains and generates the 24 V bus supply. In addition, the power pack controls the lighting loads attached to it. The advantages to using a low-voltage sensor are that it is easy to find an optimal place for the sensor, so it has a good view of the occupant, and it is simple to connect multiple sensors together to improve coverage.

Ceiling mount sensors are low-voltage sensors used even more commonly than wall mounted. These sensors can be very effective because they can be mounted close to the occupant, typically right over his or her head. They have the same characteristics as the wall mount sensor except they often use a quad-element detector. The elements are configured in a square pattern and connected in series opposition like the dual elements. Quad-element detectors have two primary advantages. First, there are four elements in the pattern, so the projection has these four elements, providing more modulation than the dual element. Second, because they are arranged in a square, moving in either the x or y direction causes modulation and as a consequence it can "see" both directions equally well.

It should be noted that there are wall mount and ceiling sensors that are line powered and have the power electronics included within them. Typical wall mount sensors of this type are used in outdoor applications. Having line power at the sensor makes it more expensive to relocate, if there is an installation problem.

Locating sensors can be problematic. If one is installing a wall box sensor, one is usually replacing an electric switch or putting it where an electrical switch would be placed. This may not be the optimal position for detecting the occupant. Placing a sensor near a source of EMI is also undesirable. For instance, placing a sensor right next to a wireless router will probably result in false tripping of the sensor. Sensors should not be placed near HVAC diffusers or returns because the sensing electronics and detector are quite sensitive to temperature differentials caused by moving cold or hot air across them, resulting in false tripping.

A sensor's field of view should not be outside the desired coverage area, which can happen in a couple of ways. First, the sensor actually has its field of view extending beyond the desired coverage area because the desired area is small or the sensor can "see" out a door. This can be remedied with a lens mask that blocks the infrared radiation from entering a lenslet. A second way that sensors see outside their intended coverage area is by reflection. In the section on materials, absorption, emission, transmission, and reflection were discussed. A table of emissivities was given in which the values for the emissivity of polished metals were presented.

The low emissivity of these polished metals means they are highly reflective of infrared and act like mirrors. If a highly polished material is in a room and is in the sensors field of view, it could be problematic. For example, consider an office with a wall box sensor that can only “see” the inside of the office. In this office, on a wall facing the sensor, hangs a picture in a polished brass frame, which can be seen from hall looking into the office. As people walk past the office, the infrared radiation from their bodies will be reflected by both the brass frame and the glass inside the frame. It is possible that this infrared signal could cause the sensor to trip inadvertently.

Another important aspect of the sensor performance relates to the optical power differential that is detected. It was previously discussed that difference in optical power on the detector surface determines the magnitude of the signal in the electronics. Refer back to the integral of Planck’s equation (Eq. 11) where the difference in optical power between the background and the object being detected is discussed. Consider the example of a person walking through the detection zones with the following parameters:

- $L = 9 \text{ m}$
- $\epsilon = 0.9$
- $T_o = 90 \text{ F}$
- $A_o = 0.2 \text{ m}^2$
- $A_{\text{lens}} = 0.5 \text{ cm}^2$

As shown in the graph of Fig. 34, as the temperature of the background approaches the temperature of the object, the difference in optical power (ΔW) approaches zero. Hence, the detector has zero output voltage and nothing is detected. This means that as the temperature of the background (ambient) approaches that of a person, then the person will become invisible to the detector. Fortunately, the temperature of human is rarely uniform and neither is the background temperature. The result is that the signals are reduced but maybe not enough to render the sensor totally inoperable. The ability to sense minor motion will be impaired but major motion could still be detectable.

Ultrasonic Sensing

The next technology discussion is on the subject of ultrasound detection of motion. Once again, this particular technology requires the object be in motion. As mentioned previously, this technology has its roots in Doppler SONAR and the discussion will largely be based on this theory. The basic physics behind motion detection is to transmit a continuous wave ultrasonic signal and detect the Doppler shift caused by human motion. The transmitted ultrasound is typically either 32 kHz or 40 kHz since these transducers are widely available and inexpensive. An important aspect of these frequencies is that they are far enough above the human ability to hear them that they will not annoy people or cause any health concerns. Codes require that the

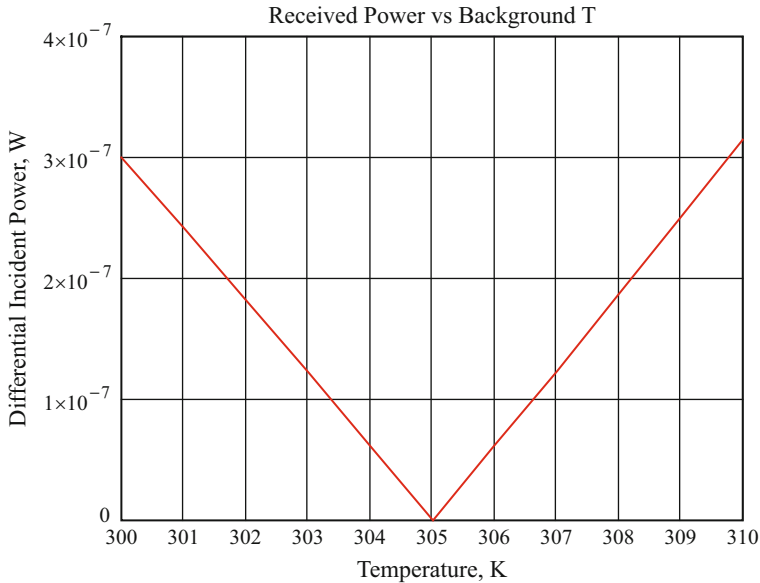


Fig. 34 Illustration of the differential optical power incident on the detector's optical surface when the background temperature is varied

sound pressure level (SPL) not exceed 115 dB SPL as measured 5 ft. from the transmitter along its principle axis for frequencies above 31.5 kHz.

Just as is the case with PIR sensors, the ultrasonic sensors can be mounted in a wall box sensor, a wall mount sensor, or a ceiling sensor. Quite often, these sensors contain both PIR and ultrasonic sensing technology and are called dual-tech sensors. Dual-tech sensors can offer better performance by combining the two technology and taking advantage of the strengths that each offer.

Ultrasonic Field Patterns

When a transducer is excited by a voltage of the proper frequency, it oscillates at that frequency. For example, to produce a good ultrasonic field from a 40 kHz transmitter, a 40 kHz sine wave could be applied to the transmitter's terminals. However, a square wave is usually applied since it gives the same results and is far easier to generate. These transducers are very narrow band and will attenuate the higher-frequency harmonics of the square wave. The transmitters produce a spherical wave front that propagates as a longitudinal wave out into the space in front of it. A graph showing the normalized measured field pattern for a 40 kHz transducer is shown in Fig. 35. The sound pressure level (SPL) produced by the transmitter is measured using a calibrated microphone and is calculated using the formula:

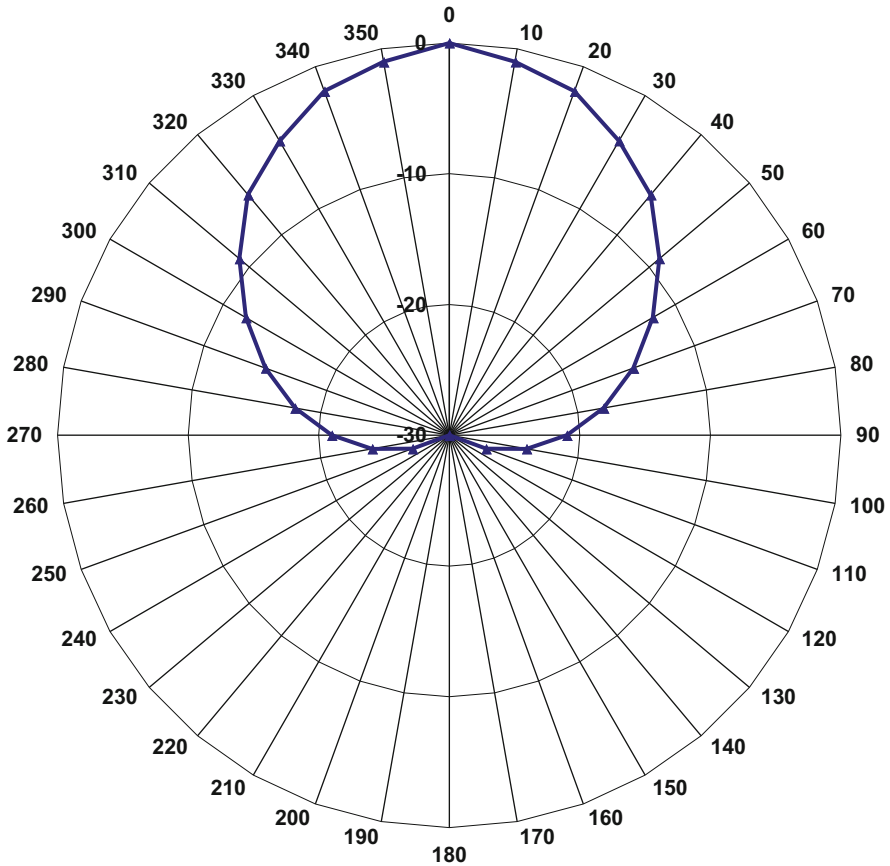


Fig. 35 Radiation pattern of a 40 kHz ultrasonic transducer

$$SPL = 20\text{Log}_{10} \left[\frac{p}{p_0} \right] \text{dB} \tag{29}$$

where p is the measured sound pressure and p_0 is the reference pressure which is $20 \mu\text{Pa}$. The manufacturers of transducers recommend that the sound pressure be measured 30 cm from the front of the transmitter. The intensity level (IL) at standard air pressure is given by (Blackstock 2000)

$$IL = SPL - 0.16 \text{ dB} \tag{30}$$

This is essentially the same as the SPL since 0.16 dB is inconsequential. When sound level SPL_{r_1} is measured at a distance r_1 , the sound level SPL_{r_2} at the distance r_2 is

$$SPL_{r_2} = SPL_{r_1} + 20\text{Log}_{10} \left(\frac{r_1}{r_2} \right) \tag{31}$$

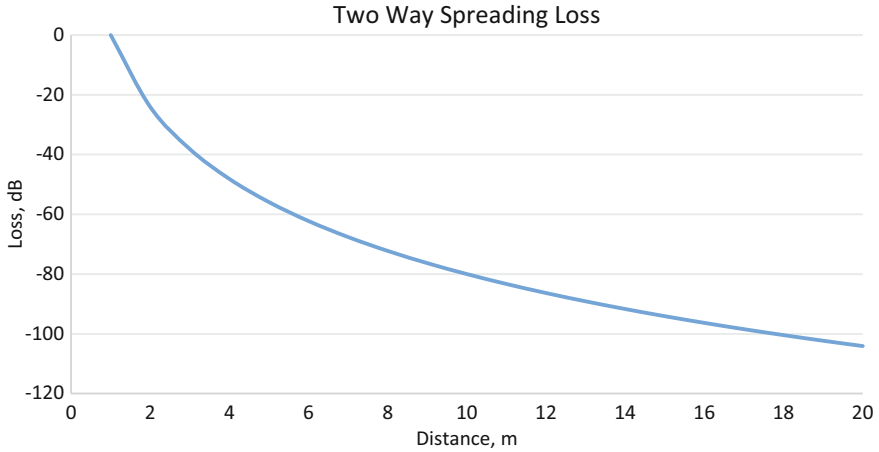


Fig. 36 Two-way spreading loss with intensity normalized to the intensity at 1 m

For example, if the *SPL* is measured at 1 m, it is expected that the *SPL* at 2 m would be -6 dB less.

As the wave front propagates, there is a spreading loss associated with it as the sphere expands. This loss in intensity for an initial intensity of I_0 is given by

$$I(r) = \frac{I_0}{r^2} \quad (32)$$

The consequence of this spreading is a rapid attenuation of the signal. As the ultrasound wave is incident on an object (in the normal direction), it will be reflected back toward the receiver. Each point on the object reflects energy and returns it as a spherical wave that also spreads with the effect that the received reflection is

$$I_R(r) = \frac{I_0}{r^4} \quad (33)$$

This two-way spreading loss is significant and is the main reason that the range of a sensor is limited (see Fig. 36). The consequence is that ultrasonic sensors are extremely sensitive at short ranges and have the ability to detect the movement of fingers. However, as the distance from the object to the transceiver increases, it is only capable of detecting major motion.

Doppler Shift

As mentioned, the sensor transmits a constant frequency sound field, and when an object in ultrasound field moves, there is an apparent change in frequency of the reflected ultrasound. Motion is sensed by detecting this change in frequency.

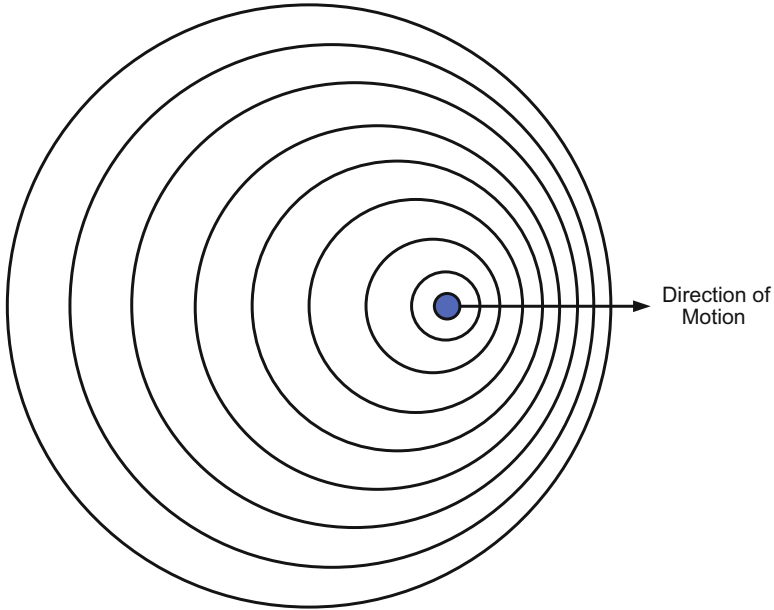


Fig. 37 Diagram illustrating the Doppler effect. Source is the *blue dot* and is moving to the right

This change in frequency is due to the Doppler effect and is the apparent change in frequency when either an object moves or the ultrasound source moves. This change in frequency is best illustrated by considering a moving sound source. Consider the isotropic sound source of Fig 37 that is emitting an ultrasonic field and is moving to the right. As can be seen, the ultrasound field to the right of the source is compressed and has a higher frequency, while the field to the left of the source of the sound field is expanded and has a lower frequency.

When considering a stationary transmitter that is collocated with a receiver, the Doppler shift is approximately given by

$$f_d = \left(\frac{2v}{c}\right)f_0 \tag{34}$$

where f_d is the amount of frequency shift of the signal, v is the velocity of the object being detected, c is the velocity of propagation of ultrasound in air, and f_0 is the transmit frequency. This approximation assumes that $c \gg v$. This also assumes that the motion is along a radial parallel to the sensor’s principle axis. The received signal is modeled as (Balleri et al. 2011)

$$\widehat{s}_R(t) = e^{j2\pi f_0 t} e^{j2\pi f_d t} \tag{35}$$

where the observed signal is $Re\{\widehat{s}_R(t)\}$. In general, human motion is a very complex phenomenon. While the person may be walking on radial course toward or away

Table 4 Male and female gait velocities and their corresponding Doppler shifts

Gender	Normal gait, m/s	Doppler shift, Hz normal gait	Maximum gait, m/s	Doppler shift, Hz maximum gait
Men	1.40	338	2.26	546
Women	1.37	331	2.08	503

from the sensor at an average velocity, the individual parts of their body will comprise many other velocities that are higher than the average velocity. This result is a description that is a sum of frequency shifts corresponding to all various velocities of the individual body parts and is given by

$$\widehat{s}_R(t) = e^{j2\pi f_0 t} e^{j2\pi f_d(t)t} \quad (36)$$

The term $f_d(t)$ is the time-varying version of the Doppler shift given by

$$f_d(t) = \sum_{i=1}^N 2f_0 \frac{v_i(t)}{c} \quad (37)$$

where the $v_i(t)$ are the individual time-varying velocities of the N various body parts. The average velocity of humans is given in Table 4 and is the average of data given in Table 4 of Bohannon (1997). Along with these velocities, the average Doppler shifts are also shown. The human gait is a periodic motion which each body part harmonically moves resulting in a rather periodic looking signal. To obtain the Doppler signal, it must be demodulated by mixing the received signal with the transmitted signal as follows:

$$\widehat{s}_R(t)\widehat{s}(t)^* = e^{j2\pi f_d(t)t} \quad (38)$$

where $\widehat{s}(t)^*$ is the complex conjugate of the transmitted signal. If there is a single motion along a radial at a single velocity, v , the signal, after low-pass filtering, would be observed as

$$s_R(t) = \cos\left(\left(\frac{2v}{c}\right)f_0 t\right) \quad (39)$$

Propagation Parameters

Propagation of ultrasound through air has two important parameters. The first is the velocity of propagation, c , which is a slight function of temperature and is given by

$$c = 331.4 + 0.6T \text{ m/s} \quad (40)$$

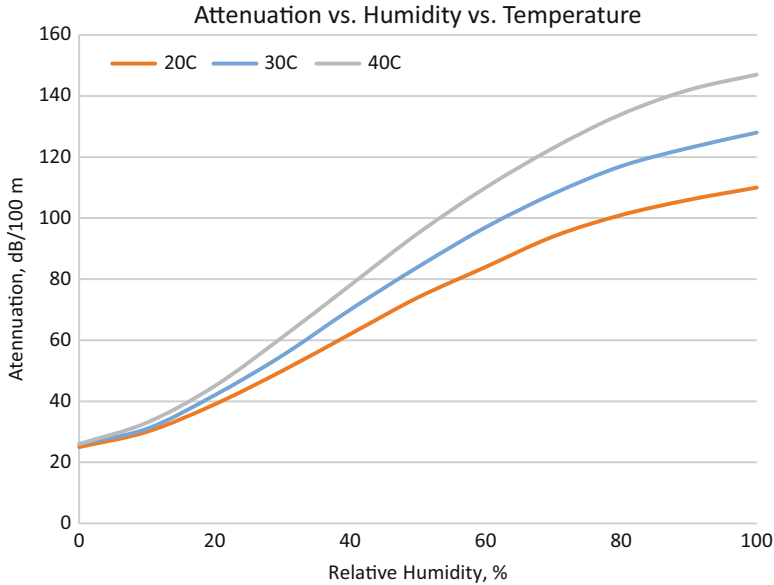


Fig. 38 Approximation of the attenuation constant at 40 kHz in air for various temperatures and humidities

The temperature in this equation is in degrees Celsius. The second parameter is the attenuation of the ultrasound as it propagates through air. The intensity varies as

$$I(z) = I_0 e^{-\alpha z} \tag{41}$$

where I_0 is the initial intensity, α is the attenuation, and z is the distance the wave has propagated. Some approximate values are shown in Fig. 38. These values are for an ultrasonic frequency of 40 kHz and are calculated (Jackett 2014; Russell 2013) using a method specified by an ANSI standard (ANSI Standard S1-26:1995). It should be noted that when determining the attenuation, the variable z should be the round trip distance to the object. The attenuation of the ultrasound as a function of distance is shown in Fig. 39 at 40 kHz and 20 °C for some typical humidities.

Diffraction

Diffraction of sound is the ability of the sound wave to “bend” around a corner. This phenomenon is the reason people can hear sounds around a corner. The phenomenon is approximately due to the Huygens–Fresnel principle. Consider a small opening upon which a sound wave is incident, Huygens proposed that as the wave encounters the opening, every point on the opening should be considered a point source from which a spherical wave emanates. As an illustration of this effect, see Fig. 40a, b in which a small opening produces spherical wave and a larger opening appears to bend

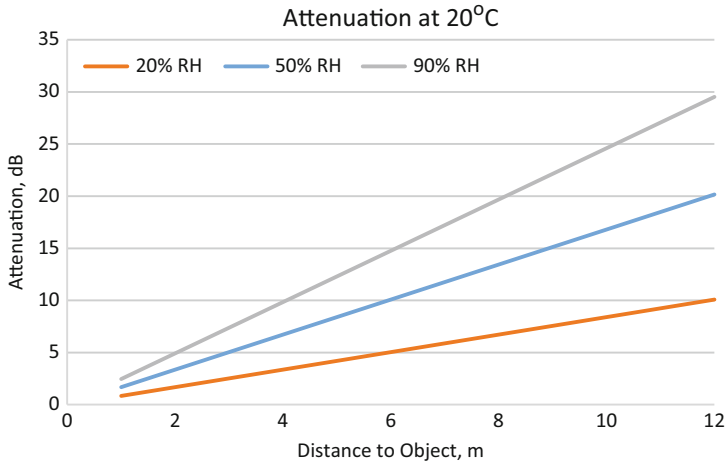


Fig. 39 Attenuation of the ultrasound signal for typical indoor environments. Note that the round trip distance is accounted for in this graph

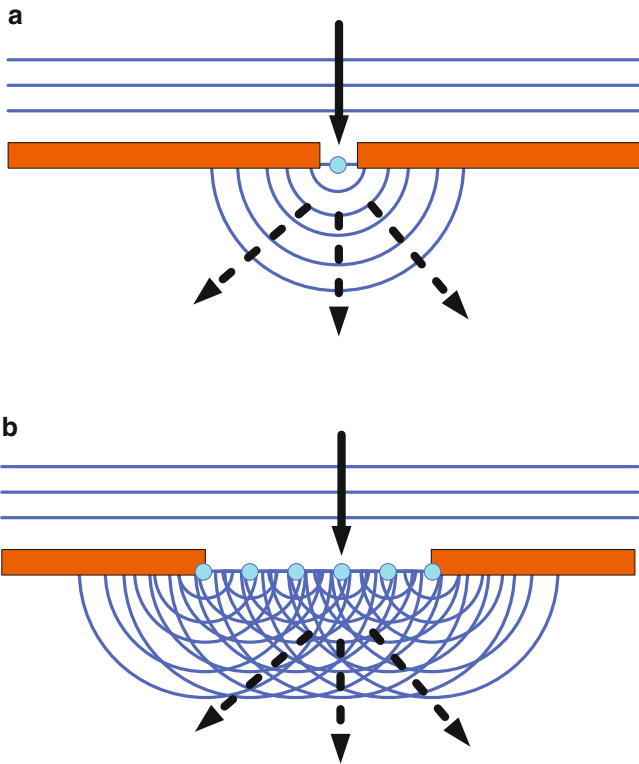


Fig. 40 Illustrations depicting diffraction of an ultrasound wave for (a) a small hole and (b) a larger slot

the wave around its corners. The exact formulation is given by the Kirchhoff's diffraction formula and is beyond the scope of this discussion. Needless to say, this phenomenon is only significant when the wavelength of the sound wave is near the size of the openings. The important application for sensors is when ultrasound with a wavelength of 0.34 in. (0.8575 cm) at 40 kHz is incident on a cubicle wall or a doorway. In these cases, the wavelength is many orders of magnitude smaller than the dimensions of these structures. As such, significant diffraction cannot be expected. It is often stated that ultrasound can “see” around corners, but diffraction is not a major contributor.

Reflections

When ultrasound propagates toward a hard surface, it will reflect back from its returning energy to the source of the ultrasound. The amount of the reflection will be determined by the reflection coefficient which is given by

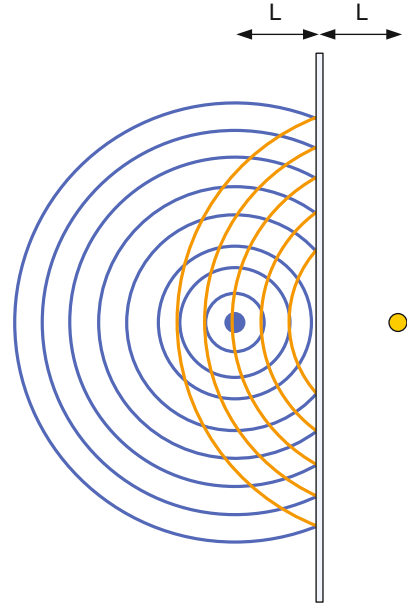
$$R = \frac{Z_{\text{air}} - Z_{\text{obj}}}{Z_{\text{air}} + Z_{\text{obj}}} \tag{42}$$

where Z_{air} is the acoustic impedance of air and Z_{obj} is the acoustic impedance of the object that is reflecting the ultrasound. This is the reflection coefficient when the direction of propagation is normal to the object. This reflection is referred to as specular reflection and is the reflection from a smooth surface that has dimension much larger than the wavelength of the ultrasonic wave. A small table of reflection coefficients for some common materials is given in Table 5 (Onda 2003). As can be seen, the acoustic impedance of air is orders of magnitude lower than most of the common materials, and hence the magnitude of the reflection coefficient is approximately one. The consequence is that most of the ultrasound energy will be reflected from objects made of these materials.

Table 5 Values of acoustic impedance for various materials

Material	Acoustic impedance, Rayls
Muscle	1.7
Skin	2
Bone	7.7
Air	0.0004
Brass	40
Brick	7.4
Glass	14
Concrete	8
Wood (pine)	1.6
Wood (oak)	2.9
Cork	0.12
PVFD (thermoplastic)	4.2

Fig. 41 Illustration depicting the image of an ultrasonic source on the other side of a hard wall



If the ultrasound is incident on a large plane object such as a wall, the ultrasound is reflected with a pattern produced as if there was another source on the other side of the wall equidistant from the actual source. This is illustrated in Fig. 41. This imaginary source is referred to as an image of the source. Recall that the two-way spreading occurs when the wave is reflected.

When the ultrasound is incident at an angle to the normal direction, the reflected wave will reflect off the surface with an angle equal to the incident wave. This behavior is described by Snell's Law (see Fig. 42). Snell's Law states that

$$\frac{\sin(\theta_i)}{v_1} = \frac{\sin(\theta_t)}{v_2} \quad (43)$$

v_1 and v_2 are the velocities of sound in the two materials. θ_i and θ_t are the angles of reflection (incidence) and refraction, respectively. When the acoustic wave is not incident on the object in the normal direction, the reflection coefficient is a function of the incident angle and is given by

$$R = \frac{Z_{\text{air}} \cos(\theta_i) - Z_{\text{obj}} \cos(\theta_t)}{Z_{\text{air}} \cos(\theta_i) + Z_{\text{obj}} \cos(\theta_t)} \quad (44)$$

Materials can also absorb ultrasound and dissipate the energy as heat. Another phenomenon that happens is scattering. When a surface is rough and has features with dimensions that are near the size of the wavelength of the ultrasound, it reflects

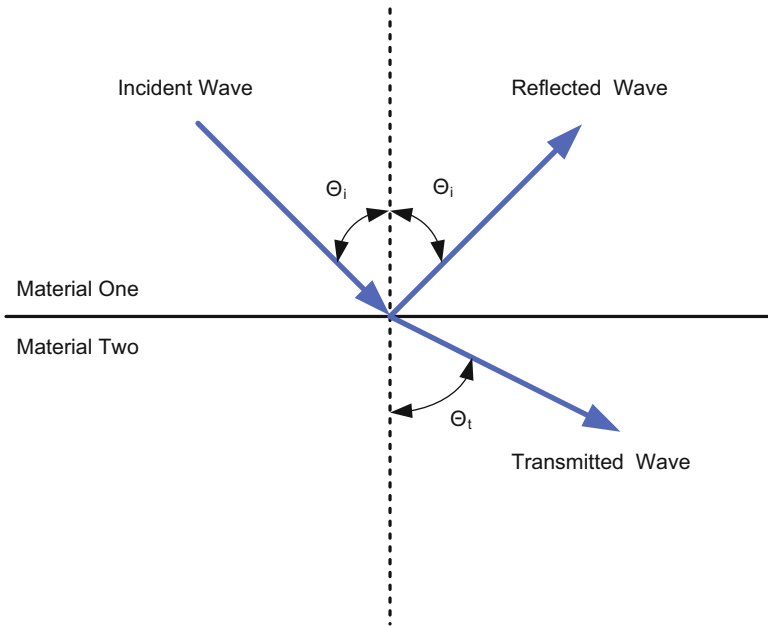


Fig. 42 Illustration of an ultrasonic wave incident on a surface at an angle

the ultrasound in many different directions. The apparent effect (from the receiver’s perspective) is that the ultrasound is attenuated since it is not reflected back to the receiver. Materials like cloth cubicle walls and carpeting absorb ultrasound, scatter it, and reflect it; however, numbers describing the amount that each contributes are difficult to find.

Returning back to the question of whether ultrasound can see around corners or over cubicle walls, the answer is sometimes. While the ultrasound may not have a direct line of sight, there may be a circuitous path through a combination of reflections. This effect is called multipath propagation. Depending upon the placement of the sensor, the effective round trip path length, the reflection coefficient of the materials involved in the reflections, the obfuscation of the transmitted beam, and the cross-sectional area of the moving object, there is a possibility that the object can be detected. An example is shown in Fig. 43, where the ultrasound has a path from the sensor to the person and back via reflections from the ceiling. At the high frequency of the ultrasound, the magnitude of the reflection coefficient of ceiling tiles is high. The major impact is the presence of the cubicle wall since it obscures a significant amount of the transmitted energy. Another typical application is a restroom which often has tile floors and walls. Tile is an excellent reflector and one would expect that the reflections from the walls, ceiling, and floor will play a significant role in the multipath propagation further enhancing the detection of a person behind the walls of the stalls.

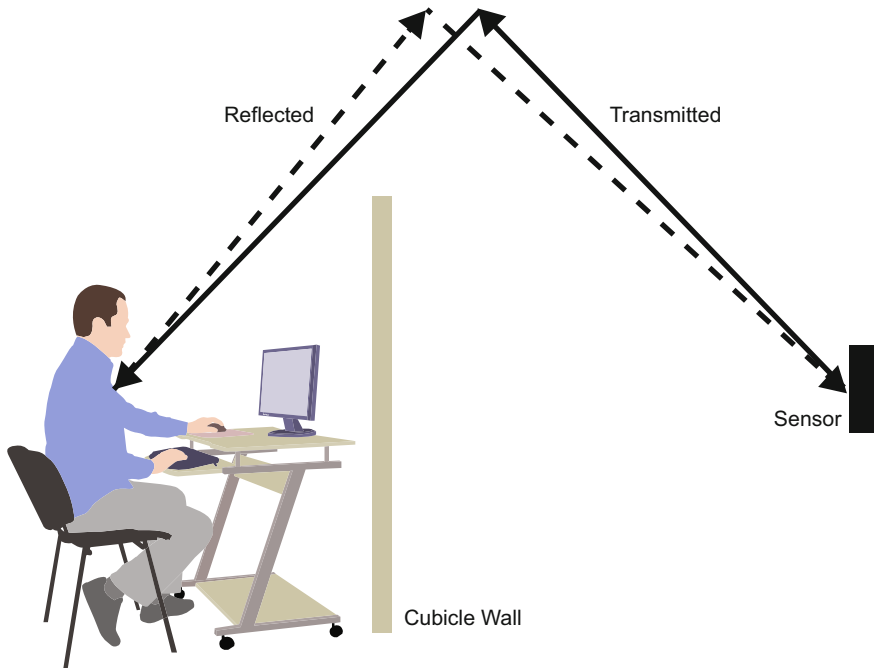


Fig. 43 Illustration of sensor “seeing” over a cubicle wall via reflections from the ceiling

References

- Aggarwal MD, Currie Jr JR, Penn BG, Batra AK, Lai RB (2007) Polymer-ceramic composite materials for pyroelectric infrared detectors: an overview, NASA/TM-2007-21590
- ANSI Standard S1-26:1995 Calculation of the absorption of sound by the atmosphere (ISO 9613-1:1996)
- Balleri A, Chetty K, Woodbridge K (2011) Classification of personnel targets by acoustic micro-Doppler signatures. *IET Radar Sonar Navig* 5(9):943–951.
- Berkley Lab (2014) Modelica Buildings Library. http://simulationresearch.lbl.gov/modelica/releases/latest/help/Buildings_HeatTransfer_Data_Glasses.html#Buildings.HeatTransfer.Data.Glasses.ID102
- Berman H (1972) Infrared intrusion detector system. US Patent 3,703,718
- Blackstock D (2000) Fundamentals of physical acoustics. Wiley, New York
- Blissett K, Dunbar R (1982) Control device responsive to infrared radiation. US Patent 4,346,427
- Bohannon R (1997) Comfortable and maximum walking speed of adults aged 20–79 years: reference values and determinants. *Age Aging* 26:15–19
- Carter B (2008) Op amp noise theory and applications, application note SLOA082, Texas instruments
- Colliard-Piraud S (2013) Signal conditioning for pyroelectric passive infrared (PIR) sensors, AN4368, ST Electronics
- Ekert G (2014) Temperature of a healthy human (Skin Temperature) energy information administration, September 2008, 2003 Commercial Building Energy Consumption Survey

- Everett P (1998) Physics of electro-optic detectors, Infrared and Electro-optic Technology Group, Everett Companies, Phoenix. <http://www.fresneltech.com/pdf/POLYIR.pdf>
- Heinze M, Yen V, Varney C (2004) Pyroelectric detectors: behavior of detector characteristics at static and dynamic temperature condition. <http://hewilliams.com/brochure/OccSensWP.pdf>, 2014
- Jackett R (2014) NPL Acoustics: Calculation of absorption of sound by the atmosphere. <http://resource.npl.co.uk/acoustics/techguides/absorption/>
- Keller HJ (2000) 30 Years of passive infrared motion detectors – a technology review. Opto/IRS² Erfurt
- Liu S (1974) Pyroelectric detector. US Patent 3,816,750
- Maniccia D, Davis R, Wolsey R, Miller N, Gross J (1997) Occupancy sensors – motion-sensing devices for lightning control, Specifier Rep 5(1):1–40
- Maniccia D, Tweed A, Von Neida B, Bierman A (2000) The effects of changing occupancy sensor timeout setting on energy savings, lamp cycling, and maintenance costs, Illuminating Engineers Society of North America Annual Conference, Paper #42, New York
- Narendran N, Yin T, Rourke CO, Bierman A, Maliyagoda N (2000) Lamp life predictor for frequently switched instant start fluorescent systems, Illuminating Engineers Society of North America Annual Conference, Paper #48, New York
- NEMA (2011) Occupancy motion sensors standard, NEMA WD7-2011
- Onda (2003) <http://www.ondacorp.com/images/Solids.pdf>
- Ravas R (1969) Movement responsive light control means. US Patent 3,459,961
- Rossin J (1973) Differential pyroelectric sensor. US Patent 3,839,640
- Russell D (2013) Absorption and Attenuation of Sound in Air, 5(1): 1–40. <http://www.acs.psu.edu/drussell/Demos/Absorption/Absorption.html>
- Schmidt W (2002) Frequency range for pyroelectric detectors for motion sensors, PerkinElmer Optoelectronics. http://kre.elf.stuba.sk/~epo/app_pyrofreqrange.pdf
- Sprout J, Berman H (1975) Dual channel infrared intrusion alarm system. US Patent 3,928,843 http://www.thermoworks.com/emissivity_table.html
- VonNeida B, Maniccia D, Tweed A (2000) An analysis of the energy and cost savings potential of occupancy sensors for commercial lighting systems, Illuminating Engineers Society of North America Annual Conference, Paper #43, New York
- Whatmore RW (1986) Pyroelectric devices and materials. Rep Prog Phys 49:1335–1386
- Philips (1985) Low-cost automatic light switching using passive infrared sensors, Philips Technical Publication 147

Ambient and Spectral Light Sensors

Sajol Ghoshal

Contents

Introduction	516
Basic Functionality of an Ambient Light Sensor	517
Filters and Multichannel Sensing	519
The Sensor Component Architecture	523
Design and Selection Considerations	524
Other Considerations	525
Field of View	525
Optical Path	525
Ripple	526
Extending the Capabilities of Lighting with an ALS	527
Creating a User Experience from the User's Perspective	528
Overview of a Sensor-Based Closed-Loop Daylight-Responsive Control Structure	529
Functional Integration	529
Component Integration of Intelligence and Sensing	531
Conclusion	533

Abstract

Silicon photocells have been used for years to inform a lighting unit of the simple presence or absence of other light in the vicinity. The most common outdoor application has been street lighting, which lends itself to the straightforward on/off control information. The needs of an advanced lighting system, however, differ from those of a streetlight or other dark/light, on/off decision scenarios. Among its other distinguishing characteristics, an advanced lighting system will be expected to adapt to specific human-centered needs within a space. While adaptation capabilities in some systems may include the ability to adjust the color qualities or color temperature of a space, it will most certainly be expected to be

S. Ghoshal (✉)
ams AG, Unterpremstaetten, Austria
e-mail: sajol.ghoshal@ams.com

able to vary illuminance levels in order to maintain a fixed level of brightness suitable to the user or task at hand, regardless of the variance in sunlight or other ambient lighting within the space. In order to accomplish this, a lighting decision engine will require knowledge of the specific amount of light within a space at any given point in time, measured in a human-centered fashion. This is the role of an ambient light sensor or ALS. This chapter will discuss the technical aspects of the ALS, including how it differs from a normal photocell, and will explore the communications and intelligence extensions that will be required for a complete sensor subsystem. A functional model will also be addressed to provide an illustration of how this important technology can be expected to be applied.

Introduction

The ability to sense the presence or absence of ambient light has been important for outdoor lighting for many decades. The introduction of smartphones and flat-panel televisions drove an entirely new need for more accurate ambient light measurement to allow the automatic adjustment of the display to enhance the user experience. Anyone who ever had to turn up the brightness of their original Personal Device Assistant (PDA) in order to view it outside, only to be blinded by the device come nightfall, understands the benefits of a more daylight-responsive display. As users have continued to demand increasingly seamless functionality from the now-ubiquitous smartphone and flat-panel TV, precision ambient light sensors have effectively become standard in even low-end models, all but eliminating the need for the user to adjust the display brightness beyond setting an initial preference. This, in turn, has increased the supply and feature sets of these devices, while driving down the cost, much to the benefit of sensor-driven lighting which currently enjoys substantially lower volumes than smartphones and TVs.

Building operators have long recognized that the users of a space tend to take a “set it and forget it” approach to their individual lighting needs. That includes simply turning on the overhead office light and leaving it on for the balance of the day. For both energy and maintenance saving reasons, in the commercial office environment, automated occupancy controls have long since replaced the simple on/off switch. While an occupancy sensor may have solved one need, one problem remains – how much light does the user really want or need? Many individuals make the adjustment themselves by turning off the overhead light and instead providing their own desk or other task lights which may serve to deliver a more comfortable light level, but unfortunately reintroduces the problem of the lights being left on when the space is unoccupied. In all cases, the incumbent lighting installations fail to address the real need, which is to deliver a user-defined amount of light, only when the space is in use, and maintain that light level as other ambient light sources, especially daylight, change.

While photocells have been in use for decades, they have typically been reserved for outdoor, simple on/off decisions related to nighttime operations. Since the

application has been driven solely by the presence or absence of either direct or indirect sunlight, the specific amount of sunlight beyond the triggering threshold has been effectively irrelevant to the functionality. In addition, there is no need to filter or distinguish the type of light since it can be safely presumed that sunlight, including its full emission range from UV to infrared, will be the only substantial source detected.

A key functional differentiator of advanced lighting will be its ability to move beyond simple on/off functionality, increased energy efficiency, or even simply better overall lighting quality into the realm of adaptation and interactivity. To accomplish this, lighting systems, or even individual luminaires and lamps, need enhanced levels of information regarding the current lighting content of the space they are tasked to illuminate. Depending upon the lighting system implementation, ambient light data can be delivered to the luminaire through a number of paths, but ultimately it will have been derived from information provided by one or more high-precision ambient light sensors.

Basic Functionality of an Ambient Light Sensor

To fully understand how to measure light, it is necessary to first discuss the units of light measurement and how they differ. Light is electromagnetic radiation that the human eye is sensitive to. Electromagnetic radiation can be characterized by its strength and how its energy is spectrally distributed. The human eye responds to energy with wavelengths between about 380 and 780 nm, with a peak around 555 nm (Fig. 1).

This is called photopic response. The strength of the electromagnetic radiation can be characterized by its power. Radiant energy, Q , measured in joules (J), is the SI

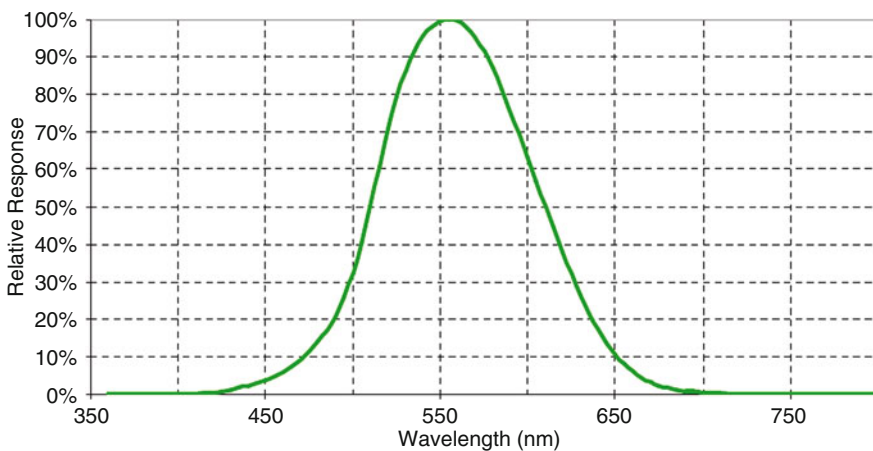


Fig. 1 Normalized photopic response

Table 1 Radiometric and photometric units

Type	Radiometric		Photometric	
	Name	SI unit	Name	SI unit
Energy	Radiant energy, Q	Joules (J)	Luminous energy, Q _v	Lumen second (lm · s)
Power	Radiant flux, ϕ	Watts (W)	Luminous flux, F	Lumen (lm)
Power/ area	Irradiance, E	W/m ²	Illuminance, E _v	Lux (lx)

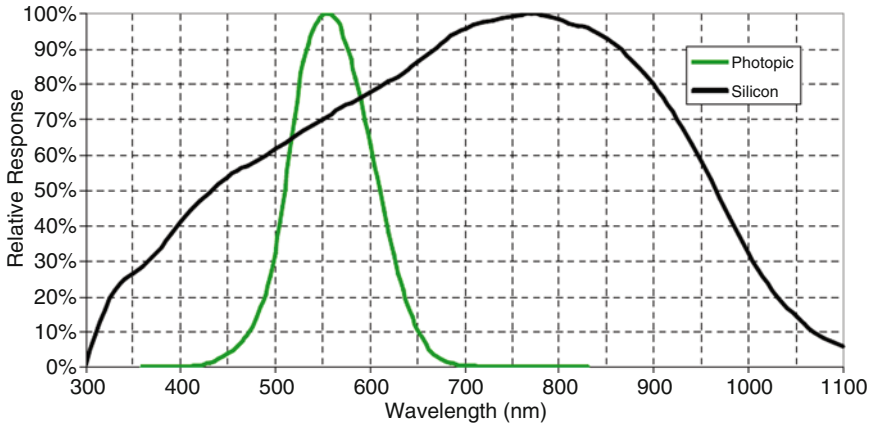


Fig. 2 Normalized silicon response compared to photopic response

unit for energy, and radiant flux, B, measured in watts (W), is the SI unit for power (energy per unit time). When radiant flux is measured per unit area, one obtains irradiance, E, measured in Watts per meter squared (W/m²).

Radiant energy, radiant flux, and irradiance are all SI units describing all electromagnetic radiation. When these measurements are weighted against the spectral response of the human eye (Fig. 1), we obtain photometric measurements. The photometric equivalents to previous units are luminous energy, Q_v, measured in lumen seconds (lm · s); luminous flux, F, measured in lumens (lm); and illuminance, E_v, measured in lux (lx). These units, both radiometric and photometric, are summarized in Table 1.

When ambient light is measured, it is the illuminance that is quantified. This gives the system the ability to make decisions based on what a user would see as a “photopic” response. The goal of the ambient light sensor is therefore to mimic the response of the human eye, in order to measure illuminance.

Silicon-based photosensors are responsive to radiation from about 300 nm to about 1100 nm. Figure 2 shows the spectral response of silicon compared to that of the eye’s photopic response.

Since silicon overlaps the photopic response, it is generally a good material for sensing ambient light. While its low cost and accessibility make it easy to implement,

the problem is that silicon is responsive to non-photopic radiation or radiation that a human eye cannot sense, such as UVA and some UVB (below 300 nm) as well as significant near infrared (NIR) below 1100 nm. The biggest component of non-photopic energy that is common in many light sources is infrared (IR) energy. Incandescent bulbs and sunlight are two prevalent examples of light sources that emit a lot of IR energy. Thus, if IR radiation is not taken into account and you were using a silicon photodiode to determine the amount of ambient light on a surface lit by an incandescent bulb, your silicon-based reading would show that the ambient light were much higher than a viewer would perceive it to be.

As a result of the non-photopic emissions within the different spectral power distribution (SPD) shapes of light sources, while it may be possible through empirical measurements to get a correlation between a photodiode sensor output and a photometric quantity for a specific light source, in general the results will have large errors and cannot accommodate an unspecified mix of different light sources (such as sunlight entering a room where there are also nearby fluorescent or incandescent luminaires). To get a photopic response with a silicon photodiode, it is necessary to filter out this unwanted radiation from either the sensor or from the integration results.

Filters and Multichannel Sensing

There are several ways to accomplish this filtering. The apparently simple and obvious choice would be the inclusion of UV- and IR-blocking filters in the optical path between the light source and the sensor. This can be an excellent way to achieve a photopic response with some of these filters are capable of effectively blocking out over 90 % of non-photopic radiation. This is the preferred method for many calibrated lux meters; however, the combination of increased production costs, including precision alignment of the filters, as well as issues such as variation over time, makes this option less desirable for the high-volume, long lifetime solutions that lighting would demand.

An alternate approach is to make use of the fact that different wavelengths of light contain different amounts of energy. Shorter wavelengths of light contain more energy than longer wavelengths, and as one would expect, the higher the energy level, the more rapidly it will be absorbed by the silicon. In other words, the light with less energy will have to penetrate into the silicon deeper than the light with more energy.

For example, imagine that a light with a peak wavelength at 450 nm and a light with a peak wavelength at 650 nm strike a silicon photodiode. For simplicity, these two sources will henceforth be referred to as the blue light and the red light, respectively, although it is not completely accurate to describe them so. Since there is an inverse relationship between the wavelength of light and the energy contained in a photon of that light, the blue light will contain more energy than the red light. According to Eq. 1, the 450-nm light delivers 2.76 eV, while the 650-nm light delivers only 1.91 eV of energy to the silicon:

$$E = \frac{hc}{\lambda} \quad (1)$$

E is the Energy of a photon (eV).

h is Planck's constant, 4.14×10^{-15} (eV s).

c is the speed of light, 3×10^{17} (nm/s).

λ is the wavelength of light (nm).

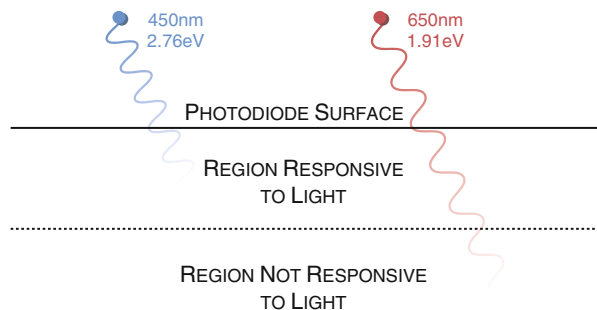
The upper limit of the silicon responsivity curve is determined by the smallest amount of energy that will elicit a response from silicon. This amount of energy is the bandgap voltage, and for silicon, it is approximately 1.12 eV at 25 °C. Using Eq. 1, it can be seen that this corresponds to a wavelength of 1109 nm, which is the general upper limit of silicon's spectral response to long wavelengths.

The higher energy of the blue light will be absorbed by the photodiode more quickly than the energy of the red light, for a given silicon thickness. In other words, the red light will penetrate into the silicon deeper than the blue light before it is completely absorbed. Since the absorbed energy can only induce an electric response within the active diode region, the amount of energy absorbed will be dependent on the depth of the active diode region. If the effective light responding region of the photodiode is not deep enough, some of the red light will not be absorbed, and the response will be less sensitive to red light. If, however, the effective region depth is sufficiently large, more of the energy in the red light will be absorbed. Figure 3 shows a simplified cross section of a silicon photodiode. In this figure, some of the red light is not being absorbed within the light responding region, while all of the blue light is.

By channelizing a photodiode, one can simultaneously measure the response to electrons generated at different silicon depths, which will indicate not only the total illuminance but also the mix of both higher and lower energy wavelengths. With this "dual photodiode approach," the result would be a two-channel response. Channel 0 is representative of the full-range silicon response, while channel 1 has an IR-only response. Figure 4 maps the normalized spectral response of both channels of an example ALS along with the standard photopic response curve.

As one can see, Channel 1, which would be measuring the response deeper in the silicon, has very little response to the more energetic UV while being substantially

Fig. 3 Absorption of different wavelengths of light in silicon



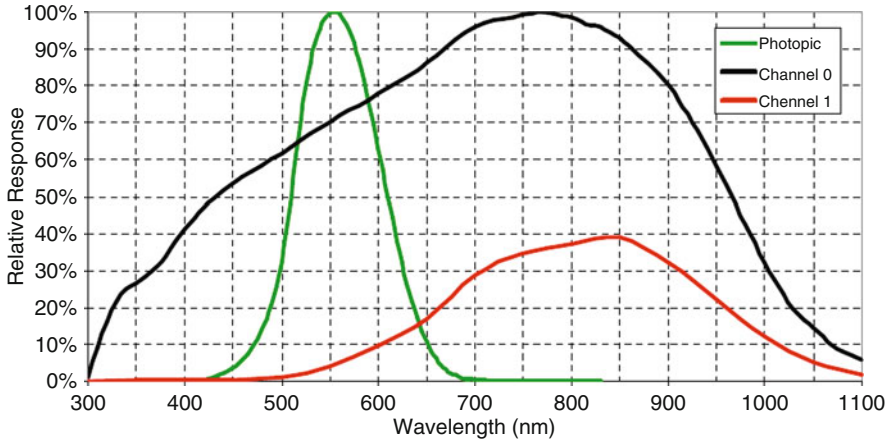


Fig. 4 Normalized silicon response compared to photopic response

responsive to the NIR region. The ratio between these two channels (Ch1/Ch0) can be used to generate a piecewise linear lux equation. By utilizing this type of lux equation developed by the manufacturer and specific to each model of ALS, the luminaire designer can synthesize an accurate picture of the photopic illuminance values. Below is an example of such a linear lux equation for a hypothetical ALS:

Normalized at $1 \times$ gain and 400 ms integration time,

$$\begin{aligned}
 &\text{For Ch1/Ch0} = 0.00 \text{ to } < = 0.25 : \\
 &\text{Lux} = 0.105 * \text{Ch0} - 0.208 * \text{Ch1} \\
 &\text{For Ch1/Ch0} = > 0.25 \text{ to } < = 0.38 : \\
 &\text{Lux} = 0.1088 * \text{Ch0} - 0.2231 * \text{Ch1} \\
 &\text{For Ch1/Ch0} = > 0.38 \text{ to } < = 0.45 : \\
 &\text{Lux} = 0.0729 * \text{Ch0} - 0.1286 * \text{Ch1} \\
 &\text{For Ch1/Ch0} = > 0.45 \text{ to } < = 0.60 : \\
 &\text{Lux} = 0.060 * \text{Ch0} - 0.10 * \text{Ch1} \\
 &\text{For Ch1/Ch0} > 0.60 : \\
 &\text{Lux/Ch0} = 0
 \end{aligned}$$

The true test of an ambient light sensor approach is how well it performs when compared to a calibrated light sensor under different lighting conditions. Figure 5 shows how well one sensor using this dual-channel approach tracks lux measured against a calibrated commercial lux meter.

A third approach to deriving ALS data under a photopic response involves using data from a color sensor. This method utilizes the separate responses from individual red-, green-, and blue-filtered photodiodes, as well as a clear channel, to allow a full comparative evaluation of the spectral power distribution, as seen in Fig. 6.

The chip scale package used to house a color sensor will typically consist of an additional glass overlay with an integrated IR-blocking filter. This allows for the

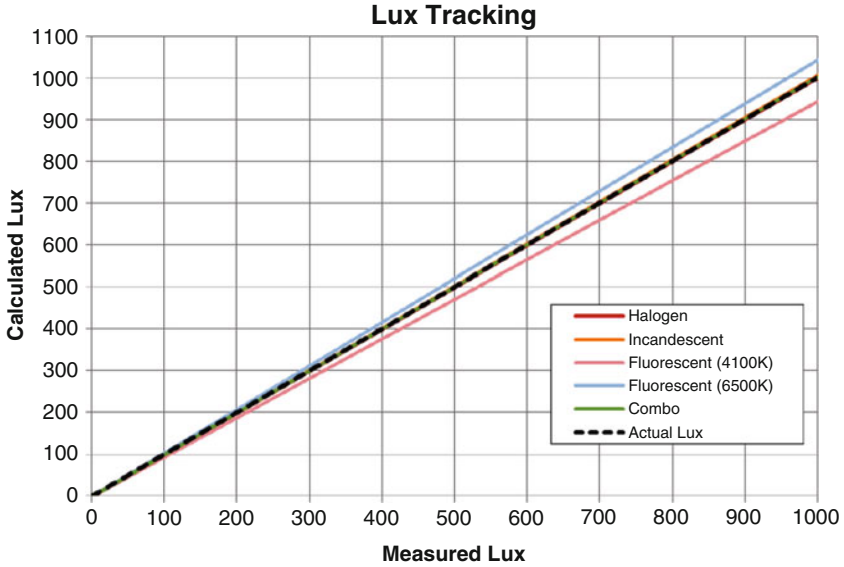


Fig. 5 Representative ALS lux calculation versus actual lux

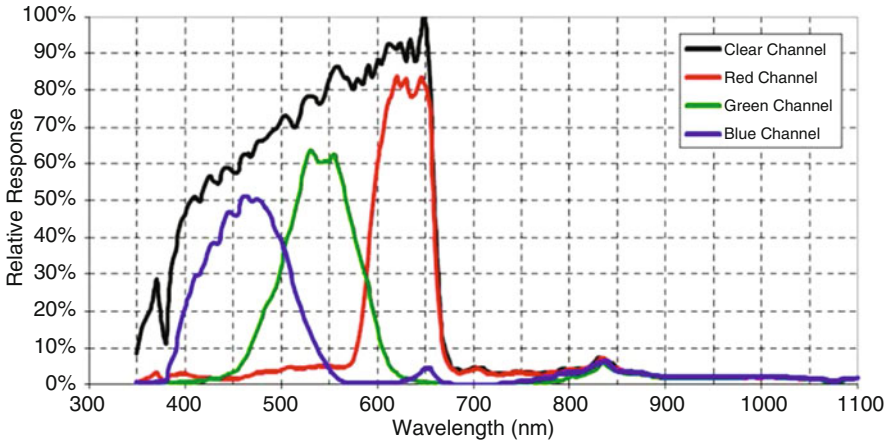


Fig. 6 Example spectral response of a color sensor

very sharp cutoff at around 660 nm. Using a weighted average of the three channels, one can obtain a fair approximation of a photopic response. When compared to an IR/UV-filtered ALS or the more cost-effective dual-channel sensors, the additional filtering and channels in a color sensor will necessarily add a substantial amount to the component cost. The benefit comes in applications which can also make use of the resulting data to calculate and subsequently respond to correlated color temperature (CCT), in addition to the lux values.

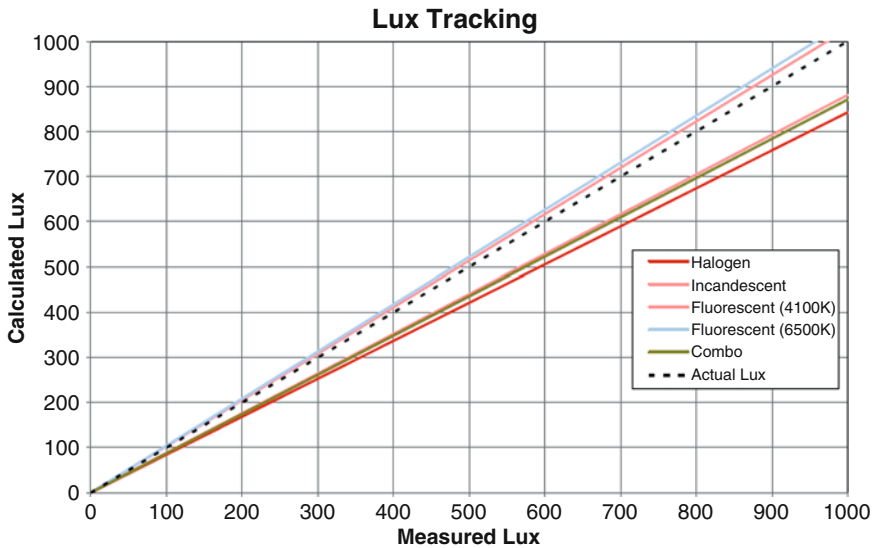


Fig. 7 Representative color sensor lux calculation versus actual lux, as measured with a calibrated lux meter

As shown in Fig. 7, the lux tracking with this method is not as good as it is using the dual-channel approach as the sensor is more sensitive to differences in the SPD of different light sources.

The Sensor Component Architecture

With an understanding of how the ALS can absorb and differentiate different types of photonic energy, it is reasonable to explore the more detailed aspects of the sensor and its integration into a lighting system or luminaire.

As discussed so far, in order to produce a photopic response, the incoming photonic energy is either filtered or spectrally limited in some fashion through the use of differing material or lenses, or the sensor response can be evaluated across multiple channels, as well as implementing a combination of these approaches. In addition, the sensors can deliver the resulting data in a variety of ways. In the simplest implementations, the sensed energy is translated by multipliers or other transformations to produce a varying analog voltage output. While that may be convenient for direct control of the output of a nightlight, in an advanced lighting system, that analog output needs to be converted to digital data through an analog-to-digital (A/D) converter in order to be put to use by intelligent decision engines.

The precision of the analog-to-digital conversion is extremely important since it will define the number of steps that can be evaluated across the sensor's lux range. Interior lighting, as an example, will usually operate in the range of 0.1–1000 lx. A 12-bit A/D converter will provide 2^{12} or 4096 steps across that range, which

translates to 1/4-lx steps. However, if that same 12-bit conversion is required to deal with sunlight, which can range up to 200,000 lx, those same 4096 steps represent 50 lx each which can be the difference between the light under a full moon and that in the family living room and therefore effectively meaningless in terms of granularity at lower lux levels. For a sensor that is capable of detecting the full range between a very dim room and full sunlight, 16-bit conversion provides a much more useful 3-lx step.

When the ultimate range of sensor is not needed, such as in an interior sensing application, an intelligent ALS will likely provide some user selectability with regard to the time over which the sensor's sampling cycle occurs. Shortening the integration time will lower the amount of energy that is sensed, effectively increasing the upper range of the ALS at the expense of accuracy. From the example above, while a 100 ms integration time may yield a range of up to 200 klx, increasing that to 400 ms will decrease the range by a factor of four, which correspondingly increases the accuracy by the same amount. The 16-bit A/D converter would then provide steps of just 3/4 lx across the revised 50 klx range.

The photodiode itself is, by its nature, an extremely low current device, operating in the range of 100 s of nanoamps. As a result, careful filtering of the photodiode's analog output is required. When implemented through A/D conversion, the signal filtering can be accomplished with signal processing techniques that can identify and eliminate signal noise that could otherwise adversely affect the lux data. With an accurate digital output to work from, a more sophisticated ALS will have on-board intelligence sufficient to actually calculate the piecewise linear lux equations that were illustrated earlier. The result in that case is a calibrated lux value, which can be output to the decision engine to implement the specific dimming strategy appropriate to the installation.

The size of an ALS is most directly related to the low end of its detection range. As we have seen, the lower the energy of the photon being detected, the thicker the silicon will need to be for accurate detection. Similarly, in order to detect extremely low lux levels, a larger sensor surface area is required. In the case of general lighting applications, the relevant minimum lux value of a space is high enough that wide-range, high-performance sensor can be quite compact. One current example from a leading manufacturer is specified for a detection range of 3–220,000 lx (very dim room to very bright sunlight), while the complete chip scale package, including photodiode array, integrating A/D converter, signal processing circuitry, lux calculation logic, and a serial output interface, is all contained in a 2 mm × 2 mm (just 4 mm² or 0.006 in.²).

Design and Selection Considerations

When selecting an ALS, a number of elements should be considered, especially in reference to the level of component integration that the luminaire designer will want to undertake. Available sensors range widely in terms of overall integration. At one end of the spectrum, an ALS may only provide the aforementioned uncompensated

analog voltage output, allowing the system integrator to fully customize the signal filtering, conversion, and lux calculations, as well as the type of output interface that data is delivered through. While that may be an exciting challenge for a handheld light meter designer, it rarely falls within the purview of a lighting system or luminaire designer. At the other end of the integrated device spectrum, the device could contain the photodiode array, A/D conversion, signal processing, lux calculation logic, and some form of standard interface all on a single CMOS integrated circuit as well as those selectable integration times to enable the integrator to adjust the sensitivity of the device. Added features can include specific power management modes, such as continuous operation, power save mode (in which the device inserts a power-saving state between each acquisition), or single-cycle operation (in which the device enters a power-down state after data acquisition until reawakened by a trigger event), both of which can reduce the already low 0.1-mA operating current to something on the order of a few micro-amp standby condition, which is primarily of benefit in portable type applications.

Other Considerations

Other ALS considerations include the device field of view and optical path, as well as the ability to reject lighting fluctuations as would be generated by sources generating 50- or 60-Hz ripple and the type of data interface choices available.

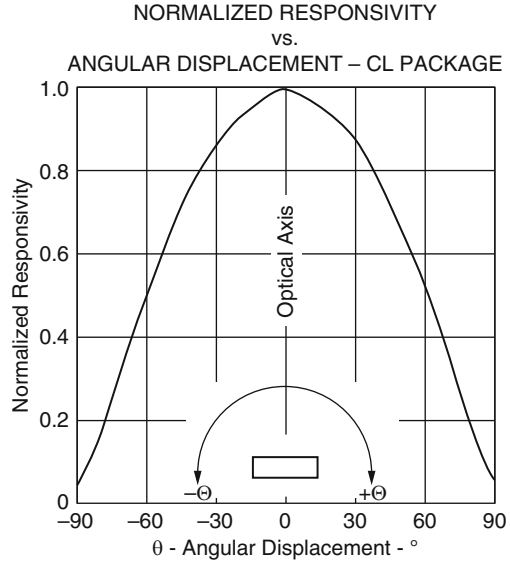
Field of View

Given that the individual or array of photodiodes in an ALS consist of flat silicon chips, it is natural to conclude that the highest level of reactivity will be the result of near vertical penetration of the photonic energy and that is indeed the case. While the literal view that an unobstructed ALS has of the space can approach 180°, the response curve diminishes rapidly off of the vertical axis and can be plotted to assist with system design considerations. An example of an angular response curve is illustrated in Fig. 8.

Optical Path

The optical path is the path from the light source to the sensor. In a real-world installation, the ALS will not be externally surface mounted on the face of a luminaire. In fact, a look at your ALS-equipped smartphone will suggest to you that it may even be mounted behind dark glass (making clear the need for low lux sensitivity even for applications that do not need to discern the difference between somewhat and really dark!). As a result of the real-world mounting and installation considerations, items such as apertures, lenses, and filters might need to be employed and can obviously have an impact on the results from the sensor. Therefore it is

Fig. 8 Normalized responsivity versus angular displacement (courtesy of ams AG)



important that the existence of these variables needs to be understood before designing them into an optical system.

An aperture is any opening, through which light is admitted. An aperture can affect the angular response of the sensor as it will limit any off-axis light. The degree to which the angular response is affected is determined by the size, shape, and depth of the aperture. A narrow and/or deep aperture would result in a very narrow angular response, whereas a very wide and/or shallow aperture would result in an angular response that more closely represents the native field of view of the stand-alone sensor package.

When light travels from one medium to another, refraction can occur if the two mediums have different indices of refraction. Lenses utilize this principle to bend light to a specific end. Lenses can be used to increase the angular response and/or the optical sensitivity of the sensor. By bending off-axis light toward the sensor, the response may be greater than it would have been otherwise. Additionally, lenses can be used in low-light applications to magnify the light incident upon the sensor.

Ripple

Although not typically visible to the eye, light sources that are powered by alternating current (AC) power lines exhibit varying light intensity as the voltage cycles across the zero value on its way toward its 110-/220-V peak. This phenomenon is known as AC ripple or AC flicker and is common in both incandescent and fluorescent bulbs. Since most AC lines in the world operate at either 50 Hz or 60 Hz, AC-powered light source intensities will vary at frequencies of 100 Hz and

120 Hz. The frequency is doubled because the light source will be the least intense at the zero crossing of the AC line and maximum intensity when the AC line is at a maximum or minimum value. One method for compensating for AC ripple is to integrate the response signal for a period of time that includes only whole integers of AC ripple periods, thereby absorbing a full AC cycle rather than a snapshot which would either be brighter or dimmer than the human eye actually perceives it to be. To accommodate both 100-Hz and 120-Hz AC ripple frequencies, a minimum integration time of 50 ms would typically be employed. Thus, any integration period of any multiple of 50 ms should be immune from variation due to AC ripple. Another method to compensate for AC ripple is to effectively slow the response time of your photodiode by incorporating an low-pass resistor-capacitor (RC) filter delay circuit. The RC circuit makes the output signal more uniform and less reactive to AC ripple.

Extending the Capabilities of Lighting with an ALS

A core attribute of an advanced lighting system will be the ability to “think” and respond autonomously to maintain the user’s preferred and productive lighting level in response to changes in the available ambient daylight. While some leading-edge lighting installations have attempted to respond to or “harvest” daylight across larger spaces, the advent of more flexibly controllable luminaires has opened the door to more granular control, which will result in substantial additional energy savings and enhance user comfort and productivity (Fig. 9).

With the basic operating principles of an ambient light sensor now understood, an example of ALS integration into luminaire, or equally feasible, an individual

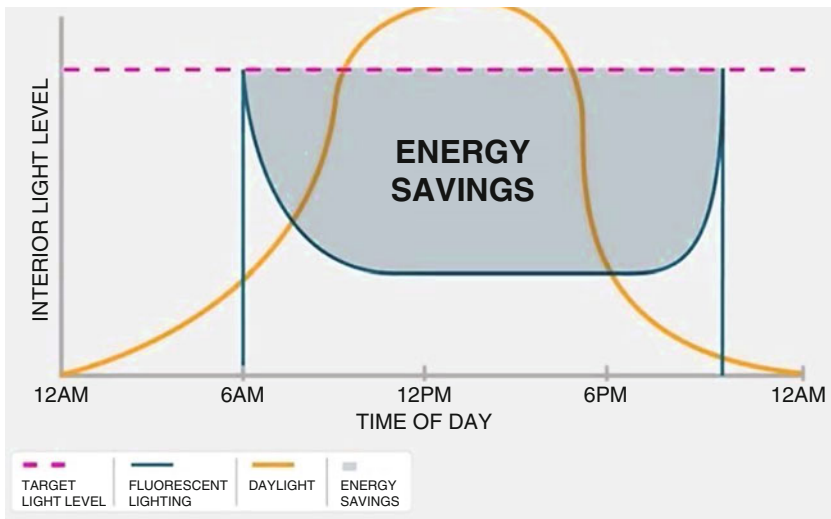


Fig. 9 Example of potential energy savings from a daylight-responsive lighting system

replacement lamp, will highlight the extensions of this technology. There are a number of important considerations in the design and implementation of a daylight-harvesting or daylight-responsive lighting system.

As a result, a daylight-responsive system design is best addressed at the luminaire or lamp level, rather than attempting to instigate daylight responsiveness at the building level. Daylight-responsive lighting system can be greatly simplified when the sensors and intelligence reside at the luminaire level. In its most basic sense, if you design a luminaire (fixture) or replacement lamp to correctly respond to and compensate for the ambient daylight that it senses, the building-level task of lighting management has been simplified by orders of magnitude.

Four basic ingredients are needed to implement a daylight-responsive luminaire design:

- Precision ambient light sensors, specifically optimized for daylight harvesting – the “daylight sensor.”
- A microcontroller-based intelligent lighting controller (ILC) that serves as the light-processing “decision engine.”
- A dimmable driver or power supply (also referred to as a “ballast” in fluorescent systems).
- The light engine – in the past that would have most likely been a fluorescent tube but with a current luminaire design. It is more likely to be an LED-based solution in which proper consideration has been given to optical design and thermal management.

Creating a User Experience from the User’s Perspective

To implement an architecture that results in a rich user experience, it is useful to clarify the functionality *from the user’s perspective*, as well as also defining how it should not behave. Current occupancy-based systems are simply automated on/off switches that pop full-on when presence is detected and turn full-off when no presence is detected after a preset period. Incumbent systems, especially the most common fluorescent-based implementations, often need to set long time-out periods in order to avoid on/off cycle times that can be detrimental to the lamps themselves. The common lifetime rating of a lamp actually assumes that switching does not exceed eight times per 24 h. An on/off switching frequency beyond that daily rate can result in a dramatic reduction of life and ultimately contributes nothing to user satisfaction.

What the user really wants is the light set to a particular level in the space and then for the luminaires to maintain that constant level as long as the space is occupied. Other than when they are actually setting the desired level, they will not want to see adjustments happening. In a sense, the “dimmer” should not be perceived as controlling the light output (in reality, the power input) for the luminaires it controls but rather as an overall ambient level controller that sets the brightness *of the space*.

Overview of a Sensor-Based Closed-Loop Daylight-Responsive Control Structure

With the dimming control function defined as the “target ambient level setting,” the responding luminaire then has a straightforward task to accomplish: Hold the illumination level of the room constant. The process of accomplishing that depends upon utilizing a high-quality ambient light sensor that can measure and integrate the lux level within the space and then adjust the luminaire’s output to maintain the target value.

The task of the daylight-responsive luminaire is to measure the lux value that is reflected back at it from the target area for a particular set point. That task is accomplished by the ambient light sensor which needs both the appropriate photopic lux range and small detection steps that are enabled by that 16-bit A/D converter. Also as earlier discussed, it is important that the ALS is only responding to a realistic average lux value and not to cyclic 50-/60-Hz peaks and valleys, such as might be generated by an older fluorescent fixture in the nearby hallway.

It should be noted that since the measurements are made on reflected lux values, there could be variations introduced based upon differing reflectivity of the surfaces in a room. A shiny conference table, for instance, would reflect back substantially more light than the carpeted floor. For practical implementations, where the ceiling height is rarely as low as 8 ft and more likely 10–12 ft in a modern building, reflective scattering helps average things out, properly allowing the more major effect to be whether the room is dominated by table tops or by open floor space, which themselves determine the overall ambient effect of any lighting in that space.

The illustration above provides an example of functionality of daylight-responsive luminaires in an office type of space. In this design, each luminaire acts independently based upon the target illumination level selected by the user. As daylight enters the room, each luminaire will operate in a closed loop mode, sensing the lux value and making subtle dimming or brightening adjustments to maintain the target level. It is important to note that occupancy sensing remains a critical function, and while illustrated as a wall-mount unit in this case (presumably co-located with the illumination level control), they could just as easily be integrated into the individual luminaires.

Functional Integration

Architecturally, the daylight-responsive luminaire (Fig. 10) will have a straightforward implementation that is greatly assisted by the intelligence and communications capabilities built-in to state-of-the-art components. In this example, the ambient light sensor is positioned to allow it a clear field of view of the target space and is connected to the ILC via an industry-standard serial bus, such as I²C. The ILC serves as the decision engine, accepting data from the ALS, as well as inputs from the occupancy sensor and illumination level control, which would typically make use of a standard 0–10-V control signals in this illustration. The ILC drives an output

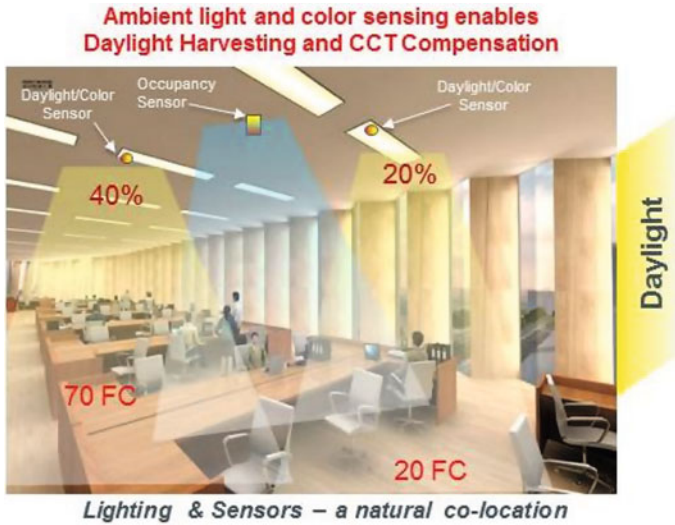


Fig. 10 Illustration of a daylight and occupancy responsive office space

0–10-V signal to any dimming ballast or LED light engine that incorporates 0–10-V input dimming controls.

In terms of specific functionality, the ILC initially compares the ambient lux as detected by the ALS to the user-selected target level of illumination, as indicated by the dimmer setting, to determine the appropriate 0–10-V output to drive the dimming ballast or LED driver. The dimmer can be replaced by a preset potentiometer in designs where direct user adjustment is not needed (such as a warehouse space) (Fig. 11).

As sunlight enters the room, the ALS senses the increase in ambient light in the room, reducing the 0–10-V signal proportionally to maintain a constant ambient lux as determined by the user-programmed target. The light engine dimming is performed on a nonlinear scale optimized to provide smooth dimming that would be a more visually pleasing experience to the human eye, enabled through programmable ramp times and adjustable lux goal targets.

In the daylight-harvesting mode, the ILC should additionally minimize abrupt changes to the lighting that could result from short-term variance in the ambient lighting conditions. The use of rapid fluctuation timers used to filter the ambient environmental noise can accomplish this by allowing the ILC to resample the environment and if the same level change is detected in sequential sampling, only then would the 0–10-V output be adjusted in the direction indicated by the change that was sensed.

Communications capabilities are very important to the daylight-driven functionality of the space, both in terms of serving building and campus-level energy management goals, such as demand-response functions, as well as to ease the overall task of integration. In the architecture above, the I²C bus serves as an effective

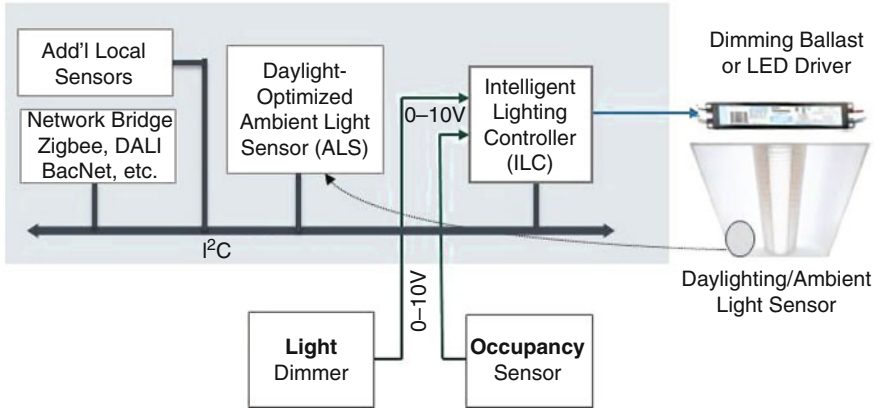


Fig. 11 Daylight-responsive luminaire control architecture

“open” platform to interconnect the ALS, as well as other sensors, with the ILC. Additionally, component-level solutions are widely available to bridge the I²C to other types of standard communications protocols that allow wider data reporting and communications, including ZigBee and DALI from the lighting side and BACnet or Ethernet/WiFi for the building management system (BMS) side of the equation.

Both approaches will be critical as good bidirectional communication will allow a more “integrated autonomy,” especially as it would apply to larger spaces where there may be a dominant daylighting influence around the building exterior, while supplemental lighting is the only influence toward the interior. The response from individual luminaires will be greatly smoothed if they “understand” what the surrounding luminaires are interpreting as the ambient conditions, allowing a more coordinated response. In addition, if a demand-response directive has been issued at a campus level, a coordinated group of luminaires could “agree” on what their cumulative contribution could be based upon ambient and occupancy conditions and report that back to the BMS before taking action. With the available contributions tabulated building wide, the BMS can issue the execution commands including adding to or reducing the amount of needed contribution from a lighting group and letting the group work out the best solution based upon actual ambient and usage conditions at the time. All this is enabled by highly localized sensing combined with intelligence and communications.

Component Integration of Intelligence and Sensing

In terms of electronic for the luminaire, Fig. 12 diagrams an example schematic to accommodate sensing, power, and drive circuitry to allow the daylight response implementation.

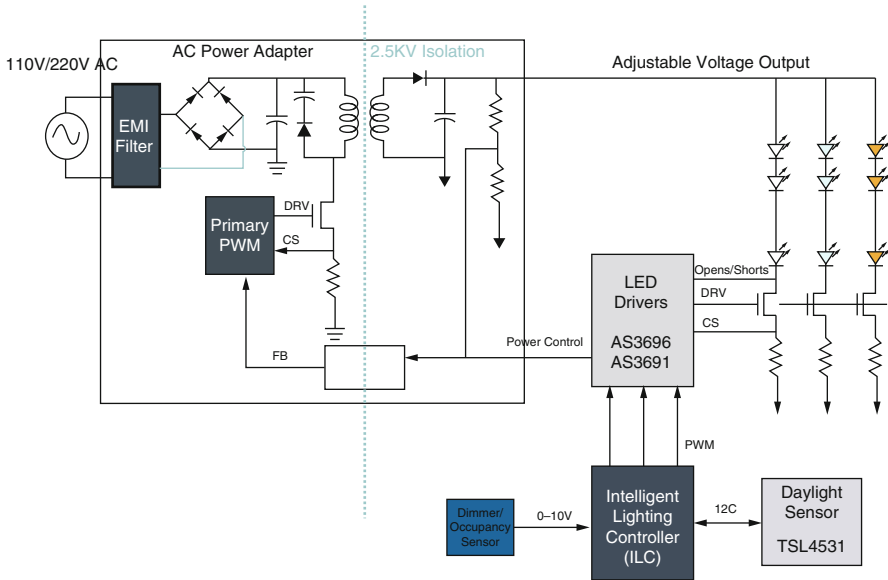


Fig. 12 A full lighting solution from AC to lumens in an LED-based luminaire

In this architecture, power is fed from the AC to DC converter, providing the 2.5 kV isolation with EMI and power factor correction. The LED driver supports two control loops: (1) The constant current driver loop which maintains a fixed current through the bottom of the LEDs and (2) the power control loop which maintains the exact output voltage headroom needed on the top of the LEDs by modulating the feedback tap of the AC/DC converter connected to the primary pulse width modulation (PWM) through the opto-coupler. This single-stage conversion delivers the highest power efficiency. The LED drivers should match the current in each string, as well as driver-to-driver, to achieve output and color consistency between strings and luminaires. For commercial quality applications, the strings should be matched to 1 % or better.

The ILC is constantly monitoring the occupancy sensor, reading information from the dimmer to determine the target illumination set point, and sensing the illumination of the room via the daylight sensor to which it is connected via an I²C port. Each occupancy event resets the “time on” timer, which should be a preset based upon the type of room and other parameters determined by the facility operator. In the on state, the direct lux output from the daylight sensor is continuously compared against the user-set target illumination level, and the light output is dimmed or increased to meet the target ambient lux value.

Conclusion

As we have learned from the smartphone revolution, the value of any service is now being judged not just by what it facilitates but also by the user experience along the way. Lighting is a critical service and one that is solely dedicated to meeting very specific user needs. With an understanding of the technical aspects of ambient light sensing, along with the an awareness of the design considerations both at the sensor and luminaire levels, a designer will be more readily equipped to integrate all of the sensing and communications functions that will be needed to meet the demands of truly advanced lighting systems.

Adaptive Distributed Sensing and Control Methods

Zhenhua Huang, Fangxu Dong, and Arthur C. Sanderson

Contents

Sensing, Estimation, and Control in Smart Lighting	536
Distributed Lighting Systems	538
Distributed Sensing of the Light Field	539
Adaptive Sampling	542
Light Field Estimation and Interpolation	546
Optimal Source Configuration	548
Adaptive Control Using Sensor Feedback	551
Future Directions for Smart Lighting Systems	555
References	555

Abstract

Innovation in solid-state lighting technology is rapidly expanding the potential for new functionality as well as the range of impacts and applications. While traditional illumination is based on a set of fixed sources that satisfy localized needs for task and ambient lighting, solid-state lighting systems propose to integrate distributed sources and sensors across the lighting space. The target “lighting field” expresses the needs of users for illumination beyond basic needs to consider energy efficiency, worker productivity, occupant health, diverse information needs, and entertainment. The lighting field is monitored by distributed sensors and closed-loop adjustments are implemented automatically to achieve the goals of the illumination system. Design of such advanced lighting

Z. Huang (✉) • F. Dong • A.C. Sanderson
NSF Engineering Research Center for Smart Lighting, Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA
e-mail: sandea@rpi.edu

systems will require new methods of representation and analysis of distributed lighting sources and sensors. This chapter introduces methods that are being explored within this new research framework. Basic methods of adaptive light field sampling, representation, estimation, and control are presented, and future directions for adaptive distributed sensing and control methods are discussed.

Sensing, Estimation, and Control in Smart Lighting

The rapid development of solid-state lighting in recent years has enabled the possibility of creating high-quality light to meet diverse requirements and provide better controllability than conventional light sources. Smart lighting technology may be designed for energy efficiency, using automated control such as automatic dimming integrated with centralized building control systems. Energy consumption is always one of the major concerns. Statistics reveal that lighting utilizes 22 % of the energy consumed by commercial and industrial markets, which consumes as much as 65 % of the total energy in the USA (Veitch et al. 1998). Smart lighting systems can sense and communicate and are integrated with novel control systems that meet the expectations and requirements of different people and different applications. These can be implemented in both open-loop mode and closed-loop mode. Research has been done on the sampling and control of smart lighting systems (Thapan et al. 2001; IESNA 2000). In open-loop mode, the requirements for energy consumption and the desired light field are utilized to calculate the configurations before sending commands to light sources. In closed-loop mode, the targets are integrated with feedback from sensor readings, and updated configurations are sent to light sources. In both modes, the process of determining optimal light source configurations from the target is necessary. The difference between them is how frequently to implement this process.

In this chapter, we address the broad problems of adaptive sensing of lighting fields and the determination of desired light source configurations to meet the requirements of a specific target light field. In this approach, a lighting domain may be defined by a discrete set of light sources, where each light source has a location (in three dimensions), an orientation, and an illumination model that describes the distribution of light rays in space (and, in general, the spectral characteristics of the light). Two key challenges occur in this process. First, the design space may be very large. The number and range of candidate solutions for the light sources may include many alternatives in the search space regions. Second, there may be more than one satisfactory solution to the design problem. The search space may have many local extrema, and several alternative design configurations may adequately meet the system performance requirements. In this case, the cost function is termed “multimodal,” and a design strategy that can explore many of these alternatives has distinct advantages.

In recent years, the goals of lighting design for both commercial and residential buildings have begun to shift from “visibility,” that is, providing necessary

illumination for occupant functions, to “lighting quality” that includes a broader range of occupant needs; economic, energy, and environmental constraints; and architectural integration. Occupant needs include the wider perspective of human health, productivity, interpersonal communication, aesthetic quality, as well as visibility and task performance. These broader considerations have led to studies of human response to lighting and their broader physiological and psychological implications (Veitch et al. 1998). Such studies suggest that many different aspects of lighting quality (IESNA 2000) should be considered in good lighting design including colors and reflectances, brightness and glare, controlled use of daylighting, and local control of artificial and natural light including temporal and spatial variations and contrasts. Quantification and metrics for lighting conditions may need to be extended as experience with lighting perception is developed. In addition, user behaviors on different tasks influence the light reaching the subject, and changing focus, head motion, eye blinks, etc., may result in different exposure.

Previous studies have explored these issues and have attempted to quantify the perceptions and effects of luminance distributions on productivity and health of occupants (Cetegen et al. 2008; Newsham et al. 2004; Mahdavi and Eissa 2002; Veitch and Newsham 1998). From these studies, it is clear that a wide variety of factors affect the preferred choice of lighting conditions, and an adaptive “smart” approach to the synthesis of lighting conditions may offer opportunities to markedly improve the effectiveness of lighting and its role in the productivity and health of occupants. While the prior work is based on studies of fixed lighting conditions, the use of adaptive distributed control will require several additional capabilities:

1. Distributed light sources with adjustable (controllable) parameters, including intensity, spectral distribution, spatial pattern, and others.
2. Adaptive smart lighting control methods will need sensory feedback indicating the actual lighting conditions in the space and algorithms for modification of distributed sources to achieve alternative spatial displays.
3. Sensing of room occupancy and task use will be important to the adaptation introduced. As discussed above, these considerations might also be individualized and depend on the group of users, the time of day, and the characterization of the tasks involved.
4. Sensors and light sources will be networked to share information and compute the control functions required.

The opportunities for distributed smart lighting systems control also encompass the goals of energy efficiency and environmental impact. Recent studies (Selkowitz 2008; Griffith et al. 2006) have attempted to assess the needs of building design to achieve improved energy and environmental performance. These studies emphasize the broad needs of cooperative economic, social, and political initiatives, in addition to technological advances. Within the technology sector, it is also clear that a newly educated workforce is an important component of these initiatives. More familiarity with new technologies will be required to

clarify performance, reduce complexity of design, and integrate new technologies with practice and regulation.

Distributed Lighting Systems

Recent advances in lighting sources, sensors, and systems have opened up opportunities for novel approaches to lighting systems that integrate component technologies in a distributed architecture. Development of integrated adaptive intelligent systems for control of energy utilization, including lighting, will reduce the barriers to custom design, construction, and maintenance for these new technologies. Adaptive smart control of distributed lighting systems will be one important component of such systems. The education of new generations of engineers and technical support workers with skills to design and implement smart energy management systems will be needed to accomplish these goals.

In recent data from the Lawrence Berkeley National Laboratory (Selkowitz 2008), Selkowitz indicates that in 2001, 39 % of total energy and 71 % of US electricity have been used in building energy use. In addition, building energy use is dominated by lighting, heating, and cooling, with lighting (28 %) the largest use of energy in commercial buildings. He indicates that major savings of lighting energy could be achieved by the application of currently available lighting control strategies including vacancy detection, dimming with daylight, demand response, and personal controls. He cites studies that show savings of 40–60 % through the use of advanced lighting controls. He also notes that the addition of wireless lighting controls to ballasts would enable more flexibility and adaptability for these systems.

Current efforts in distributed lighting systems research extend well beyond the basic control mechanisms currently available and will introduce a new formulation of the problem in terms of multiscale sensor-based control. This approach requires consideration of multisource/multisensor distributed systems and distributed algorithms that address these problems. There has been a limited amount of prior research that addresses this general problem in the context of lighting. A notable patent was issued to Lyons at Philips (Lyons 1998) in 1998. The patent describes a method for optimizing energy efficiency of a multisource lighting system specifically using a linear programming technique to adjust source intensities and satisfy a total energy consumption constraint. The problem is formulated using a set of linear models for energy allocation to each of the sources. It may be solved in several different related formats including minimum and maximum brightness or optimal brightness constraints. As presented in the patent, the approach does not use sensors or feedback in the implementation. It formulates a model-based, not a feedback control, solution to the optimization problem.

A more recent series of publications (Park et al. 2007; Singhvi et al. 2005; Granderson et al. 2004; Wen et al. 2006, 2008; Wen and Agogino 2008) addresses the problem of wireless networked lighting systems used to optimize energy savings. Wen and Agogino (2008) describe a system of lighting sources linked by wireless network technologies with controlled source intensities. The basis of their approach

is an illuminance model generator that predicts the workplane-level illuminance in an office space with known configuration and surface reflectance properties. The workplane-level illuminance is treated as a linear summation of model outcomes. The occupant's preferred light settings are also specified on the workplane surface grid, and the problem is formulated as a linear programming problem. By minimizing the norm of the vector of light output levels subject to the occupant constraints, the overall energy may be minimized. In the case that feasible solutions are not available, the settings are relaxed to form inequality constraints and solved accordingly. In this work (Wen and Agogino 2008), a hardware implementation of the proposed system was described using a wireless link to a basic actuation model for the light dimmers. In this experiment, 12 luminaires were used to provide light to 4–7 occupants. In these experiments, 50–70 % of the light energy was saved using the controlled system. This approach is also model based, not sensor based, and restricted to linear predictive models of planar illumination.

This chapter describes an approach based on the general representation of a light field incorporating spatial and spectral properties of lighting systems, including angular distribution of light rays in the field. This approach integrates sensors into the illumination space, such that the sensors will be networked directly to the sources. The resulting ad hoc network of sources and sensors provides a framework for distributed systems control. A multiresolution approach to the illumination field representation will be introduced to efficiently represent this formulation. This approach is based on previous work (Hombal et al. 2009a, b) on adaptive sampling in distributed sensor networks and has been shown to be efficient and effective for non-homogeneous function models. The optimal configuration of lighting sources is solved by evolutionary algorithms and not limited to linear programming solutions. These algorithms (Zhang and Sanderson 2009a, b) have been shown to perform well in high-dimensional multimodal search spaces. The optimization criteria for these systems may be extended to energy efficiency, environmental constraints, as well as productivity and health goals. These goals and constraints could be expressed in terms of spatial mapping properties and extended to the evolutionary optimization process.

Distributed Sensing of the Light Field

While the smart lighting system aims to produce the right light, a desired light field is usually designed to satisfy the light field requirements in a specific lighting application. This desired light field may be considered the target of lighting control, and thus the generated light field will be a set of fields similar to the target one. In this case, it is reasonable to deploy sensors based on the target field, and the objective is to generate samples to maximize the information obtained from the deployed sensor array.

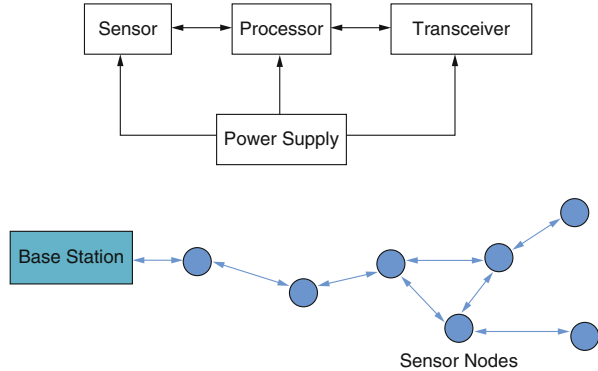
In previous research, approaches have been proposed for light field sampling. Levoy and Hanrahan (1996), while introducing the light field for image based rendering in computer graphics, proposed to use a camera to describe the phenomena

around scenes without creating a 3D geometry model of these scenes and to compute views of scenes without knowing their surface properties. Since the radiance does not change along a line unless blocked, the light field of interest may be considered as a 4D function, rather than a 5D function, between two planes. By capturing images from multiple perspectives of a single translating camera, the authors successfully represented the 4D light field. Wilbrun (Wilburn 2005) extended this idea to high-performance imaging by using large video camera arrays. The authors used different configurations of camera arrays for different purposes. Moreover, a newly released light field camera invented by Ng (2006) is also inspired by this idea. The light field camera captures the entire light field traveling in every direction in every point on the camera plane, which permits refocusing of pictures at any time after taking the pictures. The author reconstructed the light field inside the camera body by inserting an array of microlenses in front of the photosensor. Each microlens covers multiple photosensor pixels and separates the light rays from different directions. Using this information throughout the entire light field, the refocused image can be computed and displayed.

These approaches enable reconstruction of the 4D light field between two planes in free space without obstacles. In photographs taken by a single camera, the amount of light traveling along individual rays is not recorded. Instead, only the sum total of light rays striking each point in the image is recorded. In general illumination applications, the entire light field cannot be obtained by a limited number of cameras. Thus, the use of localized sensors (“point” sensors with known spectral response) instead of cameras has two advantages. First, at a specified position and viewing angle, a camera cannot provide more information about the light field than a color sensor does. In this sense, if a camera serves the same purpose as a localized sensor, a single image from a camera under a single exposure is not sufficient to provide exact light field information. A high-performance camera using multiple exposures and the synthesis of a high-dynamic range image is required. Moreover, it would be more expensive and inconvenient to deploy cameras than localized sensors. The second advantage is that additional computation is needed to obtain the light field data, since we do not need the high resolution of the camera image.

Distributed sensor networks (DSN) integrate advanced sensor technology with local communications and computing and support a growing range of applications such as observation, monitoring, and tracking of complex distributed processes (Curtin and Bellingham 2001; Estrin et al. 2001; Sanderson et al. 2006). Distributed sensor networks were originally applied in the military area such as for battlefield surveillance (He et al. 2004). However, in recent years, because of its importance in acquiring and processing information, this technology has been extended to other areas, such as industrial and civilian applications. One such scenario involves the deployment of multiple underwater vehicles such as AUVs (Curtin and Bellingham 2001) as a part of a distributed sensor network for ecological and environmental monitoring of large bodies of water, such as oceans, harbors, lakes, rivers, and estuaries. Autonomous underwater vehicles (AUVs) have been incorporated in distributed sensor networks enabling pervasive in situ observation of such processes in a wide range of spatial and temporal sampling resolutions (Sanderson et al. 2006).

Fig. 1 Components of a sensor node and the structure of a sensor network



Although the components in a sensor node network may differ in different applications, the network consists of four basic parts, sensor unit, processing unit, data transfer unit, and power supply, as shown in Fig. 1. With these components, each sensor node will perform three main tasks: sensing, computing, and communicating. These sensors are then networked together to share the information and are connected to a base station. To access the information in the sensor network, queries can be generated from the base station and routed to regions of interest with a cluster of sensors in the network. These related sensors will process data collaboratively and aggregate information locally within the cluster to reduce communications.

Because the sensor nodes in the network are embedded devices, they carry limited hardware resources including limited memory storage and battery power. There are also resource constraints on the sensor network as a whole, such as limited communication bandwidth. Therefore, efficiently utilization of resources to achieve given sensing tasks is of great importance and typically a major objective in sensor network design. Figure 1 shows the components of and a sensor node and the structure of a sensor network.

Given a set of sensor nodes, it is important to find an optimal deployment of these nodes to maximize the sensing accuracy while minimizing resource consumption. Much research has focused on the development of wireless sensor networks with local communications capability (Popa et al. 2004; Batalin and Sukhatme 2004; Isler et al. 2004). Communications networks are expected to be autonomous and ad hoc such that the node can enter and leave the network freely. Thus, when deploying sensor nodes, a set of communications and topology control problems need to be considered to maintain the network connectivity (Dharne and Jayasuriya 2006). Certain self-deployment methods such as the virtual force algorithm (Zou and Chakrabarty 2004) and potential field algorithm (Howard et al. 2002) have been developed to measure, repair, and complete the network connectivity. Furthermore, since the phenomena are usually not uniformly distributed in the environment, the density of the sensor nodes is not fixed and should adapt to this distribution. In an extremely dynamic environment, the locations of the sensors will also need to be adaptively adjusted to track the dynamic process. In these cases, dynamic and adaptive deployment of sensing resources is required to achieve the sensing task.

Vieira et al. (2004) proposed an efficient incremental deployment algorithm which uses the information of the current node density, energy level, and sensing coverage to guide the new sensor deployment. Willett et al. (2004) proposed an adaptive preview refinement approach for sensor deployment. In the preview step, by using a sparse set of sensors, an initial environment estimate is formed. This initial estimate is then utilized for determining locations of additional sensors in the refinement step. While maintaining high accuracy, this adaptive two-step approach can significantly reduce the energy and communication consumption.

Adaptive Sampling

Sampling is a broad methodology for gathering statistics about physical phenomena. It is a fundamental area of scientific activity, which generates empirical evidence for scientific models. In general, sampling refers to the methodology of selecting a finite subset from a larger set representing the total population toward an assumed scientific objective. In signal processing, such sampling is employed to select a subset of bases from a family of bases toward an efficient representation of the signal (Meyer 2001).

There are some standard sampling methods such as uniform sampling and random sampling. Uniform sampling is a classical sampling strategy, which acquires measurement at uniform spatiotemporal intervals. Two design variables are the length of transects and the separation between them. In general, there exists a tradeoff between the sampling coverage and the time taken to sample a given area which represents the efficiency of spatial monitoring. In general, it is straightforward to design a uniform sampling algorithm and easy to implement it in practice. However, due to the uniform sampling resolution, this algorithm has low sampling efficiency when the observation space (environment) has uneven features. Dense samples should be placed in regions with important features in order to have improved sampling resolution. In addition, since the sampling density is uniform everywhere, regions lacking features are probably oversampled, which results in a waste of sampling resources. On the other hand, when sampling a periodic process, there will be a potential error because of lack of randomization in the sampling algorithm. These disadvantages limit the applicability of the uniform sampling algorithm.

It is important to find adaptive sampling algorithms that can intelligently adapt to environment features and improve the estimation under resource constraints. Starting with a coarse sampling distribution, adaptive sampling algorithms incorporate environment knowledge from previous samples and utilize it to refine the sampling distribution. In contrast to uniform sampling, good adaptive sampling design should be able to allocate different sampling resolutions to regions with different features. Figure 2 shows the comparison of a uniform sampling deployment and a variation-sensitive adaptive sampling solution, in which the sample density increases in regions with high variation and decreases in regions with low variation (Hombal 2009).

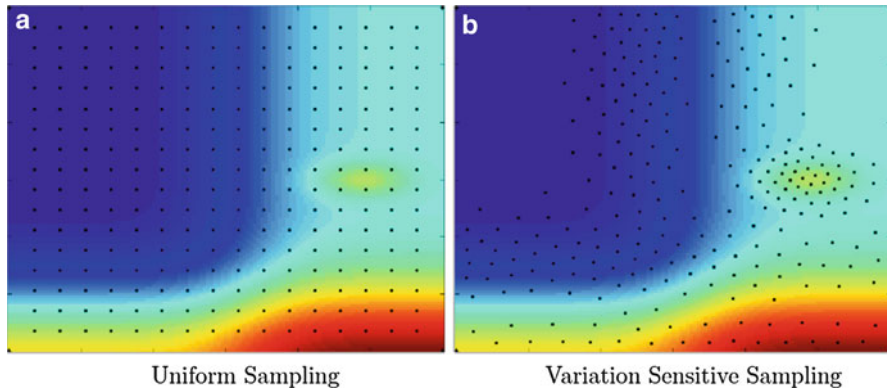


Fig. 2 Coverage and resolution of sample distributions (Hombal 2009). (a) Uniform sampling and (b) variation-sensitive sampling

The features in a practical environment are often highly nonlinear and dynamic, and there are high spatial and temporal variations in the distributions of the variables. Traditional sampling algorithms such as uniform sampling and random sampling have poor performance in these circumstances, requiring extensive sample resources to guarantee an adequate sampling resolution. For these complex environments with nonuniform features, an efficient sampling design should incorporate incremental knowledge of the environment through previous measurement to guide an adaptive sampling strategy.

Adaptive sampling refers to iterative feedback sampling algorithms in which an estimate of the underlying function is constructed from previous measurements. Usually this estimate, along with the feedback from the real environment measurements, guides the deployments of the sampling resources in the next iteration. In this approach, it is important to choose an appropriate model to estimate the underlying function based on initial samples. For complex underlying functions in the observation space, usually it is very difficult to find exact models, and thus surrogate models are considered to achieve the approximation. Surrogate models are approximation models which can approximate the characteristics of the underlying phenomena. Some popular surrogate models include response surfaces (Box and Draper 1986), support vector machines (Mangasarian 2003), artificial neural networks (Smith 1993), and radial basis functions (Duchon 1977; Unser 2000). Constructed using a data-driven approach, surrogate models are computationally feasible and convenient to implement and are widely used in data representation, optimization, and sensitivity analysis.

In adaptive sampling, it is important to choose an appropriate field model to integrate initial samples into an estimation of the underlying function. For complex underlying functions, usually it is very difficult to find exact models, and thus surrogate models are considered to achieve the approximation. Hombal (2009) introduced a multiscale surrogate model for sampling and estimation of the unknown underlying process based on localized radial basis functions. To be consistent with

this general sampling regime, Dong (2012) employs hierarchical radial basis functions (HRBF) to implement coarse-to-fine modeling of the underlying light field. The HRBF network may be viewed as a neural model for multiscale approximation of a function through multilayer decomposition of the approximation error space. Each layer of the model is approximated by a radial basis function (RBF) network with a different scale. The structural parameters in the HRBF model can be determined by a hierarchical analysis grid constructed in the problem domain.

The structural parameters of HRBF are set according to an ordering imposed by a hierarchical analysis grid defined on the problem domain. The analysis grid is such that each layer of the grid is a dyadic partition of the previous layer. The intersections of such partitions form the nodes of the analysis grid. Each layer of the analysis grid corresponds to a layer in the HRBF. The number and position of nodes at the corresponding layer in the grid determine the number of basis functions and the locations of the centers in each HRBF layer. Further, the scale parameter is set according to the density of the nodes. Figure 3 shows an example of the 1D analysis grid.

Multiscale adaptive sampling algorithm (MSAS) (Homabl 2009) is an adaptive sampling approach which achieves variation-sensitive sampling and generates a multiscale functional representation of the underlying process using localized basis functions. Starting with a sparse sample distribution, the MSAS algorithm constructs an estimate of the underlying process from existing measurements and utilizes it to guide the selection of subsequent samples for refinement. The implementation of MSAS can be facilitated by the integration of sensors on mobile robots, enabling adaptive and dynamic redeployment of sensors for new sensing tasks. This chapter provides a brief survey of MSAS as well as new results on experimental evaluation of its performance.

The multiscale adaptive sampling (MSAS) algorithm generates a variation-sensitive sample distribution and is capable of obtaining accurate multiresolution representations of the underlying functions with low sampling costs. In practice, MSAS is especially useful in guiding sensor deployment strategy for exploration of

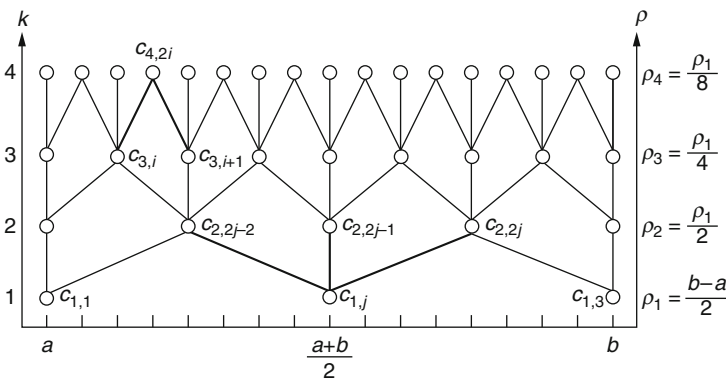


Fig. 3 The 1D analysis grid

unknown regions. However, the basic MSAS algorithm is only valid for sampling stationary underlying phenomena. Samples in MSAS are iteratively generated, and planning of sensor locations to visit the sample points is suboptimized in each iteration, but not optimized globally. In practice, subject to constraints on sensor availability, it may take a long time to deploy sensors on generated sample points for measurements. In a dynamic environment, MSAS has to globally redeploy the sample resources in response to underlying phenomenon changes, which may be time inefficient and cannot guarantee the imposed temporal sampling resolution subject to the process constraint. Since many environments are highly dynamic, this property limits the practical applications of the basic MSAS algorithm.

The adaptive sampling approach guides the systematic selection of sensors to sample the generated light field and fusion of sensor information for lighting control. This adaptive sampling and sensor fusion approach could significantly reduce the error in representation of the light field. A series of experiment results also demonstrate that the adaptive lighting control system with this adaptive light field sampling regime can consistently reduce the control error by more than 65 % relative to that with uniform sampling. However, this adaptive sampling approach may be only valid for a pristine (no disturbances) environment. Since the sample locations are predetermined and fixed based on the target field, the deployed sensors may not be able to fully capture the light field disturbance due to user activities or the presence of natural light. Among various disturbances in a smart room, daylight is a major one and the understanding of its characteristics is of great importance for the development of an energy-efficient lighting system. For an environment subject to dynamic daylight disturbance, real-time monitoring requires incorporation technologies to select or redefine the sensor network and real-time reallocation of sensor locations to adaptively sample the dynamic light field.

Dong (2012; Dong and Sanderson 2013) proposed a dynamic multiscale adaptive sampling (DMSAS) approach based on MSAS for unknown dynamic. In DMSAS, sensors deployed at the last time step first take samples of the new underlying function and employ the sensing data with corresponding RBFs to construct a coarse multiscale estimation. This estimation can be viewed as a measurement of the new underlying process and is corrupted with sampling noise. With this measurement and prior knowledge of the process evolution model, a Kalman filter may be designed to derive a refined estimate of the new underlying function and determine the new position of sensors for multiresolution sampling. In this way, the route of currently deployed sensors to their destinations can be optimally planned with an objective to minimize the traveling time. Different from globally redeploying all the sensors in MSAS, DMSAS involves locally moving related sensors to capture local features of the underlying process. The relocation of sensors in DMSAS is much more time efficient and consistent with dynamic sampling. The resulting approach maintains a linear evolving model of the dynamic daylight and recursively identifies the model parameters based on previous daylight estimation. To sample a new daylight field, a model-based prediction is first achieved through Kalman filtering, and new sample locations are generated based on this prediction. The mobile sensor can then be repositioned to take all the samples and refine the daylight estimation.

These procedures are repeated to plan and manage mobile sensors to track and estimate the dynamic daylight field.

Light Field Estimation and Interpolation

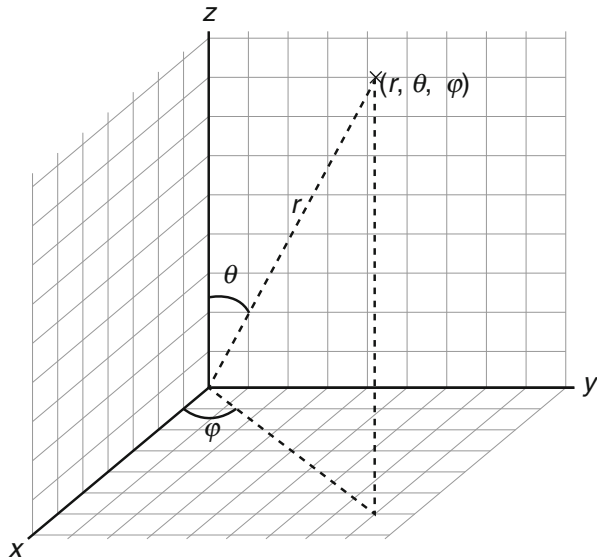
Currently, in general illumination applications, the measurement of light distributions and control of light sources are typically conducted with respect to a two-dimensional reference plane in the illumination field. Light sensors are placed either on a plane or pointing at the same angle, and much of the light distribution information from different viewing angles is neglected. The results of light source analysis, design, or control will no longer be valid if the viewing position or the viewing angle is changed. Therefore, it is necessary to take into consideration light distributions from all directions for better design and control of light sources to meet diverse target light field requirements. In order to address these design and control issues in the more general case, a consistent representation and set of analysis tools for the five-dimensional field are needed.

For analysis and design purposes, it is desirable to reconstruct (interpolate) the light field based on a set of discrete sample points of the simulated data or experimental measurements. In this case, the sampled data occur in the five-dimensional space, three dimensions in spatial position and two dimensions in direction. The sampled data may include the property of anisotropy. That is, there may be different models for estimation in different directions. Adaptive sampling of the light field (Dong 2012) and adaptive control of color-tunable lighting systems (Afshari et al. 2012) are examples where such an interpolated model may be required.

Most previous research on light field acquisition has been applied to the problem of image rendering in computer graphics (Foley et al. 1995; Hill 2001). Rendering may be viewed as a reconstruction of the light field forming an estimate of the perceived image around the objects (Goral et al. 1984; Phong 1975). However, these techniques are not suitable for the study of the general illumination applications considered here. First, the goal in image rendering is to achieve realistic images when viewed from a specified viewing angle. The goal in general illumination applications is to achieve a full description of the light field in space rather than the light field viewed from a specified angle. Second, in applications to image rendering, there is a requirement for a high-resolution representation of the image light field. In the case of general illumination, the required resolution of the desired light field is usually greatly reduced. The goals of general illumination, such as ambient, task, and accent lighting, require a lower spatial resolution but span a larger spatial volume and range of angles.

For image synthesis in computer graphics, Greger et al. (1998) proposed the idea of irradiance volume to calculate the global illumination. They sampled the radiance at sample points and directions and computed the irradiance distribution function to build the irradiance volume (Greger 1996; Greger et al. 1998). However, the query and estimation of the irradiance volume is completed by trilinear interpolation, without considering the anisotropic characteristic of light transport. Huang (2013)

Fig. 4 Spherical coordinates with two angle variables



proposed a spatial estimation techniques based on Kriging techniques to address the problem of anisotropy.

Kriging (Matheron 1963) is a technique developed for spatial interpolation of sampled data and has often been applied to fields of environmental science (Bayraktar and Turalioglu 2005), hydrogeology (Chiles and Delfiner 1999), and mining (Richmond 2003). Kriging techniques estimate the value of an unknown function expressed by a combination of two components: (1) a deterministic component related to the estimate of a localized mean and (2) a stochastic component which is dependent on the covariance of the function.

A five-dimensional function of the light field, which is a subset of the seven-dimensional plenoptic function of the light field widely accepted in computer graphics (Gershun 1936; Levoy and Hanrahan 1996; Adelson and Bergen 1991), is utilized to describe the light field. The light field can be represented as a five-dimensional plenoptic function in terms of spatial position (x, y, z) and spatial angle (θ, φ) . The sample space is therefore expressed as a five-dimensional vector $x = (x, y, z, \theta, \varphi)$. Figure 4 shows an example of spherical coordinates. The function value $L(x)$ is the plenoptic function of spatial position (x, y, z) viewed at angle (θ, φ) .

For the illumination application, the light field is sampled at a finite set of points, and the plenoptic function is estimated at other locations and angles using the technique of universal Kriging. A second-order polynomial is used for the trend, and there are no cross-terms considered between different independent dimensions. In order to account for the anisotropic properties of light, a three-dimensional anisotropic model based on the spherical semivariogram model is used. The spherical semivariogram is defined by a parameter, and the anisotropic model is defined by additional three parameters: two angle variables and the fraction of anisotropy. Thus, the range of variance changes with direction.

Fig. 5 A 3D view of the five-dimensional light field sampled in a horizontal plane in three-dimensional space with diffuse reflective surface

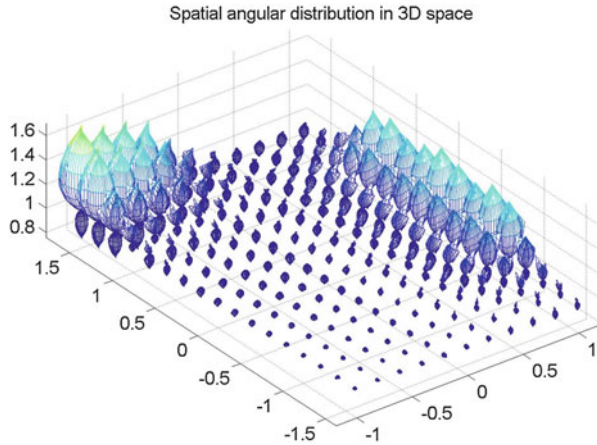


Figure 5 shows a 3D view of the five-dimensional light field distribution in the space along a horizontal plane. Each volume pair (upper and lower) in this figure represents the light field that can be detected at the specified position, which is the point of intersection of the upper and lower volume. Any point on a volume represents the light that can be detected from a viewing angle, which is determined by the direction from the point of intersection to the point on the volume, and the distance between represents the value of light measurement such as illuminance.

Optimal Source Configuration

Optimal light source configurations are needed to meet the requirements of light performance and energy consumption. Figure 6 shows a schematic example of light source configurations and the target light field. The search process can be formulated as an optimization problem based on analytical models of light sources, light propagation characteristics, and the user requirements.

To formulate this optimization problem, the mathematical definition of the target requirements is important. In previous illumination applications, the target is represented by a one-dimensional array or a two-dimensional matrix of illumination values which can be detected based on two-dimensional sensor sampling. However, this is not sufficient to achieve the goals of smart lighting, because there is a lack of spatial information including the directional characteristics of light sources. A five-dimensional function of the light field is utilized to describe the target. The five dimensions include the appropriate values for spatial position (x , y , z) of the specified point and the azimuth angle and zenith angle (θ , φ) of the specified direction. The light field at any position at any viewing angle is expressed as $L(x, y, z, \theta, \varphi)$, which can express the target accurately. Figure 7 shows a schematic drawing of five dimensions of a light field detected at a single location in space.

Fig. 6 Schematic example of light source and target light field

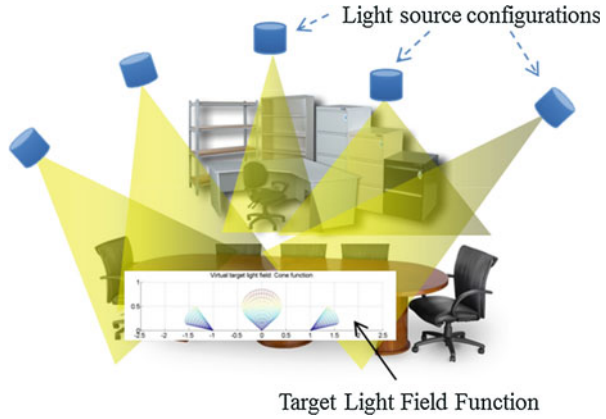
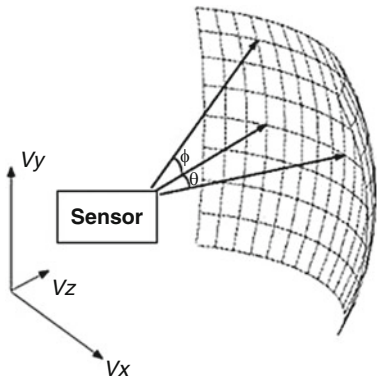


Fig. 7 Schematic drawing of five dimensions of a light field



Since the angular information is included in the target function, the angular information in the light source configurations is part of the corresponding target light field. The light source is characterized by the following parameters: the spatial position (x, y, z) , which can also be expressed by the spherical coordinate (R, θ_1, ϕ_1) , the directional angle (θ_d, ϕ_d) , the power (intensity) of the light source I , and the beam angle of the light source ω . These seven parameters constitute the basic seven-dimensional search space of a single light source. The estimation of the five-dimensional light field in the computation is accomplished by spatial sampling and spatial estimation using the Kriging technique described above.

Consider the case of a desired light field $L_d(\mathbf{q})$ at each spatial position \mathbf{q} with N light sources and the light field generated from the i th light sources $L_i(\mathbf{q})$. The optimization problem can be expressed as

$$\min \left(\sum_{\mathbf{q}} \left\| L_d(\mathbf{q}) - \sum_{i=1}^N L_i(\mathbf{q}) \right\| \right)$$

According to the light transport characteristics, the result of multiple light sources is the superposition of the fields generated by all the light sources. On one hand, for each light source, there is a seven-dimensional search space. The dimension increases dramatically with the increase in the number of light sources. On the other hand, there may be more than one satisfactory configuration to the problem. There is therefore a need to optimize with respect to problems of high dimension with highly multimodal objective functions.

The Speciated Parameter Adaptive Differential Evolution (SPADE) algorithm is one of the optimization algorithms developed for this purpose (Huang 2013). The SPADE algorithm is a novel multiple population DE algorithm proposed to identify different species which occur as distinct subpopulations of the evolutionary process. In this approach, the population is partitioned using an unsupervised clustering method, and the subpopulations are able to track alternative solutions associated with local minima of the objective function. These species evolve separately based on parameter adaptive differential evolution with occasional crossover interactions across subpopulations. Therefore, each species can be tracked separately by the corresponding population to achieve multimodal optimization.

In the optimization of light source configurations, the SPADE algorithm utilizes a five-dimensional light field estimation. A model of the light source and the light sensor response according to the light source is predefined by an eight-dimensional matrix. This eight-dimensional matrix is generated in simulation by importing luminous intensity models of the light source and applying light sensors. The model variables include the distance between the light source and the light sensor, the spread cone angle of the light source, the two spatial angles of light sensors with respect to the light source, the two orientation angles of the light source, and the two orientation angles of the light sensor.

Compared to general ray tracing methods, this eight-dimensional matrix model-based approach saves a lot of computation time. However, this model-based approach has the problem of sensitivity to model variation and model errors. First, the optimization procedure is sensitive to the luminous intensity model of the light source, which defines the entire the light distribution of the light sources. In general, there are different types of light sources, such as point light sources, spot light sources, and area light sources. If the luminous intensity model is very different from what is used in practice, there would be a large error caused to the result of optimization. In this thesis, an approximation of the luminous intensity model of the light source is provided by the designer. Second, among the model parameters that can be adjusted in the procedure, the sensitivity to the parameter of source spread angle is high, while the sensitivity to the distance between the light source and the light sensor is low. For better performance of the optimization, it may also be worthwhile to take the optimization of the luminous intensity model of the light source into the procedure in future work. In addition, a more extensive quantitative evaluation of model sensitivity and its impact on optimization may be carried out.

For the optimization of light source configurations, the SPADE algorithm offers substantial advantages. The computational time is affected by the dimension of the

search space and the number of species generated. The dimension of the search space makes a larger impact on the computational time than the number of species generated. In illumination applications, additional computation is required for the light sensor response caused by the light source. In order to save computational time, an approximation of the model was used to integrate the ray tracing program into the optimization procedure. However, the model is still of high dimension, and there are additional requirements for the computation of applying the model to generate the results and provide a comparison with the five-dimensional target light field. This additional computation cost is relatively small compared with the effect of additional dimensions added in the search space.

Figure 8 shows an example of a light source configuration optimization. In this example, the target light field is obtained with the condition of three predefined light sources, as shown in the figure on the right. The search space is a 15-dimensional space, including the distance away from the center, the angular position in the spherical coordinate, and the orientation angle of all three light sources. The optimization results are determined with the average error less than 25 %. The figure on the left shows the light field distribution after optimization. Similarly, each volume represents the light field that is detected at the specified position.

Adaptive Control Using Sensor Feedback

The multiscale functional approximation of the light field can be achieved by fusing multiple sensor measurements, and it can be utilized in a feedback mode for lighting control. Dong and Sanderson (2013) incorporated the sampling methodology and adaptive lighting control in a smart space testbed, as shown in Fig. 9. The lighting system consists of multispectral LED modules as light sources and a set of color sensors used for light field sampling and estimation.

Effective feedback control in a lighting system requires knowledge of the light propagation under given conditions. For an LED lighting system, (Afshari et al. 2012) introduced a linear model to characterize the light propagation from the system input to the generated light field on discrete sensors. A similar model can be constructed for mapping the LED input to the generated light field on the discrete domain.

$$f_{\text{RGB}} = \mathbf{G}u_{\text{RGB}} + \omega$$

where f_{RGB} is the generated light field represented in RGB color space, u_{RGB} is the RGB LED lighting input, \mathbf{G} is the light transport matrix (LTM), and ω denotes light disturbances in the system. The LTM depends on the room configuration and can be identified using a least square approach.

In this lighting system, light field sampling is conducted by a distributed sensor network. Based on the sampling tasks, the sensors are divided into two subsets. In one of the subsets, the sensors are set static and adaptively deployed based on the

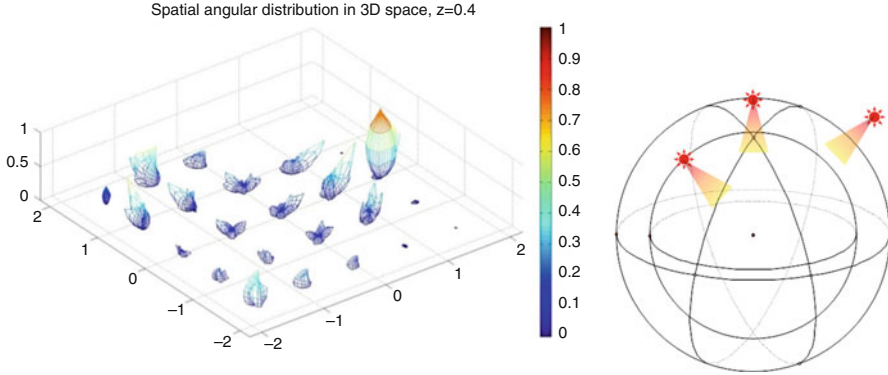
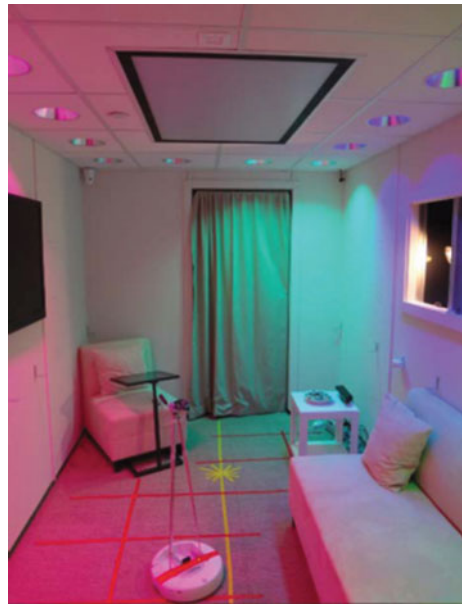


Fig. 8 An example of the light source configuration optimization

Fig. 9 Smart space testbed for lighting control



target light field. The main task is to sample and capture features of generated light field. The other subset of sensors is mobile, which are routed and repositioned with time by the proposed dynamic adaptive sampling approach to track and sample the dynamic disturbance. The actual light field is measured by all sensors, and a functional light field approximation can be derived by an HRBF interpolation on the sensor measurements. As a fusion of all the sensor measurements, this light field approximation should be capable of capturing features of the actual light field, which is a combination of the generated field by light sources and the dynamic disturbance.

Once the approximation of the light field is constructed, it can be used in feedback mode for adaptive lighting control. The objective of the control is to minimize the perceptual difference between the estimated light field and target light field.

$$\begin{aligned} \min J(\mathbf{u}) &= \sum \|\mathbf{f}_{\text{Lab}}^{\text{tar}}(\mathbf{x}) - \hat{\mathbf{f}}_{\text{Lab}}(\mathbf{x})\|_2^2 \\ \text{st. } \mathbf{f}_{\text{RGB}} &= \mathbf{G}\mathbf{u} + \mathbf{d}_{\text{RGB}} + \boldsymbol{\omega} \\ \hat{\mathbf{f}}_{\text{Lab}} &= \mathbf{h}(\mathbf{f}_{\text{RGB}}) \end{aligned}$$

where $J(\mathbf{u})$ is the cost function, $\mathbf{f}_{\text{Lab}}^{\text{tar}}(\mathbf{x})$ is the target light field represented in CIELAB color space, $\hat{\mathbf{f}}_{\text{Lab}}(\mathbf{x})$ is the estimation of target light field represented in CIELAB color space, \mathbf{f}_{RGB} is the generated light field represented in RGB color space, \mathbf{u} is the RGB LED lighting input, \mathbf{G} is the light transport matrix (LTM), \mathbf{d}_{RGB} represents the daylight in RGB color space, $\boldsymbol{\omega}$ denotes light disturbances in the system, and $\mathbf{h}(\cdot)$ is a nonlinear mapping from the RGB color space to the CIELAB color space.

With this formulation, a gradient-based control method is developed to iteratively obtain the optimal LED input to iteratively obtain the optimal LED input to generate the target light field. The control input at the i^{th} time step is calculated as

$$\mathbf{u}_{i+1} = \mathbf{u}_i + \epsilon \cdot \begin{bmatrix} \nabla_{\mathbf{u}}(\mathbf{L}_{\text{tar}} - \mathbf{L}(\mathbf{u})) \\ \nabla_{\mathbf{u}}(\mathbf{a}_{\text{tar}} - \mathbf{a}(\mathbf{u})) \\ \nabla_{\mathbf{u}}(\mathbf{b}_{\text{tar}} - \mathbf{b}(\mathbf{u})) \end{bmatrix}_i^+ \begin{bmatrix} (\mathbf{L}_{\text{tar}} - \hat{\mathbf{L}}) \\ (\mathbf{a}_{\text{tar}} - \hat{\mathbf{a}}) \\ (\mathbf{b}_{\text{tar}} - \hat{\mathbf{b}}) \end{bmatrix}_i$$

where \mathbf{u}_i is the i^{th} time-step input, ϵ is the tunable step size, $[\cdot]^+$ is the pseudo-inverse matrix, and $\nabla_{\mathbf{u}}(\cdot)$ is the gradient with respect to the input.

An experiment has been implemented to evaluate the dynamic adaptive sampling approach in lighting control with dynamic daylight disturbance. The light field is observed on a 2D horizontal plane in the smart space, and a 2D target light field is specified on the observation plane. Initially the lighting control system is tested without presence of daylight, and a distribution of 12 samples is generated by adaptive sampling based on the target field. The light field estimation is utilized in feedback mode by the control system, which adaptively tunes the light sources to reproduce the target field. Figure 10a shows the distribution of the steady-state lighting control error over the observation plane, which is defined as the Euclidean distance between the target and actual light field in the CIELAB color space. The lighting control error is very small in the region of observation, which implies that with accurate feedback of the light fields, the lighting system is capable of reproducing the target field with minimal perceptual distortion. For comparison, Fig. 10b shows the distribution of the lighting control error with uniform light field sampling. In this case, the lighting system fails to reproduce the target field in the region due to the lack of sensors and incomplete feedback of the generated light field. The mean control error on the problem domain is five times higher than that in a system with adaptive sampling.

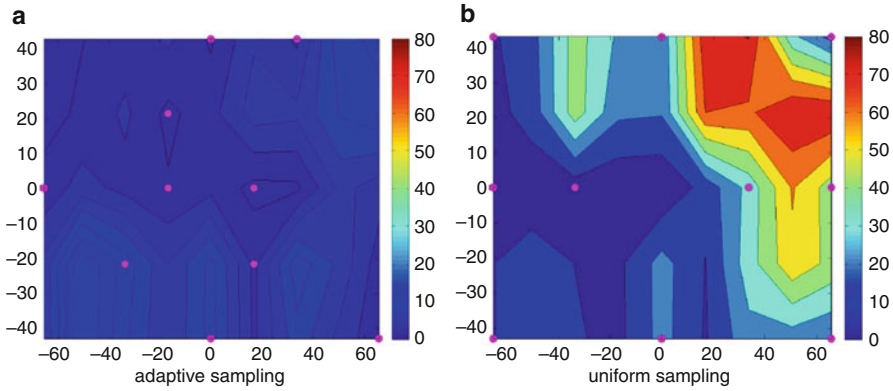
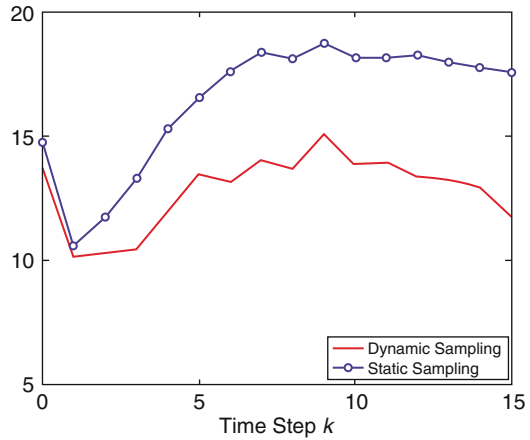


Fig. 10 Control error comparison between lighting systems with adaptive and uniform sampling methods. (a) Adaptive sampling and (b) uniform sampling

Fig. 11 An example of light control comparison between dynamic and static methods



At the following time steps in the experiment, the lighting system is tested with dynamic daylight disturbance. At each time step, eight samples are generated based on the target field by adaptive sampling, and their locations are fixed during the experiment. Another four new samples are generated by the dynamic adaptive sampling approach to track the daylight field. Based on the measured light field, the control system aims to compensate the daylight disturbance and maintain the actual light field as the target one. For comparison, a parallel experiment is conducted to test the lighting system with 12 static sensor samples, whose locations are predetermined to sample the target light field by the adaptive sampling approach. Figure 11 shows the mean control errors of the lighting control systems with dynamic and static sampling versus time. The lighting system with dynamic approach is capable of tracking the dynamic circumstances and reproducing the target light field with little perception distortions. In comparison, the lighting system

with static approach fails to adequately reproduce the target light field due to the lack of sensors and insufficient feedback of dynamic information from surroundings. The lighting system with dynamic adaptive light fielding sampling reduces the control error by 20 % relative to that with static sampling.

Future Directions for Smart Lighting Systems

The rapid development of new technologies for lighting sources, sensors, network infrastructure, and architecture underlies a broad exploration of novel systems concepts and new models for lighting design and deployment. The principles of adaptive distributed sensing described in this chapter will be central to these new approaches and will lead to opportunities for adaptive and intelligent environments. Such environments will require integrated sensing capability without intrusion into the living space. In addition, lighting sources are expected to continue to become more diverse and innovative. Sources that provide more adaptive access to spectral and spatial characteristics will be important. Flexible networking of sources and sensors will provide the backbone of infrastructure to support adaptive changes responding to use, occupancy, health, and task-related needs.

Acknowledgments This research was supported in part by the Engineering Research Centers Program of the National Science Foundation under NSF Cooperative Agreement No. EEC-0812056 and in part by New York State under NYSTAR contract C090145.

References

- Adelson E, Bergen J (1991) The plenoptic function and the elements of early vision. In: Landy M, Movshon JA (eds) *Computation models of visual processing*. MIT Press, Cambridge
- Afshari S, Mishra S, Julius A, Lizarralde F, Wen JT (2012) Modeling and feedback control of color-tunable LED lighting systems. Paper presented at 2012 American control conference, Montreal
- Batalin MA, Sukhatme GS (2004) Coverage exploration and deployment by a mobile robot and communication network. *Telecommun Syst J* 26(5):181–196
- Bayraktar H, Turalioglu FS (2005) A Kriging-based approach for locating a sampling site – in the assessment of air quality. *SERRA* 19(4):301–305
- Box GE, Draper NR (1986) *Empirical model-building and response surface*. Wiley, New York
- Cetegen D, Veitch JA, Newsham GR (2008) View size and office luminance effects on employee satisfaction. *Proceedings of Balkan Light 2008*, Ljubljana, 7–10 Oct 2008, pp 243–252
- Chiles JP, Delfiner P (1999) *Geostatistics, modeling spatial uncertainty*. Wiley Interscience, New York
- Curtin TB, Bellingham JG (2001) Guest editorial-autonomous ocean-sampling networks. *IEEE J Ocean Eng* 26(4):421–423
- Dhame AG, Jayasuriya S (2006) A new protocol for the development and maintenance of autonomous mobile sensor networks. Paper presented at American control conference 2006, pp 3494–3499
- Dong F (2012) *Dynamic adaptive sampling for environmental and smart lighting distributed sensor applications*. PhD dissertation, Rensselaer Polytechnic Institute
- Dong F, Sanderson AC (2013) A dynamic adaptive light field and daylight sampling approach for smart lighting control. *Light Res Technol*. doi:[10.1177/1477153513502030](https://doi.org/10.1177/1477153513502030)

- Duchon J (1977) Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In: Schempp W, Zeller K (eds) *Constructive theory of functions of several variables*. Oberwolfach, 1976. Springer, Berlin, pp 85–100
- Estrin D, Borriello G, Colwell R, Fiddler J, Horowitz M, Kaiser W, Leveson N, Liskov B, Lucas P, Maher D (2001) *Embedded, everywhere: a research agenda for networked systems of embedded computers*. Computer Science and Telecommunications Board (CSTB) Report
- Foley JD, Dam AV, Feiner SK, Hughes JF (1995) *Computer graphics: principle and practice*. Addison Wesley, Reading
- Fung G, Mangasarian OL (2004) A feature selection newton method for support vector machine classification. *J Comput Optim Appl* 28(2)
- Gershun A (1936) *The light field*, Moscow (trans: Moon P, Timoshenko G). *Math Phys*, 1939 18:51–151
- Goral CM, Torrance KE, Greenberg DP, Battaile B (1984) Modeling the interaction of light between diffuse surfaces. *Comput Graph* 18:213–322
- Granderson J, Wen YJ, Agogino AM, and Goebel K (2004) Towards demand-responsive intelligent lighting with wireless sensing and actuation. In: *Proceedings of the IESNA (Illuminating Engineering Society of North America) 2004 annual conference*, Tampa
- Greger G (1996) *The irradiance volume*. Computer science. Cornell University
- Greger G, Shirley P, Hubbard PM, Greenberg DP (1998) *The irradiance volume*. *Comput Graph Appl* 18:32–43
- Griffith B, Torcellini P, Long H, Crawley D, Ryan J (2006) Assessment of the technical potential for achieving zero-energy commercial buildings. In: *Proceedings of the 2006 ACEEE summer study on energy efficiency in buildings*. American Council for an Energy-Efficient Economy, Washington, DC
- He T et al (2004) Energy efficient surveillance system using wireless sensor networks. In: *2nd international conference mobile system, application and services, wide-area monitoring of mobile objects*, pp 270–283
- Hill FS (2001) *Computer graphics*. Prentice Hall, Upper Saddle River
- Hombal VK (2009) *Adaptive sampling in robotic sensor networks for environmental applications*. PhD dissertation, Rensselaer Polytechnic Institute
- Hombal V, Sanderson AC, Blidberg DR (2009a) Adaptive multiscale sampling in robotic sensor networks. In: *Proceedings 2009 IEEE/RSJ conference on intelligent robots and systems*, St. Louis
- Hombal V, Sanderson AC, Blidberg DR (2009b) Adaptive sampling in robotic sensor networks for environmental robotics. In: *Proceedings 2009 conference on unmanned undersea systems and technology*, Lee
- Howard A et al (2002) Mobile sensor network deployment using potential fields: a distributed, scalable solution to the area coverage problem. In: *6th international symposium distributed autonomous robotics system*, Fukuoka
- Huang Z (2013) *Parameter adaptive multimodal optimization and its application in smart lighting*. PhD dissertation, Rensselaer Polytechnic Institute
- Huang Z, Sanderson AC (2014) Light field modeling and interpolation using Kriging techniques. *Light Res Technol* 46(2):219–237
- Illuminating Engineering Society of North America (IESNA) (2000) *Lighting handbook: reference & application*, 9th edn. Illuminating Engineering Society of North America, New York
- Isler V et al (2004) Sampling based sensor-network deployment. In: *Intelligent Robots and Systems, 2004 (IROS 2004)*. *Proceedings. 2004 IEEE/RSJ*
- Levoy M, Hanrahan P (1996) Light field rendering. In: *Proceedings of the Association of Computer Machinery's Special Interest Group on computer graphics and interactive technology (ACM SIGGRAPH)*. ACM Press, pp 31–42
- Lyons D (1998) *Computer software for optimizing energy*. US Patent 5,812,422, Philips Electronics North America Corporation, 22 Sept 1998

- Mahdavi A, Eissa H (2002) Subjective evaluation of architectural lighting via computationally rendered images. *J Illum Eng Soc* 31(2):11–20
- Mangasarian OL (2003) Support vector machine classification via parameterless robust linear programming. Data mining institute technical report
- Matheron G (1963) Principles of geostatistics. *Econ Geol* 58:1246–1266
- Meyer Y (2001) Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth dean Jacqueline B. Lewis memorial lectures. American Mathematical Society
- Newsham GR, Marchand RG, Veitch JA (2004) Preferred surface luminances in offices, by evolution. *J Illum Eng Soc* 33(1):14–29
- Ng R (2006) Digital light field Photography. PhD dissertation, Stanford University
- Park H, Burke J, Srivastava MB (2007) Design and implementation of a wireless sensor network for intelligent light control. In: Proceedings of the 6th international conference on information processing in sensor networks, Cambridge, MA, pp 370–379
- Phong B (1975) Illumination for computer generated pictures. *Commun ACM* 18(6):311–317
- Popa DO, Stephanou HE, Helm C, Sanderson AC (2004) Robotic deployment of sensor networks using potential fields. In: IEEE international conference on robotics and automation
- Richmond A (2003) Financially efficient ore selection incorporating grade uncertainty. *Math Geol* 35(2):195–215
- Sanderson AC, Hombal VK, Fries DP, Broadbent HA, Wilson JA, Bhanushali PI, Ivanov SZ, Luther M, and Meyers S (2006) Distributed environmental sensor network: design and experiments. multisensor fusion and integration for intelligent systems, 2006 I.E. international conference, Sept 2006, pp 79–84
- Selkowitz S (2008) Energy efficiency perspectives: intelligent networks and the challenge of zero energy buildings. Presentation to connected urban development global conference
- Selkowitz SE, Granderson J, Haves P, Mathew PA, Harris JP (2008) Scale matters: an action plan for realizing sector-wide “zero energy” performance goals in commercial buildings. In: Proceedings of the 2008 ACEEE summer study on energy efficiency in buildings, Pacific Grove, pp 17–22
- Singhvi V, Krause A, Guestrin C, James H, Garrett JR, Matthews HS (2005) Intelligent light control using sensor networks. In: Proceedings of the 3rd international conference on embedded networked sensor systems. ACM, San Diego
- Smith M (1993) Neural networks for statistical modeling. Wiley, New York
- Thapan K, Arendt J, Skene DJ (2001) An action spectrum for melatonin suppression: evidence for a novel non-rod, non-cone photoreceptor system in humans. *J Physiol* 535:261–267
- Unser M (2000) Sampling 50 years after Shannon. *Proc IEEE* 88(4):569–587
- Veitch JA, Newsham GR (1998) Lighting quality and energy-efficiency effects on task performance, mood, health, satisfaction and comfort. *JIES* 27:107–129
- Veitch JA, Julian W, Slater AI (1998) A framework for understanding and promoting lighting quality. In: Veitch JA (ed) Proceedings of the first CIE symposium on lighting quality. CIE Central Bureau, Vienna, pp 237–241
- Vieira LF et al (2004) Efficient incremental sensor network deployment algorithm. In: Proceedings Brazilian symposium computer networks, pp 3–14
- Isler V, Kannan S, Daniilidis K (2004) Sampling based sensor-network deployment. In: Intelligent Robots and Systems (IROS 2004). Proceedings. 2004 IEEE/RSJ
- Wen YJ, Agogino AM (2008) Wireless networked lighting systems for optimizing energy savings and user satisfaction. In: Proceedings of 2008 I.E. wireless hive networks conference, Austin
- Wen YJ, Granderson J, Agogino AM (2006) Towards embedded wireless-networked intelligent daylighting systems for commercial buildings. In: Proceedings of the IEEE international conference on sensor networks, ubiquitous, and trustworthy computing (SUTC'06), vol. 1. IEEE Computer Society, Taichung
- Wen YJ, Bonnell J, and Agogino AM (2008) Energy conservation utilizing wireless dimmable lighting control in a shared-space office. In: Proceedings of the 2008 annual conference of the Illuminating Engineering Society, Savannah

- Wilburn B (2005) High-performance imaging using large camera arrays. In: Proceedings of the ACM computer graphics, pp 765–776
- Willett R, Martin A, Nowak R (2004) Backcasting: adaptive sampling. In: Information Processing in Sensor Networks, IPSN 2004. Third international symposium, 26–27 Apr 2004, pp 124–133
- Zhang J, Sanderson AC (2009a) Adaptive differential evolution. A robust approach to multimodal optimization. Springer
- Zhang J, Sanderson AC (2009b) JADE: Adaptive differential evolution with optional external archive. *IEEE Trans Evol Comput* 13(5):945–958
- Zou Y, Chakrabarty K (2004) Sensor deployment and target localization in distributed sensor networks. *ACM Trans Embed Comput Syst (TECS)* 3(1):61–91

Lighting Control Protocols and Standards

Maulin Patel and Satyen Mukherjee

Contents

Introduction	560
Proprietary Versus Standardized Protocols	562
Standardized Wired Protocols	563
0–10 V DC	563
DALI	564
DMX512	565
RDM	565
Architecture for Control Networks (ACN)	566
BACnet	567
LonWorks	569
KNX	570
X10	572
HomePlug	573
G.hn	573
Standardized Wireless Protocols	574
ZigBee	574
6LoWPAN	576
Z-Wave	577
EnOcean	578
Standardized Application Protocol	578
TALQ	578
Summary	579
References	582

Abstract

The technological advances in LED light sources, sensors, and control systems have ushered a new era of smart lighting control systems that provide the right

M. Patel (✉) • S. Mukherjee

Intelligent Enterprises, Current, Powered by GE, San Ramon, CA, USA

e-mail: maulin@gmail.com; satyen.mukherjee@philips.com

quantity and quality of light when and where it is needed in order to enhance the energy savings, maintenance savings, comfort, health, productivity, safety, well-being, and user satisfaction. A networked lighting control system can collect a wide variety of data including presence, ambient light, dimming level, power consumption, user interactions, user preferences, device status, and device failures. These data can be exploited to offer energy management, space optimization, demand response, trend analysis, user comfort maintenance, automatic fault detections and diagnosis, predictive maintenance, and other advanced applications and services.

In order to realize the full potential of smart lighting systems, we need very efficient and scalable networking technologies that can support building-wide, enterprise-wide, and city-wide connectivity. The industry has realized these needs leading to the development of a plethora of networking protocols, both proprietary and standardized. Although the benefits of standardized systems over proprietary technologies are recognized by stakeholders, no standard has emerged as the dominant choice leading to a very fragmented market landscape. In this chapter, we provide a brief overview of standardized networking technologies suitable for lighting controls applications. We also discuss their key features and commercialization prospects.

Introduction

Advanced lighting controls offer one of the most cost-effective means to reduce the energy, carbon footprint, operation, and maintenance costs of buildings and public spaces. Lighting controls exploit various strategies to regulate the timing and intensity of light in order to provide the right amount of light when and where it is needed. In addition to saving energy and maintenance costs, advanced controls can improve user comfort, well-being, productivity, safety, and satisfaction by providing the right quantity and quality of light at the appropriate time and place.

A meta-analysis of energy savings from lighting controls documented in 88 case studies in literature found average energy savings potential of 24 % for occupancy-based control, 28 % for daylight harvesting control, 31 % for personal tuning, and 36 % for institutional tuning Williams et al. (2012). Average savings potential of 38 % was found in case studies using more than one control strategy in commercial buildings. Energy savings in the range of 40–87 % due to schedule, occupancy, and daylight adaptive lighting have also been reported for outdoor lighting used in parking lots, streets, and highways [DOD/ESTCP project EW-201141 final report]. In addition to opportunistic dimming, many networked control systems are also capable of shedding additional lighting load when electricity prices spike or supply is limited. Such demand-responsive lighting systems help in maintaining the stability of power grid.

Adoption of SSL lighting technology enables advanced control strategies which harness nonvisual benefits of lighting for health and well-being. Examples include scene setting, color temperature control and spectral tuning strategies which in

addition to energy savings also provide occupant comfort and circadian rhythm-based productivity and performance augmentation.

Networked systems provide the most flexible means to realize the full potential of lighting control and support advanced applications. The networked lighting control systems connect digital dimming ballasts or digital LED drivers to sensors, room controllers, remote controls, and user interfaces in a scalable architecture that can support small stand-alone spaces to multiple rooms, areas, floors, and entire buildings. More advanced systems extend their reach beyond one building to support enterprise-wide lighting management deployed over the Internet. These systems are capable of collecting a variety of data including presence, ambient light, temperature, humidity, noise, air quality, dimming level, power consumption, user interactions, user preferences, device status, and device failures.

Similarly, a wide-scale deployment of outdoor lighting control networks in urban spaces can facilitate collection of rich sensing data about the environment and infrastructure such as light levels, air quality, traffic, weather, road conditions, and emergency situations. The ability to measure real-time spatiotemporal data combined with historic data can provide deep insights, enabling optimization of light levels, lighting system maintenance schedule, traffic planning, and road maintenance planning.

The state-of-the-art systems exploit these data for energy management, demand response, trend analysis, user comfort maintenance, automatic fault detections and diagnosis, and predictive maintenance. The optimal operational parameters, strategies, and schedules are decided based on actual facility needs, current occupancy patterns, weather conditions, onsite power generation, and demand management policies. Sophisticated algorithms are implemented to detect malfunctioning sensors, communication failures, battery depletions, tempering of sensors, outliers in occupancy, illuminance, temperature, and other anomalies and then alert the facility manager. Advance failure predictions enable scheduled maintenance actions which are more cost-effective. Suggestions to overcome performance bottlenecks are provided to the facility managers to ensure that savings persists and maintenance costs are reduced.

Intuitive visualizations for monitoring, situation awareness, and alerts are cornerstones of modern lighting control systems. Notifications of alarms via emails and smart phones are also provided for quick incident response. Some systems also support operational, financial, and environmental reporting and work order management for streamlining the maintenance processes. Costs are attached to issues to facilitate budget justification and spending prioritization.

In the longer term, intelligent lighting controls will act as eyes, ears, noses, and brains of the buildings and public places. They will play pivotal role in responding to emergencies, enhancing safety, and enriching the environment. For example, occupancy sensors will detect the presence of trapped humans in the building in case of disasters such as fires or earthquakes. Outdoor lights will direct the emergency crew to accidents or crime scenes. The intelligent control strategies will be able to aggregate data from multiple unobtrusive sensors embedded in the environment to learn user's preferences. The autonomous learning algorithms will then anticipate

user's needs and personalized user's environment by adapting light scenes for activity type, time of the day, room temperature, occupancy, daylight, glare, mood, etc.

Lighting will be increasingly used for applications beyond illumination. One great example is visible light communication (VLC) where data bits are embedded in the light. VLC can provide high-speed wireless connectivity to augment overloaded RF connectivity channels such as Wi-Fi. In order to embed high-speed data in light, VLC-enabled luminaire will need high-speed network connectivity which underscores the need for high-speed communications technologies for the lighting control systems. VLC also enables accurate indoor location tracking for retail shops, malls, hospital, museum, and public buildings.

Inspired by the benefits lighting control strategies offer to the occupants, owners, and society, many state and local energy codes and standards have mandated some form of lighting controls, for example, California Title 24. Similarly, many lighting control measures have been recognized as essential components of overall energy efficiency by many certification and accreditation agencies, e.g., Green Building Certification Institute.

In order to support the vision of smart lighting systems, we need very efficient and scalable networking technologies that can support building-wide, enterprise-wide, and city-wide connectivity. Lighting control technology landscape is highly fragmented, and there are many technology choices available to the designers, system developers, buyers, and users. The purpose of this chapter is to provide a brief overview of standardized networking technologies suitable for lighting control applications.

Proprietary Versus Standardized Protocols

A proprietary protocol is a closed protocol that is developed by a single manufacturer. The knowledge of underlying technology rests with the manufacturer who developed it. Developers of proprietary protocols sometimes make their technology available to other implementers under the contract or by opening their specification. The downside of the proprietary system is that the buyer is locked into products of a single manufacture. Although the initial costs of proprietary systems can be competitive, in the long run they may turn out to be expensive as some vendors take advantage of being the sole supplier for additions and maintenance by engaging in price gouging. Another risk of proprietary systems is that if the manufacturer goes out of business or stops the production of the components, then buyer will not have alternative supplier for the expansion or upkeep of the system. On the upside, the proprietary technology could be superior to standardized alternatives available in the market. Currently, the lighting controls market is very fragmented, and there are many proprietary systems available in the market.

An open protocol is typically a standard that is developed, published, and maintained by a well-recognized organization (e.g., ISO, IEC, IEEE, ANSI, ASHRAE, etc.) or an industrial alliance (e.g., ZigBee, BACnet, HomePlug, etc.).

Standards are developed by industry professionals through open and democratic processes. Any change to the standardized protocol typically requires addressing the comments of users, implementers, and professionals before being approved by the standardization body. It is possible that proprietary intellectual properties are included in a standard, but usually they can be licensed. Nominal fees might be associated with usage of the standard to cover the cost of development and administration. Standardized protocols typically have a testing and certification body which tests and certifies the protocol implementations to ensure interoperability among devices from different vendors.

Standardized protocols offer vendor independence which is the key benefit over proprietary protocols. Standardization enables multiple vendors to bring products to the market which fosters competition and keeps prices in check. Standardized products offer more buying choices with varying functions, features, and price tags. By the virtue of being open, standardized protocols facilitate integration of various domain-specific subsystems such as HVAC, lighting, and shading. Standardization also nurtures the ecosystem of user, designer, developer, engineer, implementer, and installer communities. It also provides freedom for knowledge sharing, skill enhancement, training, open-source tools, online resources, and FAQs. Industry benefits from the vibrant ecosystem and shared learning. Hence, in this chapter, we focus mainly on the standardized protocols.

Standardized Wired Protocols

Early lighting control systems were based on wired analog technologies which were later replaced by digital systems. Most digital wired systems allow devices to share a physical cable for communication which simplifies the wiring and reduces installation errors compared to analog systems. Typically, digital wired systems offer more bandwidth and high reliability. They are less susceptible to interferences and offer robust connectivity. However, laying wires in retrofit buildings, streets, highways, and parking garages can be very difficult, incurring high installation costs. Wired systems are less flexible to reconfigure compared to wireless systems. Nevertheless wired systems are widely deployed for lighting controls applications. In this section, we review the standardized wired protocols suitable for lighting control applications.

0–10 V DC

0–10 V DC (ANSI E1.3–2001 (R2011)) is one of the oldest signaling methods widely used in commercial and industrial lighting control applications. As the name suggests, the signal comprises of a DC voltage which ranges between 0 and 10 V. In one variant, the voltage varies from 0 to +10 V, and in another variant the voltage varies from 0 to –10 V.

When a 10 V signal is applied to the lighting system, it sets its output to 100 % (i.e., max output), and when a 0 V signal is applied, it sets its output to 0 % (i.e., off).

Dimmers interpret an intermediate voltage as a dimming signal and set the output in proportion to the input signal voltage. Depending on the design of the dimming device the output voltage, power or lumens is linearly dependent on the input signal voltage. Some of the dimming technologies do not support full range of dimming. For example, many types of fluorescent ballasts cannot be dimmed below 10 %. In that case, a switch or a relay is used to terminate the power supply to the lights to turn them off when input voltage is set to 0 V.

The main advantages of 0–10 V DC are simple design, implementation, and diagnosis. The main limitations are due to wiring complexity. Since it requires one wire per control channel and a common return wire, a large-scale installation could have lots of wires and connectors which increase not only material costs but also installation labor costs. Long cables induce voltage drops which may require calibration of receiving device to compensate for voltage drop. Moreover, the signal is susceptible to interference from adjacent AC power cables.

DALI

Digital Addressable Lighting Interface (DALI) is an international standard (IEC 62386) for the networked lighting control. DALI specifies an asynchronous, half-duplex, serial protocol over a two-wire differential bus which supports the data rate of 1200 bit/s. The devices can be networked in a daisy chain or star topology or a combination of these. Due to the diode bridge in the interface circuitry, DALI wiring is polarity insensitive. Signal interface is galvanically separated and doesn't need any termination resistors. DALI bus requires a power supply at 16 V DC, and supply current can be no more than 250 mA. This voltage appears on the data cables and can be used to supply power to peripherals such as motion detectors. DALI employs Manchester encoding to mitigate interference caused by electrical noise. The network cable is mains-rated with 600 V isolation, and it can run in the same raceway that includes mains power. These features simplify cabling and reduce installation costs.

A DALI controller can monitor and control lighting devices using the two-way communication. The controller can query individual lighting devices, and the specific device responds with the requested information such as lamp status. DALI network comprises of up to 64 lighting devices that have DALI interfaces. Each lighting device is assigned a unique static address in the range 0–63. DALI devices can be addressed individually or using group and scene broadcast message to address multiple devices simultaneously (e.g., “Group 1 set to 100 %” or “Set scene A”).

DALI defines 256 dimming intensity levels between off and 100 %. These levels are translated to ballast power levels via a logarithmic dimming curve which adapts larger increments in brightness at high dim levels and smaller increments at low dim levels. This is to adjust the dimming levels in a way that appears linear to the human eye.

Industry alliance AG DALI (<http://www.dali-ag.org/>) promotes adoption of DALI protocol. Alliance members (about 120 companies as of Feb. 2015) are allowed to use the DALI trademark on DALI compliant devices.

DMX512

DMX512 is a communication standard (ANSI E1.11–2008 (R2013)) widely used for controlling lighting and special effects in theatrical applications. It is also used in architecture lighting and entertainment applications.

DMX512 employs RS 485 differential signaling over twisted pair as its physical layer. DMX512 supports unidirectional asynchronous serial communication at 250 Kbit/s data rate. It does not have built-in error checking so it is not suitable for controlling hazardous applications such as pyrotechnics.

A DMX512 network is called a “DMX universe.” The 512 after the DMX refers to the number of control channels used in one universe. One device may use one or more channels (e.g., three channels for controlling red, green, and blue independently). In a DMX512 network, the devices are daisy chained, and a terminator is connected to the last device on the daisy chain to absorb signal reflection. A single DMX512 controller (e.g., a lighting console) acts as a master, and all other devices (e.g., dimming lights, fog machines, etc.) act as slave devices. No more than 32 units of load can be on a single bus. To support more than 32 units of the load, the network can be expanded across parallel buses using DMX splitters. The DMX512 standard specifies a five-pin XLR-type connector where only the first three pins are actually used. DMX512 standard is maintained by a trade association named Professional Lighting and Sound Association (PLASA, <https://www.plasa.org/>).

RDM

Remote Device Management (RDM) is an ANSI standard (ANSI E1.20–2010) which enhances DMX512 to enable bidirectional communication between an RDM compliant lighting controller and RDM compliant lighting fixtures over a DMX line. RDM is designed to be compatible with DMX512 equipment. Thus, RDM compliant controller and RDM complaint responders (receivers) can operate on a DMX line without interfering with the normal operation of standard DMX512 devices that do not recognize the RDM protocol. Moreover, DMX512 and RDM receivers can be used with a legacy DMX512 controller to form a DMX512 only system.

RDM packets are relayed in between the DMX512 data packets. The DMX512 data packets use the default Start Code 0×00 , whereas RDM packets use the start code $0 \times CC$. Hence, the legacy device unaware of RDM will not be able to read RDM packets.

RDM communication can be broadly classified in 3 types: discovery, unicast, and broadcast. In discovery, the controller discovers all the devices connected to the bus. The controller broadcasts a discovery command and waits for a response. If more than one device attempts to respond simultaneously, then collision will occur and controller will not receive a response. The controller will refine its search to a small range of addresses following a binary search algorithm. Once a device is found, it is prohibited from responding again to discovery messages.

In unicast communication, a pair of devices communicates with each other. To address collisions during unicast, RDM authorizes only one device to transmit at any given time. Only controller can send a request to a device. After the request has been sent, the controller relinquishes control of the DMX line for a given period of time, so the device can transmit its response. If controller does not receive a response in a given time, then it may retry. In broadcast communication, the controller sends a message to multiple fixtures. Responses are prohibited in broadcast communication except during the discovery.

RDM is also maintained by PLASA. In order to test interoperability among devices of different vendors, PLASA holds RDM plugfests several times a year.

Architecture for Control Networks (ACN)

Architecture for Control Networks (ACN) is a suite of network protocols designed primarily for theatrical lighting applications where lighting effects are synchronized with audio, video, automation, special effects, and pyrotechnics. ACN is an ANSI standard (ANSI E1.17–2010). The protocol leverages mainstream networking protocols (e.g., UDP/IP), enabling it to run over commercial off-the-shelf inexpensive hardware (e.g., Ethernet and Wi-Fi). In a typical ACN application, a controller discovers devices, automatically learns how to control them (by reading device descriptions), and then controls devices across the network by getting and setting values of properties of those devices.

Protocol Architecture

ACN builds upon widely used Internet protocols. Figure 1 shows the layered architecture of ACN. The lowest ACN layer above the underlying transport (e.g., UDP) is the Root Layer Protocol (RLP). RLP specifies the methods for combining

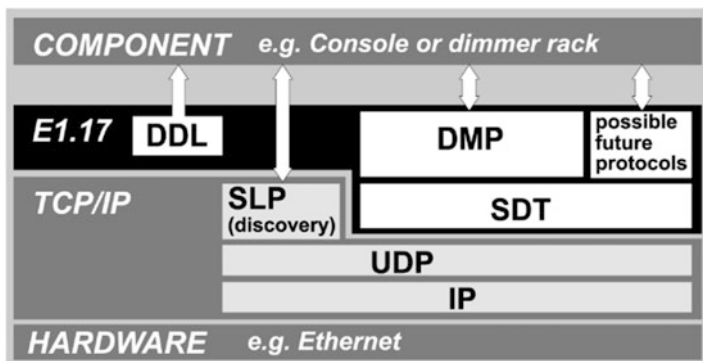


Fig. 1 Architecture of control networks protocol stack

protocol data units (PDUs) from different ACN subprotocols into packets for transmission via the transport. RLP also defined methods for parsing and processing PDUs upon receipt. ACN uses Service Location Protocol (SLP) for discovering devices on the ACN network. ACN defines the Session Data Transport (SDT) subprotocol over UDP to support ordered delivery, reliability, and online status. Ordered delivery is achieved with the help of sequence numbers, and reliability is achieved by negatively acknowledging missed messages.

ACN defines Device Description Language (DDL) which describes controllable devices in a machine readable language, enabling controllers to automatically interface to them. DDL is based on the Extensible Markup Language (XML) syntax. Device Management Protocol (DMP) provides configuration, monitoring, and control of devices. DMP defines method for getting and setting values of device properties. It also defines addressing structure necessary to identify individual properties and messages used to manipulate them.

ACN also defines interoperability profiles which specify the combinations of layers and operating parameters (e.g., specific values of timing parameter) to enable interoperability in the desired application environment. ACN also supports integration with other protocols such as MIDI, SMPTE, DMX512-A, RDM, and RS232 through gateways that translate between ACN and other protocols.

ACN is also maintained by PLASA. There are several ongoing projects which are developing the open-source libraries of the ACN protocols for various computing platforms.

BACnet

Building automation and control networks (**BACnet**) is a communication protocol designed for building automation and control systems such as heating, ventilating, and air-conditioning (HVAC) control, lighting control, access control, and fire detection systems. BACnet is an ASHRAE and ([ANSI standard 135–2012](#)) protocol.

The BACnet protocol architecture has four layers which correspond to the physical, data link, network, and application layers of the OSI reference model.

Physical and Data Link Layer

The physical and data link layer provides the capability to unicast and broadcast messages to the devices on the same network. BACnet provides many options for physical and data link layers. First two options are Ethernet and ARCNET. Third option is the Master-slave/Token-Passing (MS/TP) protocol defined in BACnet standard which runs over an EIA-485 physical layer. The fourth option is the Point-To-Point protocol which provides methods for hardwired or serial (RS-232) asynchronous communication. BACnet also specifies LonTalk, Internet Protocol, and ZigBee as additional options. All these technologies provide a choice of master/slave, token-passing, or contention-based medium access control over twisted-pair, coax, fiber optic, or wireless media.

Network Layer

BACnet protocol defines a network layer which enables messages to travel from the source devices to destination devices irrespective of the BACnet data link layer technology used by intermediate nodes. Routers are used to connect disparate BACnet LANs (e.g., Ethernet and MS/TP) and relay messages between them. Routers automatically build and maintain their routing tables to facilitate message flows.

BACnet messages can also be relayed over the networks that use the IP as their networking protocol in two distinct ways. The first method is IP tunneling where BACnet messages from a source are encapsulated inside IP packets by the ingress IP router. The ingress IP router relays the resulting IP packets toward the egress IP router over the IP networked. The egress IP router recovers the encapsulated BACnet messages and forwards them to the destination on the BACnet LAN.

The second method is called BACnet/IP. In BACnet/IP, each BACnet device is an IP node which has its own IP address and IP protocol stack. BACnet/IP defines a new protocol layer called the BACnet Virtual Link Layer (BVLL). BVLL uses UDP/IP to send and receive messages over the IP network. Since IP routers typically don't forward the broadcast messages, BACnet defines a BACnet Broadcast Management Device which handles the broadcast traffic.

Application Layer

Objects: An object is a data structure which models a device and stores device information in the form of properties. Typical objects include analog inputs, analog outputs, binary inputs, binary outputs, calendar, event-enrollment, file, and schedule. Recently, lighting centric objects such as lighting output and channel are developed for lighting control applications. Objects have properties, for example, the "present value" property of an analog input object which typically represents data generated by a sensor or the state of a physical device (e.g., lights). Other properties associated with an analog input object include fault status, reliability, object name, minimum and maximum limits, etc. The "lighting command" property of lighting output object provides lighting specific functions such as ramping, stepping, fading, and blink-warn. Each property has an identifier, a data type, and a conformance code that indicates whether it is mandatory or optional and whether it is read-only or writeable.

Services: BACnet also defines a number of services such as object access services, alarm and event services, file access services, remote device management services, and more. Object access services are used to read or write properties of objects. Alarm and event services are used to notify a remote device if a specified value has been changed or a specified event has occurred. The file access services can be used to read or write the contents of the file. Remote device management services are used to discover devices and objects on the network, time synchronization, starting and stopping communication, etc. These services are implemented using a client-server paradigm where the client device sends a service request to the server device and the server responds with a service response.

Interoperability

BACnet ensures interoperability of devices using three paradigms. The first is interoperability areas (IAs) which define the set of capabilities in terms of specific BACnet elements that must be implemented in a particular device to enable interoperability. These elements are listed in the device profiles specified in the standard. The second is definition of “BACnet Interoperability Building Blocks” or “BIBBs” which specify whether supporting device is expected to initiate or execute a particular service. The third is the definition of a set of BACnet Device Profiles which specifies the expected capabilities of a device in each of the five interoperability areas.

BACnet International is an industry organization that promotes the use of BACnet in building automation and control systems through trade shows, publications, and educational programs. BACnet Testing Laboratories (BTL) was established by BACnet International to perform compliance and interoperability testing activities. As of Feb. 2015, the BTL product listings contain 625 active products from 107 distinct manufacturers.

LonWorks

LonWorks is a suit of open protocols originally developed by Echelon Corporation and currently standardized by (ISO/IEC 14908, Parts 1, 2, 3, and 4). LonWorks is widely used for HVAC control, lighting control, security, fire detection, and other building automation systems. LonWorks specifies the control network protocol (CNP) which follows the seven-layer OSI model. CNP offers several choices for physical layers technology including two variants of fiber optic which support 1.25 Mbit/s, three variants of Power Line which support up to 3.987 Kbit/s, free topology twisted pair which supports 78.13 Kbit/s, RS-485 twisted pair which supports 39.06 Kbit/s, and transformer isolated twisted pair which supports 1.25 Mbit/s ([Introduction to the LonWorks Platform](#)).

CNP uses a MAC algorithm called predictive p-persistent CSMA for its link layer. In the p-persistent CSMA, a device transmits with a fixed probability p if the channel is idle and defers a transmission with probability $(1-p)$ when the channel is busy. In the predictive p-CSMA, the probability p is variable and dynamically adjusted based on the expected traffic load. The protocol supports optional priority mechanism where priority messages are given earlier access to the medium than nonpriority messages.

The network layer of CNP defines naming and addressing of devices to facilitate message exchange. Hierarchical addresses are used which enable the router to filter messages based on their destination address. This layer also defines methods for routing messages from a source device to one or more destination devices including the scenario where the destination device is using a different physical layer technology.

CNP transport layer offers three types of message services. Acknowledged service where the sending device waits for an acknowledgement from receiver and if the acknowledgement does not arrive, then it resends the message up to a configurable number of retries. If acknowledgments are not required, then unacknowledged or repeated service can be used to send a message for configurable number of times without waiting for an acknowledgement.

Session layer offers the request/response service used for device management, fetching values, and other remote actions. It also defines an authentication protocol that enables receivers to determine if the sender is authorized to send the message. The presentation layer encodes the message as a network variable, application message, or foreign frame. Network variables are used by applications to exchange data. A standard network variable types (SNVTs) are defined to promote interoperability by ensuring that applications use a common interpretation of data exchanged through network variables. SNVT defines standard units, ranges, and resolution. For example, illumination measurements can be reported by the network variable type called SNVT_lux, ranging from 1 to 65535 lux with 1 lux resolution.

Connections can be established between output and input network variables on different devices or between output and input network variables on the same device. For example, a lighting device has a switch-type input network variable, and a dimmer switch device has a switch-type output network variable. These two devices can be connected through a process called binding, allowing the switch to control the lighting device.

The application layer defines a set of standard network services which include network configuration, network diagnostic, file transfer, application configuration, application management, alarming, scheduling, data logging, etc. The application layer services ensure that devices created by different manufacturers can interoperate with each other and can be configured using standard tools. LonWorks also defines IP tunneling standard to support IP-aware applications and remote network-management tools.

LonMark International is a global industry organization which defines device SNVTs, objects, profiles, and IP connectivity. It also provides interoperability design guidelines, product conformance testing, marketing assistance, educational programs, and product certification. As of Feb. 2015, LonMark International and its affiliate organizations have about 400 members worldwide and 521 LonMark-certified products.

KNX

KNX is an [ISO/IEC 14543-3](#) standard for home and building automation. KNX was born out of convergence of three previous standards: the European Home Systems Protocol (EHS), BâtiBUS, and the European Installation Bus (EIB or Instabus). The KNX protocol is widely used for lighting controls in continental Europe. Other applications include HVAC control, motorized blind control, safety and security, remote access control, and energy management.

KNX standard defines several options for physical and link layer technologies. It defines two variants of twisted-pair technology where data and power can be carried over one pair so that devices with limited power consumption can be powered by the bus. Both support bus, star, and tree topologies in any combination. They provide asynchronous character-oriented data transfer and half-duplex bidirectional communication at 4.8 Kbit/s and 9.6 Kbit/s. Both implement CSMA/CA methods for collision avoidance.

Two variants of power line communications are also defined which implement asynchronous half-duplex bidirectional communication at the data rates of 1.2 Kbit/s and 2.4 Kbit/s. Both implement CSMA methods for medium access. The RF technology specified by KNX implements asynchronous half-duplex bidirectional communication in 868 MHz frequency band at the data rate of 38.4 Kbit/s. Medium access follows CSMA mechanisms. These physical layer technologies support very low data rate which is not suitable for high data rate streaming or multimedia application. KNX also supports other physical layer technologies such as IR and Ethernet.

KNX network layer supports routing, hop limit, and segment-wise acknowledged datagram. Transport layer enables multicast, broadcast, connectionless, and connection-oriented services.

KNX applications implement data points which represent the processes, control variables, parameters, and diagnostic data in the system. These data points are contained in the group objects and interface object properties. Interoperability is achieved through standardized data-point types which are grouped into functional blocks. Data points can be accessed and modified through unicast or multicast mechanisms. Data points of applications on the networks are logically linked through binding mechanisms.

KNX groups the features, services, and objects into profiles to enable interoperability and certification. Network management processes are also included in profile definitions to facilitate integration. The manufacturer declares which profiles the device conform to and testing is done accordingly.

To facilitate discovery, a KNX device exposes essential information about itself (e.g., the profile it implements) to peers on the KNX network. To enable this, a set of device descriptor fields and objects with corresponding discovery procedures are standardized.

The KNX standard specifies three configuration modes to choose from when developing a KNX compatible device. The S (“system”)-mode is meant for well-trained installers to link products and configure them using PC-based tool to implement sophisticated building control functions. The E (“easy”)-mode devices are already preprogrammed and loaded with a default set of parameters. Using a simple configurator (need not be a PC tool), each component can be easily reconfigured by an installer with basic KNX training. The (“automatic”)-mode achieves “plug-and-play” configuration meant for consumer products. The three configuration modes share common run-time environment and networking.

KNX association (<http://www.knx.org/>) promotes the adoption of KNX through training, periodicals, conferences, meetings, and software development activities.

For system configuration, KNX association offers a vendor neutral tool called Engineering Tool Software (ETS). According to KNX association, as of Feb. 2015, 340 KNX member companies worldwide offer almost 7,000 KNX-certified product groups in their catalogues.

X10

X10 is a legacy industry standard communication protocol that uses power line wiring and RF wireless for signaling. It is designed for home automation applications. The main intent behind X10 protocol is to overcome the difficulty in installing new wires for automation and control by leveraging the existing power supply wiring for signaling.

The data is encoded on a 120 kHz carrier signal which is transmitted as 1 ms burst at zero crossing of 50/60 Hz alternating current waveform. All messages are sent twice in succession to improve reliability. After accounting for signaling and retransmission overhead, X10 supports the throughput of about 20 bit/s. Since the throughput is very low, only simple operations such as switching household appliances, thermostats, plug-load controls, and audible alarms are supported. X10-based lighting control applications can support on, off, dimming, and scene setting features. X10 sensing modules for temperature, light, infrared, motion, and contact closures/openings are also available.

To support wireless keypads, remote controls, motion sensors, burglar alarms, and other types of devices, an RF protocol is also defined. A bridge translates the X10 RF packets into X10 power line control messages. The RF protocol operates at a frequency of 310 MHz in the USA, 418 MHz in Britain and Europe, and 433.92 MHz in Europe.

Legacy X10 protocol has several limitations. The standard X10 power line and RF protocols do not support encryption. Unfiltered X10 power line signal from close neighbors who happen to use the same device addresses could inadvertently control each other's devices. To address this issue, inductive filters are used which attenuate the signal leaving or entering the premises. Unintentional (or malicious) control of neighbor's devices can also happen when someone with an X10 RF remote control adjusts X10 devices using RF to power line bridge.

Another major limitation of X10 technology is that 120 kHz carrier signal cannot pass through the high impedance of the distribution transformer winding. To address this issue, capacitors are installed between two leg wires to provide path for X10 signals. Active X10 repeaters are needed to support interphase communication in homes that have 3-phase power.

Many power supplies used in electronic devices such as TV and computers attenuate X10 signals by providing a low impedance path to high-frequency signals. Sometimes filters are also used in certain types of power supply that attenuate X10 signals traveling on that branch circuit blocking the communication.

HomePlug

HomePlug is a family of various power line communication standards that exploit existing electrical wiring for home networking. Since new wires need not be installed, the installation is quick, easy, and relatively inexpensive. HomePlug AV specification is aimed at multimedia applications such as in-home distribution of gaming, Internet, HDTV, and DVR content. HomePlug AV specification supports peak data rate of 200 Mbit/s at physical layer and 80 Mbit/s second at MAC layer. HomePlug AV2 specification supports data rate of 1.26 Gbit/s at physical layer. HomePlug AV is defined as the baseline technology for the FFT-OFDM PHY within the IEEE 1901 Broadband Power Line Standard.

HomePlug Green PHY specification is targeted at low-power and low data rate applications such as smart metering, smart energy, security, lighting, HVAC, and plug-in vehicles. HomePlug Green PHY supports a peak data rate of 10 Mbit/s at physical layer. The three main specifications published by HomePlug (HomePlug AV, HomePlug AV2, and HomePlug Green PHY) are interoperable. All the HomePlug standards implement AES 128 encryption.

HomePlug Alliance (<http://www.homeplug.org/>) is an industry group of 60 companies working together to develop technology specifications, testing, and certification for power line communications. As of Feb. 2015, there are 182 HomePlug-certified products from HomePlug alliance members. Many of these products are adapters that plug into wall outlet (or power strips or extension cords) and are used to transfer high-speed data to outlet connected devices.

G.hn

G.hn is the common name for family of home networking standards developed by ITU and commercialized by HomeGrid alliance. The main motivation behind development of these standards was IPTV, especially when offered by service provider as a bundle of voice, video, and data service. Another focus area is smart home and smart grid applications.

The G.hn family of standards defines communication protocols over power lines, phone lines, and coaxial cables with data rates up to 1 Gbit/s. The intent is to enable a single G.hn semiconductor device to support multiple wire types, thereby reducing the development costs for device manufacturers and lowering the deployment cost for service providers (by enabling customer self-install).

G.hn MAC layer uses TDMA scheme. G.hn defines the concept of relay nodes which are capable of forwarding the messages, thereby enabling extended coverage of the network. G.hn utilizes AES 128 encryption for security. G.hn also defines profiles to manage device complexity and enable interoperability. Home automation, home security, and smart grid devices are example devices based on low complexity profiles.

The HomeGrid forum (<http://www.homegridforum.org/>) is an industry alliance of about 70 member companies responsible for marketing, compliance testing, certification, and promotion of G.hn. As of Feb. 2015, about 22 vendors offer HomeGrid certified products.

Standardized Wireless Protocols

Rapid evolution of wireless technology has inspired lighting controls industry to develop wireless lighting control protocols. One of the main benefits of wireless technology is that it enables easy and flexible installation of devices. Devices can be placed where they are needed including areas that are difficult to wire. After the installation, the devices can be easily moved or reconfigured when the layout of the space changes. By eliminating wires needed for communication, wireless technology saves on labor costs of installing wires, which can be substantial in retrofit buildings, parking garages, and street/highway lighting. Without the need to open ceilings, dig earth, or punch holes to pull wires, the wireless systems can be installed quickly, causing little disruption to normal operations.

On the other hand, wireless systems offer low data rate and are susceptible to interferences. The range of wireless signal can be greatly reduced due to obstacles in the environment requiring signal repeaters which would add to the cost. These issues could introduce delay in command execution. Battery-operated wireless devices need periodic battery replacements (typically after many years).

ZigBee

ZigBee is a low-power, low data-rate wireless standard designed for wide range of applications including home automation, smart buildings, industrial automation, home entertainment, patient monitoring, smart meters, and other consumer and industrial applications.

ZigBee defines a network, security, and application layer protocol suite on top of the PHY and MAC layers defined by the IEEE 802.15.4 wireless personal area network (WPAN) standard. The PHY layer exploits the direct sequence spread spectrum technique for interference tolerance, and MAC exploits CSMA with collision avoidance (CSMA/CA) for channel access. IEEE 802.15.4 defines the data rates of 250 Kbit/s in 2.4 GHz band, 40 Kbit/s in the 915 MHz band, and 20 Kbit/s in the 868 MHz band. It provides transmission range of 10 m–100 m (depending on power output and environmental characteristics) at the data rate of 250 Kbit/s in 2.4 GHz band.

ZigBee supports flexible network topologies including star, tree, and mesh. When two devices are not within the direct transmission range of each other, they can communicate by using intermediate nodes as routers (provided at least one route can be found). ZigBee routers form a mesh network to which low-duty-cycle ZigBee end devices connect as leaf nodes to form a hybrid topology. ZigBee is optimized for

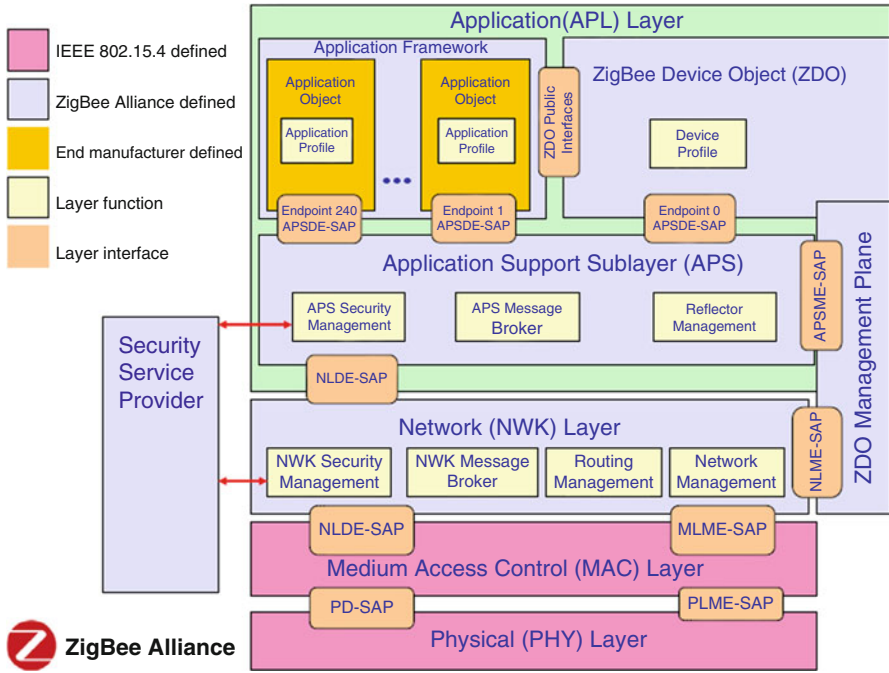


Fig. 2 ZigBee protocol architecture (source <http://www.zigbee.org/>)

low-duty cycle operation of end devices (i.e., an end device can turn off the radio most of the time) to enable long battery life. A ZigBee end device does not take part in routing of messages which reduces the memory requirements, thereby reducing the manufacturing cost compared to routers. In the non-beacon-enabled ZigBee networks, the routers have their transceivers continuously active, requiring continuous power supply.

This type of architecture supports heterogeneous networks in which routers are listening for incoming messages, whereas sleepy end devices wake up and transmit when an external stimulus is applied. The typical application of such architecture is battery-operated light switch which when pressed sends commands to a ZigBee router embedded in mains powered lamps. The ZigBee end node in the switch will wake up when switch is pressed, send commands to the lamp, receive acknowledgements, and then return to the sleep.

A ZigBee network has one coordinator device which is responsible for network creation and maintenance. It also serves as the trust center and repository for security keys. New nodes can join the network typically through a sequence of button presses on the device already part of the network and the device being added. ZigBee uses AODV as the routing protocol to support mesh routing. ZigBee Pro added many-to-one and source routing features to reduce the storage space required and routing overhead. The ZigBee security suite is built on the AES-128 Standard.

The ZigBee application layer defines several advanced features and services to facilitate information exchange among networked devices. These include addressing, application commands, device discovery services, service discovery services, binding services, and management services.

ZigBee defines application profiles which are groups of messages, formats, conventions, and security features targeted for specific application domain such as home automation, personal home, hospital care, etc. ZigBee-enabled applications from different manufacturers can interoperate if they implement the same profile.

The ZigBee Alliance (<http://www.zigbee.org/>) is formed by a group of companies to develop, maintain, and promote the ZigBee standard and also to test and certify ZigBee products. As of Feb. 2015, the alliance has about 400 members and claims over 1000 ZigBee-certified products. The Alliance publishes public application profiles that allow multiple OEM vendors to create interoperable products. Public application profiles relevant for lighting controls include home automation which defines devices intended for use in the home such as lights and switches, blinds, wall outlets, remotes, thermostats, and HVAC equipment. Another public profile, commercial building automation, defines devices such as advanced lights and switches, keyless entry, and security systems. Light Link is another public application profile targeted specifically for lighting control applications. ZigBee members may also apply for manufacturer-specific profiles which are not defined by ZigBee Alliance but instead are defined by the OEMs making the products. Private profiles are used for those applications that do not need to interoperate with other manufacturer's products.

6LoWPAN

6LoWPAN stands for IPv6 over low-power wireless personal area networks. The intent of 6LoWPAN is to enable low-power, low-data rate wireless devices with limited capabilities to take part in Internet of Things. The 6LoWPAN working group in IETF defines the header compression, fragmentation and reassembly, and routing methods that enable IPv6 packets to be transported over low-power low data rate IEEE 802.15.4-based WPAN. This technology is targeted for home automation, smart meters, building controls, industrial automation, asset management, and other low data rate automation applications.

6LoWPAN addresses the key challenges associated with transmitting large IP messages over low-power, low data rate WPAN. The IPv6 requires that links support a maximum transmission unit (MTU) of at least 1280 octets. On the other hand, the size of IEEE 802.15.4 packet is 127 octets. The MAC layer overheads of IEEE 802.15.4 can be 25 octets without security which leaves 102 octets for payload. If AES-CCM-128 security is used at the MAC layer, then overhead reaches to 46 octets, leaving only 81 octets for upper layers. Standard IPv6 header is 40 octets and UDP header is 8 octets which will leave very little room for application data.

To address these issues, stateless header compression methods are defined which compress the size of IPv6 and UDP headers. Similarly, to transport large IP packages

over IEEE 802.15.4-based network, 6LoWPAN defines fragmentation and reassembly methods. The fragmented IPv6 packets have to be routed within PAN for which mesh routing protocols are specified. Neighbor discovery protocol is also part of the specifications to facilitate network autoconfiguration. 6LoWPAN network connects to the Internet through a gateway (or boarder router). 6LoWPAN utilizes AES-128 security features built into IEEE 802.15.4 standard.

Several open-source and commercial implementations of 6LoWPAN are available in the market. Noteworthy projects using 6LoWPAN technology include the Smart Energy and Home Automation: Restful Architecture ([SAHARA](#)) project which uses web services for smart energy and home automation. Similarly, the European Union FP7 HOBNET project ([HOBNET](#)) uses 6LoWPAN technology for the automation and energy management of smart and green building. 6LoWPAN technology is also widely used in many commercial large-scale smart meter deployments.

In 2014 a new industry alliance named Thread (<http://www.threadgroup.org/>) was formed to promote interoperability among home automation devices that use 6LoWPAN for communication. Target applications include home appliances, climate control, lighting, energy management, safety, and security.

Z-Wave

Z-Wave is a low-power, low data rate wireless communications technology designed for home automation applications to communicate with lights, HVAC, automated window treatments, smoke alarms, sensors, access controllers, entertainment systems, and other domestic appliances. Z-Wave is also used in battery-operated consumer electronics devices such as remote controls.

Z-Wave supports data rates of 9.6 Kbit/s, 40 Kbit/s, and 100 Kbit/s in 900 MHz band. Transmission range is 30 m in free space, with reduced range in indoor environments. Z-Wave defines two types of nodes: controllers and slave devices. Each network can support up to 232 nodes in mesh network architecture and has one primary controller and zero or more secondary controllers that control routing and security.

Devices are added to the Z-Wave network by pressing a sequence of buttons on the controller and the device being added (also known as “pairing” and “adding”). Z-Wave provides source-routed mesh network to route around household obstacles and radio dead zones, ensuring sufficient coverage for most residential houses. In order to route the messages, the intermediate node should be awake which may not be the case for battery-operated devices. Therefore, battery-operated devices are not designed as routers. Z-Wave supports AES-128 for security.

Z-Wave was developed by a startup named Zensys that was later acquired by Sigma Designs. The Z-Wave Alliance (<http://z-wavealliance.org/>) is a consortium of about 300 manufacturers who intend to develop products based on the Z-Wave standard. According to Z-Wave alliance, as of Feb. 2015, there are more than 1200 different products certified and over 35 million products sold since 2005.

EnOcean

The EnOcean is an energy-harvesting wireless technology aimed at building automation systems. It utilizes micro energy converters to harvest energy which is then used to drive ultra-low-power wireless modules to communicate among battery-less wireless sensors, switches, controllers, and gateways. EnOcean is an international standard (ISO/IEC 14543-3-10) which defines physical, data link, and networking layers.

The EnOcean technology is designed to support battery-less maintenance-free operation. It can operate on electrical energy harvested from mechanical motion, indoor light or temperature difference using electromagnetic, and photovoltaic or thermoelectric energy converters. Design choices are made to make the overall system sustainable even with very small and intermittent amounts of energy harvesting. The size of wireless data packet is only 14 bytes which are transmitted at 125 Kbit/s. Instead of requiring the acknowledgement for reliability, three identical copies of the packets are sent at pseudo-random intervals. EnOcean uses 902 MHz, 928.35 MHz, 868.3 MHz, and 315 MHz spectrum for transmission. The transmission range is up to 300 m in the open and up to 30 m inside buildings. EnOcean provides repeaters to extend the range of communication. EnOcean offers various options for security including AES-128 encryption.

In lighting control applications, battery-free wireless switch eliminates the need for installing wires between the switch and controller (e.g., luminaire) which saves time and material used. EnOcean technology is also used to develop battery-free occupancy sensors, photo-sensors, temperature sensors, CO₂ sensors, and power metering sensors for lighting control and HVAC applications.

EnOcean GmbH is a spin-off company of Siemens founded in 2001. In order to promote the technology, the EnOcean Alliance (<https://www.enocean.com/en/home/>) has been formed. As of Feb. 2015, EnOcean alliance claims about 350 member companies, 1500 interoperable products, and over 250,000 EnOcean-enabled buildings.

Standardized Application Protocol

TALQ

TALQ is an industry standard (<http://www.talq-consortium.org/>) for outdoor lighting control applications intended for enabling interoperability between Outdoor Lighting Networks (OLN) of different vendors and the Central Management System (CMS). TALQ focuses on application layer interface that enables management of multi-vendor OLN by a single CMS. TALQ standardizes the interface for information exchange between heterogeneous OLN and CMS. It does not specify the underlying communication technology within the OLN. Vendors are free to adopt suitable communication technology (e.g., RF, power line, cellular), commissioning methods, and security features within their OLN. The TALQ protocol is built on

widely used Internet protocols (TCP/IP, HTTP, and XML) and security standards (transport layer security).

The interface specification between OLN and CMS defines two end points, one residing at the OLN side called TALQ Bridge and another residing at CMS side called CMS. These end points utilize the underlying network and transport layer services to support reliable bidirectional communication between the OLN and CMS. This enables the control of devices inside the OLN through the TALQ bridge.

TALQ standard supports lighting control features including scheduling, sensor-based control, configuration, and monitoring. Standard includes advanced features for asset management, event handling, energy savings, failure notifications, and maintenance optimization.

Summary

The technological advances in LED light sources, sensors, and control systems have ushered a new era of smart lighting control systems that provide the right quantity and quality of light when and where it is needed in order to enhance the energy savings, maintenance savings, comfort, health, productivity, safety, well-being, and user satisfaction. State-of-the-art lighting control systems collect presence, temperature, humidity, air quality, ambient light, dimming level, power consumption, user interactions, user preferences, device status, device failures, and many other types of data points. These data can be exploited to offer energy management, space optimization, demand response, trend analysis, user comfort maintenance, automatic fault detections and diagnosis, predictive maintenance, and other advanced applications and services.

In order to realize the full potential of smart lighting systems, we need very efficient and scalable networking technologies that can support building-wide, enterprise-wide, and city-wide connectivity. The building automation industry has realized the benefits of deploying standardized communication protocols over proprietary solutions, leading to the development of a plethora of communications standards. However, so far, no standard has emerged as the dominant choice, leading to a very fragmented market landscape. In Table 1 we have summarized the key features of standardized communications protocols we have reviewed in this chapter.

In coming years, lighting control industry is expected to slowly migrate from proprietary systems toward standardized systems. In the near term, the BACnet, KNX and LonWorks will remain the major forces in building automation for large commercial buildings as they facilitate easier integration among the lighting, HVAC, safety, security, and building management systems. With increasing acceptance for wireless technologies due to improved reliability and lower installation costs especially in retrofit buildings, ZigBee and 6LoWPAN is expected to see an uptake in demand. The number of connected devices in smart homes is also exploding which would increase the market for standardized wireless connectivity solutions such as ZigBee and 6LoWPAN.

Table 1 Key features of standardized communications protocols

Protocol	Physical media	Data rate	Topology	Application domain	Lighting applications
0–10 V DC	2 wires	N/A	N/A	Analog lighting control	+++++
DALI	2 wire differential bus	1.2 Kbit/s	Star and daisy chain	Digital lighting control	+++++
DMX512	RS 485	250 Kbit/s	Daisy chain	Theatrical, entertainment, and architectural lighting control	+++++
RDM	RS 485	250 Kbit/s	Daisy chain	Theatrical, entertainment, and architectural lighting control	+++++
ACN	Ethernet/Wi-Fi	~1 Gbit/s	Star, tree	Theatrical, entertainment, and architectural lighting control	++
BACnet	RS 485	115 Kbit/s	Daisy chain	Lighting, HVAC, security, building management systems	+++
	Ethernet	~1 Gbit/s	Star, tree		
LonWorks	Fiber optics	1.25 Mbit/s	Daisy chain	Lighting, HVAC, security, home automation, building management system	++++
	Power line	4 Kbit/s	Bus		
	Free topology twisted pair	78.13 Kbit/s	Free		
	Transformer-isolated twisted pair	1.25 Mbit/s	Bus		
	Ethernet/Wi-Fi (IP tunneling)	~1 Gbit/s	Star, tree		
KNX	Twisted pair	4.8 and 9.6 Kbit/s	Bus, star, tree	Lighting, HVAC, security, building management system	+++
	Power line	1.2 and 2.4 Kbit/s	Bus		
	RF	38.4 Kbit/s	Star		
	Ethernet	~1 Gbit/s	Star, tree		
X10	Power line	20 bit/s	Bus	Lighting, home automation	+++++

(continued)

Table 1 (continued)

Protocol	Physical media	Data rate	Topology	Application domain	Lighting applications
HomePlug	Power line (Green PHY)	10 Mbit/s	Bus	Smart home, smart grid, plug-in vehicle charging	+
	Power line (AV)	200 Mbit/s	Bus	In-home multimedia content distribution, gaming, Internet	
	Power line (AV2)	1.26 Gbit/s	Bus		
G.hn	Power lines, phone lines, and coaxial cables	1 Gbit/s	Bus	IPTV, voice, video, Internet, home automation, smart energy	+
ZigBee	RF IEEE 802.15.4	20 Kbit/s, 40 Kbit/s and 250 Kbit/s	Star, tree, and mesh	Lighting, HVAC, security, home automation, smart meters, industrial automation, building automation	+++++
6LoWPAN	RF IEEE 802.15.4	250 Kbit/s	Star, tree, and mesh	Lighting, HVAC, security, home automation, smart meters, industrial automation, building automation	+
Z-Wave	RF	9.6 Kbit/s, 40 Kbit/s, and 100 Kbit/s	Mesh	Lighting, HVAC, security, home automation	+++++
EnOcean	RF	125 Kbit/s	Star	Lighting, HVAC, home automation, building automation	+++++
TALQ	N/A	N/A	N/A	Outdoor light management	+++
Rarely used for lighting applications					+
A few lighting applications					++
Some lighting applications exists					+++
Many lighting applications exists					++++
Widely deployed lighting applications					+++++

While Internet protocol has been at the forefront of information technology revolution, their influence in building automation arena has been modest. Although IP protocols provide excellent interoperable connectivity solution, connectivity alone is not enough to develop interoperable applications. On the top of connectivity provided by IP technology, the industry needs to define standardized application profiles, commands, data structures, and semantics to develop interoperable applications. There have been many attempts to exploit the IP technology for lighting controls and building automation such as ACN, 6LoWPAN, BACnet over IP, ZigBee over IP, and a few more, but the progress has been limited. Some vendors have developed proprietary end-to-end IP-based lighting control systems (e.g., power over Ethernet-based LM-IP system from Philips). In the longer term, more IP-based solutions are expected to emerge and stake the claim at building automation market due to their inherent strengths such as ubiquity, large installed base, robustness, technological maturity, availability of skilled manpower and tools, and ease of integration with other IT systems.

References

- ANSI/ASHRAE Standard 135–2012 BACnet – a data communication protocol for building automation and control networks
- ANSI E1.11 – 2008 (R2013), Entertainment technology – USITT DMX512-A, asynchronous serial digital data transmission standard for controlling lighting equipment and accessories
- ANSI E1.17 – 2010 Entertainment technology – architecture for control networks (ACN)
- ANSI E1.20 – 2010 Entertainment technology-RDM-remote device management over USITT DMX512 networks
- ANSI E1.3 – 2001 (R2011) Entertainment technology – lighting control systems – 0 to 10V analog control specification
- HOBNET: HOlistic Platform Design for Smart Buildings of the Future InterNET. http://hobnet-testbeds.eu/index.php?option=com_content&view=article&id=4&Itemid=12
- IEC 62386 Digital addressable lighting interface standards
- Introduction to BACnet For building owners and engineers. www.BACnetinternational.org
- Introduction to the LonWorks Platform, Rev 2. Echelon Corporation
- ISO/IEC 14543-3 Information technology – Home electronic system (HES) architecture
- ISO/IEC 14543-3-10:2012 Information technology – home electronic systems (HES) – Part 3–10: wireless short-packet (WSP) protocol optimized for energy harvesting – architecture and lower layer protocols
- ISO/IEC 14908-X:2012 Information technology – control network protocol. Part 1, 2, 3 and 4
- SAHARA:Smart energy and home automation: restful architecture. <http://sahara.tzi.org/?lang=en>
- Williams A, Atkinson B, Garbesi K, Page E, Rubinstein F (2012) Lighting control in commercial buildings. *Leukos* 9(3):161–180

Adaptive Control Technology for Lighting Systems

Francis Rubinstein

Contents

Introduction	584
Impact of New Technologies on Adaptive Controls for Lighting	585
Evolution of Adaptive Controls for Commercial Building Lighting	586
Early Adaptive Controls	586
Electronic Ballasts for Fluorescent Lamps	587
Lighting Control Systems	588
Barriers to Adaptive Control Systems	589
The Need for Dimming	590
Analog Dimmable Ballasts	591
Analog LED Drivers	591
Digital Control	593
DALI	593
Wireless Load Controllers	594
Energy Saving Strategies for Adaptive Controls	594
Demand Response	595
Lighting Control Networks	596
Luminaire Level Lighting Controls	597
Measured Energy Savings from Adaptive Controls	599
Two Demonstrations of Luminaire Level Lighting Controls	600
Results	601
Separating the Effect of Adaptive Controls from the Light Source Change	602
References	605

Abstract

This chapter examines the effect of major technology trends on the use of adaptive controls in commercial and institutional buildings. Adaptive controls were originally designed to reduce the energy wasted by building lighting

F. Rubinstein (✉)

Lawrence Berkeley National Laboratory, Berkeley, CA, USA

e-mail: fmrubinstein@icloud.com

systems. Office buildings, in particular, were notorious for leaving lights burning all night. This conspicuous waste led lighting companies to develop lighting schedulers and occupant sensors to begin to deal with the most obvious problem. These early lighting control products formed the basis of a slowly growing lighting controls market. As digital control technology gradually supplanted analog control, manufacturers would field control systems of increasing complexity and capabilities, such as daylight harvesting systems with automated shading and demand response-enabled controllers.

Introduction

This chapter examines the effect of major technology trends on the use of adaptive controls in commercial and institutional buildings. Adaptive controls were originally designed to reduce the energy wasted by building lighting systems. Office buildings, in particular, were notorious for leaving lights burning all night. This conspicuous waste led lighting companies to develop lighting schedulers and occupant sensors to begin to deal with the most obvious problem. These early lighting control products formed the basis of a slowly growing lighting controls market. As digital control technology gradually supplanted analog control, manufacturers would field control systems of increasing complexity and capabilities, such as daylight harvesting systems with automated shading and demand response-enabled controllers.

Prior to the growth of LEDs (from 2010 and on) adaptive control technologies were used primarily to effectively manage building lighting in larger institutions and facilities and to optimize the building's lighting energy use by the use of different control strategies. But high installation costs and uncertainty about how effective controls would be in typical applications resulted in only slow growth of the adaptive controls market during the 1990s and early 2000s. Stricter building codes, such as ASHRAE 90–2010 and California's Title 24, began to accelerate the uptake of adaptive controls in new buildings since new codes effectively required the use of daylighting, demand response, and occupancy-responsive lighting strategies. Equally important, full-dimming capabilities are being required for most new installed lighting. But the real acceleration of adaptive controls into commercial buildings is beginning only now as LED lighting establishes a toehold in the general lighting market.

As LED lighting evolved, adaptive lighting control technologies were rapidly developed to leverage the best features of this new dynamic light source. LED lighting is much easier to control and dim than legacy light sources such as fluorescent or HID lighting. Compared to fluorescent lighting, LEDs can be dimmed over a larger range including all the way off and can be rapidly switched ON and OFF with impunity. It is a great advantage to the LED industry that the small, low-voltage nature of LEDs is so amenable to easy, inexpensive dimming. As LED lighting increasingly infiltrates general lighting, manufacturers began making lighting in different shapes and sizes whose color temperature could be varied from warm

to cool on command. This ability to change color temperature on demand from wall display, automation system, or smart phone added a fundamentally new dimension to adaptive lighting.

A key advantage of LEDs from an adaptive controls perspective is that LEDs are an intrinsically low-voltage light source, so they can be easily adapted to microprocessors, connectivity software, and the Internet. This improved connectivity is leading to panoply of new adaptive control products that are well-suited to applications in the commercial (as well as residential and outdoor) sectors. Most importantly, many of these new applications were designed more with improving lighting quality and leveraging the advantage of the color shifting properties of LEDs and less with simply creating lighting that was energy-efficient and nothing more. With these new adaptive control technologies, lighting designers and specifiers now have the tools to create lighting experiences, which are energy efficient, comfortable for the occupants, and have minimal impact on the environment.

Impact of New Technologies on Adaptive Controls for Lighting

The exponential growth of LEDs into general lighting is driven by the five major technology trends shown in Table 1.

First, emerging LED fixtures and lamps developed over the past few years are poised to rapidly displace legacy fluorescent lighting in the commercial sector. Although much of the appeal of LED lighting is its high efficacy and the ability to change the color of the light, LEDs are also much easier to dim from a technical

Table 1 Key technology trends in adaptive controls

Current practice	Future trend	Rationale for shift
Static (nondimming) fluorescent	Full-range dim-to-off LED lighting with color tuning as option	Improved dimming range, efficiency, and performance with LED drivers
Microprocessor in central controller	Microcontroller embedded into each light	Allows addressability and connectivity at the fixture level. Allows standardized interfaces and minimizes rewiring costs
Area-based controls where one sensor controls large block or row of lights	Sensing and control at the individual fixture	Reduced commissioning costs and setup times
Separate controller and driver (ballast)	Microcontroller and connectivity integrated into driver (ballast)	Lower capital and labor costs
Wired lighting zones	Wireless and powerline-controlled lighting and hybrid schemes combining wired device networks and wireless controllers	Reduced wiring and labor costs, especially in existing buildings

standpoint than fluorescent lighting. Also significant is that the incremental cost of adding dimming to an LED driver is much less than adding dimming to a fluorescent ballast.

Secondly, microprocessors are increasingly being attached to or embedded into each lighting fixture. Previously, microprocessors were installed only at the controller level. But as microprocessor costs have dropped, it is increasingly cost-effective to install microprocessors in the fixture itself. Thus, groups of lighting fixtures can be physically connected with CAT-5 type cabling and form digital control networks capable of addressing individual fixtures and exchanging data in a robust and reliable manner. As the density of microprocessors increases and intelligence now resides within each light, digital control techniques (individual addressing, bidirectional signal flow, etc.) are now available to lighting controls and fixture manufacturers.

Third, the commercial linear fixture market is moving away from area-based controls and trending towards fixtures with embedded sensors and controls. By increasing the density of occupancy and light sensors, the adaptive controls system can sense the luminous environment at a small, human scale, which can improve the occupants' buy-in to the new lighting.

Fourth, and most compelling from the smart lighting perspective, is that LED lights are increasingly being embedded with a microprocessor that can control both the intensity and the color of individual fixtures or lamps.

Finally, new communications technologies allow lighting fixtures and lamps to be addressed over the Internet by means of physical wires, powerline carrier, or wirelessly. By connected lighting systems and lamps to the Internet, lighting systems will form the digital beachhead to interconnect all network services in our buildings, homes, and streets as we move towards the Internet of Things (or the Internet of Everything). The following sections describe how these five technology trends have shaped adaptive lighting controls in commercial applications.

Evolution of Adaptive Controls for Commercial Building Lighting

Historically, lighting controls have been viewed as a necessary but unremarkable adjunct to a building's electrical system. Strategically located wall switches, along with appropriately circuited and zoned lighting systems, have been a mainstay of electric lighting operations for over 100 years.

Early Adaptive Controls

Basic lighting systems in older commercial buildings typically include manual wall switches to control individual office fixtures and larger zones of fixtures in open plan areas. Very often, some form of automated lighting schedule is included in the controls scheme to turn lights off after hours, based on occupancy schedules set on timers that control circuits, zones, or entire floors. These can be either manual on and automatic off, requiring occupants to turn lights on upon entering a space so that

lights are not turned on automatically when the operating schedule begins even if occupants have not yet arrived, or automatic on and off. Normally, either option can be overridden by wall or zone switches or relays if after-hours occupancy is necessary. Emergency lights that stay on 24 h a day to illuminate ingress and egress zones are common in these floor spaces, as well, and normally operate on separate, dedicated circuits not subject to the automated schedules.

As long as electricity costs were low, little attention was paid to how building lighting operated. All that changed with the oil embargo in 1973. The subsequent disruptions to petroleum supplies between 1973 and 1979 caused a real shock to the energy economy as a whole. Long lines at the gas pumps were just the most obvious manifestation of the disruption caused by spiking oil prices. By the end of the 1970s, the real price of electricity went up nearly 20 %. This increase in electricity costs spurred development of electronic ballasts for fluorescent lamps and improved phosphor lamps, and several new lighting control innovations hit the market. The first generation of occupant sensors and daylight sensors were produced and hastily brought to market in an effort to reduce the energy wasted by lighting systems.

Occupancy sensors that automatically turn lights off after a space is vacated are less common than simple manual switches and automated schedules but are implemented in many commercial office buildings. These are most common in individual private offices, where typical operation is manual on, automatic off, giving occupants the option of using or not using their overhead lights. Occupancy sensors are less commonly deployed in open office areas; when they are, it is typically in an automatic on and off configuration. Occupancy sensors in open offices are usually installed so that a zone of fixtures covering multiple workstations is configured to be controlled from a single sensor.

Fluorescent lighting continues to dominate lighting in the commercial sector, so it is no surprise that adaptive controls originally focused on this key light source. Although fluorescent lighting is the workhorse of building lighting, dimming fluorescent lighting is expensive and is difficult to install for reasons that are outlined below. It would not be until 2010 when emerging LED lighting began shifting the calculus so that dimming could cost-effectively be prescribed everywhere.

Electronic Ballasts for Fluorescent Lamps

The electronic ballast for fluorescent lamps was a particularly important enabling technology that was developed and perfected during the 1980s. Electronic ballasts, combined with improved fluorescent lamps, were a significant breakthrough at the time and resulted in major increases in both lamp/ballast system efficacy and quality of light at a modest cost premium. During the late 1980s and early 1990s, spurred on by the restructuring of the utility sector, some utilities began to incentivize property owners to upgrade their fluorescent lighting systems from magnetically ballasted T-12 fluorescent systems to electronic ballasts and T-8 fluorescent lamps. The combination of better improved system efficacy (lumens per watt) and better quality light (due to improved rare-earth phosphors) was sufficiently compelling that by the

Table 2 Major energy savings strategies for adaptive controls

Energy or demand reduction strategy	Definition	Implementation issues
Occupancy	Reduces lighting energy and unnecessary operation by lowering light levels or turning lights off in offices and zones when occupants leave an area. Reduces demand by taking advantage of variable occupancy patterns within individual zones throughout an office or building	Occupant sensing at the individual fixture level improves reliability
Daylighting	Allows lighting systems to save energy by taking advantage of the available natural light. Photosensors detect the level of illumination along the perimeter of the building floor and adjust the electric light output to achieve a target light level	Light sensing at the individual fixture level improves overall performance
Personal controls	Adjustment of light levels by occupants according to personal preference. Applies to all spaces where occupants have a sense of ownership of the lighting	Fully leverage smart phone technology to take advantage of LED color tuning
Institutional tuning	Allows building managers or tenants with a dimmable lighting system to decrease light levels and lighting energy consumption by programming default power levels for fixture zones or individual fixtures at a lower level than maximum power and light output to reflect actual building lighting needs and policies regarding light levels provided	Excellent retrofit strategy since light levels may often be reduced
Demand response	Ability to adjust light levels based on electric grid reliability, energy price, or other factor. Incentives for reducing lighting demand when requested are set according to a demand response tariff	Auto demand response significantly improves reliability and customer acceptance of DR

early 2000s, electronic ballasts and T-8 fluorescent lamps had saturated the market and largely displaced magnetic ballasts in many commercial applications. The first fully-dimmable electronic ballasts were also developed during this time as well as the first integrated lighting control systems that could implement all the major lighting controls strategies (see Table 2).

Lighting Control Systems

During the 1980s and 1990s, more sophisticated lighting control systems emerged. These systems were usually revolved around a proprietary network technology and often required components sourced from a single manufacturer. These systems were

a combination of network-connected field application panels with embedded time schedulers, relay panels, and connected occupancy sensors and daylight sensors.

Most conventional lighting systems can only be switched off and not dimmed although this is changing now with the availability of LED systems. Before LEDs, dimming and multilevel fluorescent lighting was available but at a significant cost premium compared to nondimming equipment.

In some applications (conference rooms for example), dimming may have been required for functional reasons (i.e., scene setting) or other aesthetic reasons. To support these specialized functions, manufacturers fielded proprietary controls solutions that consisted of fully-dimmable (or multilevel) ballasts, dimmer switches, multibutton scene selectors, and other analog components. Then, as State and National Energy Codes began to drive the direction of the lighting market in general, these systems were enhanced and expanded to allow selective implementation of combinations of lighting control strategies over the entire buildings area according to specific requirements.

Early installations of adaptive controls often only attempted to implement one or two control strategies (such as daylight harvesting in daylighted areas and occupancy sensing in private offices). These simpler installations did not require a large networked lighting control system, which resulted in less expensive yet less capable that were difficult to maintain. To implement several control strategies in combination requires a networked system of lighting controllers, networks, and sensors as well as the microprocessor power to implement different operational sequences for all the connected lighting. Networked systems are powerful and are capable of managing large areas of lighting for improved functionality, user comfort, and operational efficiency.

Compared to manual controls and automated schedules typical in offices, networked lighting controls can better match lighting system operation to the needs of the occupants, providing light when needed and at illumination levels more appropriate to the conditions of the space. Controls save energy by operating lights only when needed and at no more power than necessary to do the job. These systems operate at a higher spatial and temporal resolution than basic lighting controls and can give users better access to local lights.

Barriers to Adaptive Control Systems

Despite the availability of advanced lighting controls, only 2 % of commercial buildings in the USA employ photosensors for daylighting control and only 1 % utilize installed energy management and lighting control systems (Williams et al. 2012). Advanced lighting controls uptake in the commercial market has been hindered by high installation costs, which can include high equipment costs as well as high labor costs, due to factors such as extensive controls wiring, system complexity, laborer unfamiliarity, and commissioning requirements.

Because of the added expense and real (and perceived) complexities associated with installing well-functioning adaptive controls, the penetration of advanced

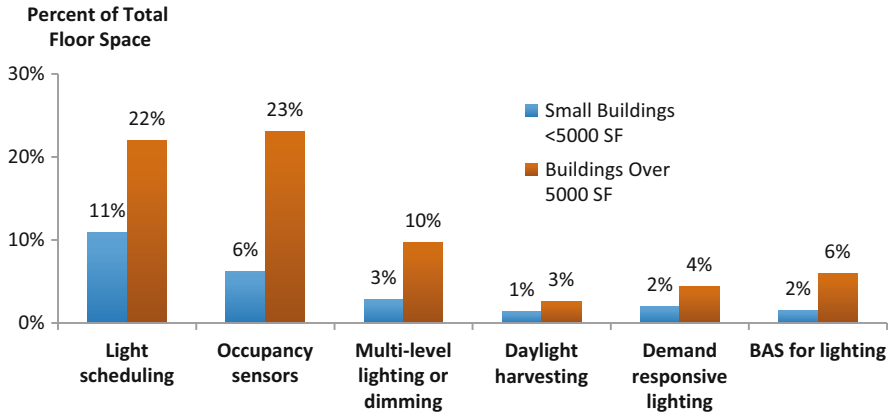


Fig. 1 Penetration of different types of adaptive controls in commercial buildings under and over 5000 ft². Slightly over half (2.8 million) of the 5.5 million commercial buildings in USA are less than 5000 ft²

control strategies has been quite slow, especially in the very small building market (Energy Information Administration 2015). Buildings under 5,000 ft² are about 50 % of the total building stock. Yet, the use of control strategies such as dimming, daylighting, or demand response is in the single digits in these buildings (Fig. 1).

The Need for Dimming

For adaptive controls to fully penetrate the commercial sector, building lighting needs to become fully dimmable in operation. Lighting that can only be switched ON and OFF is too inflexible to be useful in implementing complex, automated lighting control strategies, such as daylighting and demand response. Even occupancy sensing, which really only requires ON/OFF operation to work, can be more comfortable for the occupants if fade-to-off and other gentle fades are implemented. More important, in order for daylighting and demand response to be acceptable to occupants in typical settings, automated changes of light level must be as inconspicuous as possible. Only dimming equipment can provide the gradual changes in output that are so necessary to successful lighting services in modern buildings.

Dimming control gear is available both for fluorescent and LED light sources. Fluorescent lamps use dimming ballasts as control gear while LED arrays are controlled using specialized power supplies called drivers. Analog ballasts and LED drivers use a 0–10 VDC circuit to vary the light output of the source. Analog-controlled ballasts and drivers are the most common form of dimmable control gear, but there are several digital control schemes that are increasingly being used instead, particularly DALI and DMX.

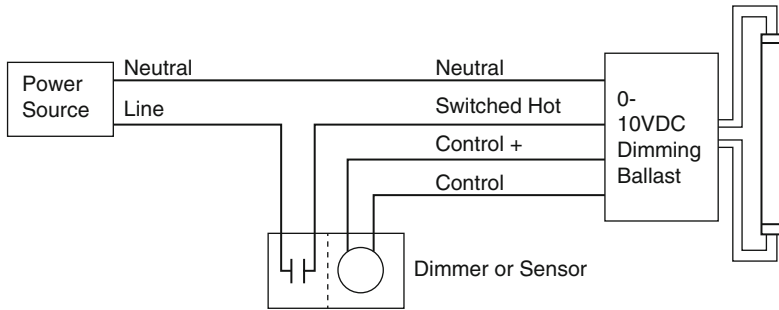


Fig. 2 Typical wiring diagram for a 0–10 VDC dimming fluorescent ballast. Each 0–10 VDC dimming ballast within a given zone is controlled by a two-wire, polarity sensitive parallel bus. A separate switch or relay is needed to energize and de-energize the ballast

Analog Dimmable Ballasts

Analog dimming ballasts vary the fluorescent lamp output from 100 % to 10 % light output using a 0–10 VDC control circuit. The simplest method for dimming blocks of lights in unison uses an analog control loop with the analog ballasts connected in parallel. Every connected ballast operates in lockstep with all the other ballasts on the same control loop.

Controlling groups of lights in this manner is often the least costly path but also the least flexible. The zoning of the lights is entirely determined by how the wiring is installed. There is no addressing of individual ballasts within the group nor can zones be configured after installation without expensive rewiring.

A complication with analog fluorescent ballasts is that they cannot be switched ON and OFF from the (low voltage) analog loop. Rather, a separate switch (or relay) must be used to energize the power wiring (see Fig. 2). This complicates the control wiring and further increases the cost premium of dimming, which is already higher than for static systems.

Analog LED Drivers

Why Dimming Is Better with LEDs

The fundamental difference between LED and fluorescent is that LEDs are solid-state devices while fluorescent lamps are gas discharge devices. Gas discharge lamps work by running electric current through an evacuated glass tube containing tiny amounts of mercury and other elements to produce a glowing plasma. Since plasmas are nonlinear loads, a ballast is used to carefully regulate the amount of current flowing to the lamps. Fluorescent lamps also require relatively high voltages (100 VAC) to start and then

(continued)

maintain the plasma. LEDs, on the other hand, being solid-state devices are intrinsically low voltage and can be easily dimmed from a less complex LED driver. Dimming ballasts, because they operate tunable resonant circuits, are inherently more complex in design than static ballasts or LED drivers. Dimming fluorescent lamps requires varying the frequency of the drive current as well as its magnitude. To maintain good lamp life, dimming ballasts must also carefully control the amount of voltage being supplied to the lamp filaments. Taken together, these factors mean that dimming fluorescent lamps is intrinsically more difficult, less reliable, and significantly more costly than dimming LEDs (Figs. 3 and 4).

For these reasons, dimming of fluorescent lamps, especially at lowest output, tends to be suboptimal in terms of energy efficiency and lamp stability.

Fig. 3 Relative efficacy as a function of relative input power for 2 T-8 lamps operated by electronic dimming ballast. At minimum dim level, the system consumes 27 % of full power

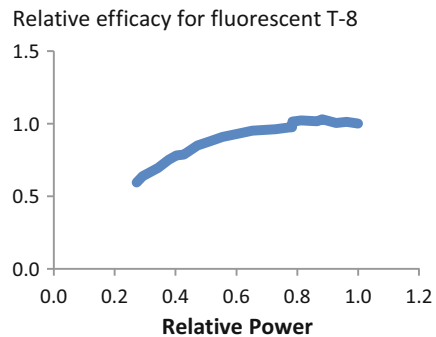
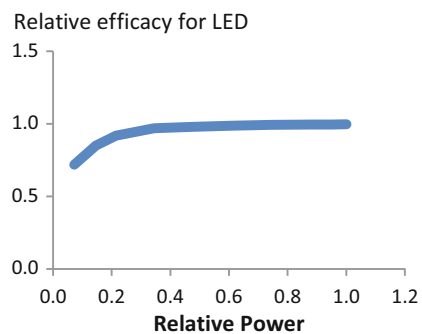


Fig. 4 Relative efficacy as a function of relative input power for a PWM (pulse-width modulated) LED driver. At minimum dim level, the system consumes 7 % of full power



Unlike fluorescent ballasts, which are mostly static, LED drivers are either capable of 50 % output (in addition to full output) or are fully dimmable using 0–10 VDC. Furthermore, most LED drivers do not require a separate relay for the power, which simplifies control of this light source. Analog LED drivers are convenient stepping

stones for lighting controls manufacturers striving to embed adaptive controls into luminaires. By adding a microprocessor to the luminaire, either by means of an additional fixture-installed controller connected to the LED driver or embedded in the driver itself, LED luminaires can be “adaptive controls enabled.”

As the industry moves towards individually addressable lighting, the use of analog control loops is too inflexible to accommodate today’s control requirements. However, each analog-controlled LED driver can be fitted with a module that combines a microprocessor and an analog circuit for controlling the driver. In this way, the analog to digital frontier is pushed all the way out to the individual driver. This improvement in performance is also driven by the trend to install microprocessors at the device level rather than at a centralized controller. In the near future, the LED source, driver, microprocessor, and embedded sensors will all be combine on one board, thus driving down cost further and improving performance.

Digital Control

There are many different digital control protocols that can be used for lighting. DALI is the most commonly used digital protocol for fluorescent ballasts (and now LED drivers) although other protocols, such as DMX, are used in some higher-end applications (Illuminating Engineering Society of North America 2011).

DALI

DALI (Digital Addressable Lighting Interface) is a simple digital protocol originally designed to dim groups fluorescent dimming ballasts on a simple two-wire control bus. Each DALI ballast has a unique address and all DALI ballasts are connected using a two-wire, polarity insensitive bus capable of supporting up to 64 devices on one network bus (Fig. 5).

Different lighting scenes can be invoked from a wall-mounted scene controller. The scene controller allows the user or the lighting automation system to call

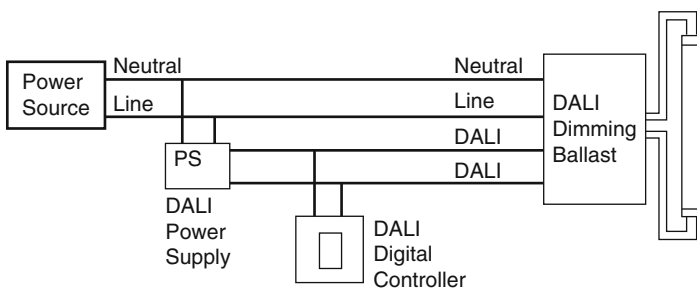


Fig. 5 Circuit showing DALI dimming ballast controlled by a controller and powered by a power supply

different scenes for different groups of lights. Additionally, newer implementations of DALI (DALI 2.0) allow for sensors, such as occupancy sensors or daylight sensors to be directly operated on a DALI circuit. DALI 2.0 adds a third byte to the original two-byte DALI protocol. This greatly increases the potential usefulness of DALI for adaptive controls purposes. But even with the additional features of DALI 2.0, the 128 device limit (2×64 devices on a typical DALI router), the slow speed (1200 BAUD), and other technical limitations means that most DALI networks need to be wrapped in higher order software that can exploit the full potential of DALI.

Wireless Load Controllers

Several lighting controls manufacturers have chosen to use 0–10 VDC LED drivers (or fluorescent ballasts) that are connected, at the fixture level, to luminaire adaptor. This adaptor, or load controller, is able to apply the correct analog voltage to drive the light source according to a signal delivered as input. This input can be a wireless protocol, such as ZigBee or Wi-Fi. By combining wireless control of individual luminaires with wireless sensors and switches, a flexible adaptive control network that can be adapted to many commercial applications.

Energy Saving Strategies for Adaptive Controls

The main purpose for adaptive controls in commercial buildings has been to implement different combinations of energy savings strategies to reduce the energy footprint of the building lighting.

All the major lighting control companies offer lighting products and systems capable of implementing all the above control strategies and others as well. However, the networking methods, control protocols, and system architecture vary considerably from system to system.

While energy savings has been the main reason for using adaptive controls in the first place, with dimmable LED lighting dropping in price, it is becoming easy to use adaptive controls for more than just minimizing operating costs. Occupant comfort improved amenity and occupant access to local lighting are becoming increasingly important, especially for knowledge workers, hospitality industry, healthcare, and education sectors.

Software is playing an increasingly important role in strengthening uptake of adaptive controls in commercial lighting systems. Since every lighting installation is different, well-crafted software is necessary to allow easy and efficient management of the lighting throughout a facility.

With larger networked systems, the lighting control system software is critical not only to the day-to-day management of the lighting but also the “look-and-feel” of the overall system to the facilities managers. Control software that is designed to interoperate with other automated systems and is well connected to the Internet

will be more useful and successful in the long run than a system designed to operate in isolation.

Demand Response

Demand response is a load management strategy that aims to reduce the building electric demand during times of high electricity demand, high electricity price, or other grid stress. While the utility incentives to customers participating in demand response programs vary widely across the country, most programs reward the customer for decreasing their electric demand for 4–6 h on critical peak days (Piette et al. 2009a). As a cost saving strategy for the customer, demand response is much more reliable if the response to the demand alert is automated and preprogrammed to reflect the customers load priorities.

Open source solutions for automated demand response (or OpenADR) are available through the Demand Response Research Center. OpenADR makes it relatively simple for networked lighting control manufacturers to add demand response capabilities to their Internet-connected control systems. Demand response offers a whole new set of lighting-as-a-service opportunities for adaptive controls manufacturers. For example, some adaptive control software allows the normal operation of a building's lighting control systems to be modified during times of grid stress as announced by the OpenADR signal. Table 3 shows examples of how specific operational sequences for the different energy savings control strategies can be modified in response to demand response events.

Under normal conditions, the control system operates with the usual strategy, such as dimming electric lighting in concert with available daylight to provide a given light level. During times of grid stress, high electricity costs, or other triggering event, the lighting controller would modify the setpoint to a lower level for the

Table 3 Control strategy design intent and modifications during demand response events

Energy savings strategy	Design intent of strategy during normal operation (typical example)	Modification during times of grid stress
Daylighting	Automatically balance available daylight using dimmable electric lighting to provide 500 lx at all occupied times	Reduce setpoint to 200 lx during times of grid stress
Occupancy sensing	Reduce light levels in corridors, stairwells, and unoccupied spaces during times of vacancy	Turn lights off in corridors, stairwells, and unoccupied spaces when vacant OR reduce timeout intervals
Institutional tuning	Reduce light levels on area basis to correct for over-lighting, lamp CCT, institutional policy, or workgroup preference	Reduce light levels even further
Personal controls	Light levels are selected based on local user preference	Light levels prevented from exceeding set level

duration of the demand response event. Thus, the DR controls would maintain control of lighting loads but operate the lighting at a lower level of service.

The operational sequences for all the control strategies can be likewise modified during times of grid stress using the modifications described in Table 3. By treating demand response as an overlay on top of already existing control strategies, the reach and effectiveness of adaptive controls is increased without any increase in cost.

Lighting Control Networks

With recent advances in electronics, microprocessors, sensors, and connectivity, lighting control manufacturers have perfected large, networked lighting control systems that can manage large complex lighting installations according to any desired programming. These systems are capable of handling such diverse applications as running airport lighting to a college campus to a hospital. The specific combination of control strategies to be used in any given application depends fundamentally on what the programming requirements are for that space.

Networked lighting systems are especially useful in commercial and institutional building with large areas of managed lighting where occupants do not have a sense of “ownership” of the lighting. Large offices, hotels, airports, and campus complexes are all good examples where networked lighting controls are the most appropriate solution. Furthermore, most of these buildings have in-house facilities managers. Networked lighting systems generally require qualified personnel to operate the equipment and keep control software up-to-date.

Wired, wireless, or hybrid topologies that combine wired and wireless techniques are all used for network control of lighting systems.

The simplest system consists of load controllers, occupancy and daylight sensors, and switches and dimmable lights assembled into small, independent zones. Examples are private offices, hotel rooms, classrooms, etc. where the occupants or users of the space have a sense of ownership of the lighting. A wireless network system that implements most of control strategies for this first category is shown in Fig. 3. The gateway shown in the figure is the only component with a connection to the Internet. All the other components shown communicate wirelessly only with the gateway. There are several choices of wireless protocol although ZigBee, followed by Wi-Fi would be the most common (Fig. 6).

This same basic architecture can be implemented with a wired digital network (such as DALI) or even with a hybrid architecture in which wired DALI lighting is controlled wirelessly and/or uses wireless sensors that may be self-powered.

Since these systems use gateways, all the gateways can be tied to an RS-485 control bus which in turn is connected to the Internet. In this way, small zones can operate independently under most circumstances, but receive additional inputs from say a demand response automation signal to modify operational sequences under exceptional conditions.

For managing larger installations where the occupants have no sense of “ownership” of the lighting (such as airports or hospitals), a full-featured adaptive control

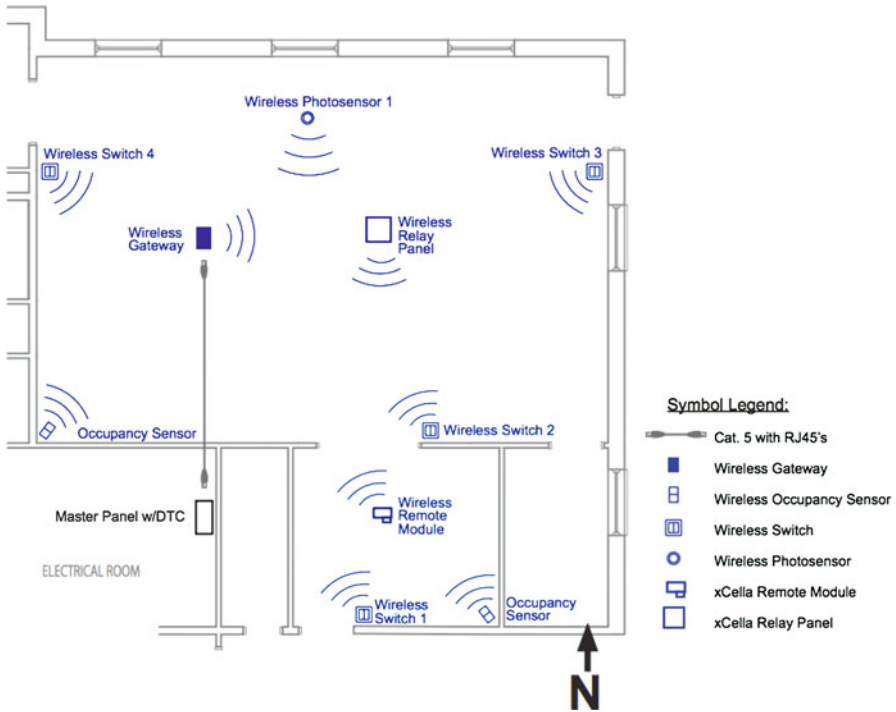


Fig. 6 Sample layout of wireless area-based control system with photosensors, switches, and occupancy sensors as indicated. The gateway is the only device that is connected to the Internet

system would be the preferred and most flexible solution. These systems tend to be zone based in that long rows of lights are controlled in unison from a zone-based daylight sensor or other input (such as demand response).

Luminaire Level Lighting Controls

Another important development in adaptive controls is the embedding of occupancy and daylight sensors within each luminaire rather than using area-based sensing and control as network lighting controls do. Individually controlled pendant-lighting has been available since the late 1990s. These early systems used dimming fluorescent lighting with an occupancy and daylight sensor integrated into the body of the fixture. Commissioning was accomplished using a PC application. As LED lighting began to penetrate the general lighting market, manufacturers began designing Luminaire Level Lighting Controls (LLLC) optimized for LED lighting, which could replace in-place fluorescent troffers and parabolic type fixtures on a one-for-one basis.

Fig. 7 Close-up of multisensor embedded in fixture housing. The small dome senses occupancy using passive infrared radiation. The dark aperture to the immediate right is for a light sensor. The sensor is wired to a controller (not visible) that is concealed in the ballast channel



LLLC is defined (Navigant Consulting Luminaire level lighting controls market baseline 2014) as having the following features:

- Full-range dimming
- Each luminaire embedded with a multisensor capable of measuring at a minimum: light level and occupancy status and optionally power consumption
- Ability to transmit and receive occupancy status (at a minimum) with other lights either wirelessly (e.g., ZigBee) or using other methods, such as IR or Visible Light Communications (VLC)
- Distributed control at each luminaire without need of a central controller

Commissioning for LLLCs is simplified because each fixture is typically capable of calibrating its response to changes in light level automatically with only minimal human intervention. Commissioning is generally limited to setting the spatial extent of a “work group” and adjusting the default ON and background levels (Fig. 7).

LLLC are well suited to existing buildings where it is not economical to rearrange fixture spacing and a “drop-in” replacement for the existing recessed fluorescent lights is most expedient.

Since LLLC luminaires each contain a light level sensor as well as a local control loop with commissionable setpoints, it is straightforward to calibrate the lighting operation so that the desired light levels (background level, default full ON, OFF etc.) are automatically provided as required with no human intervention. Thus, LLLCs provide a simple way to provide appropriate light levels in localized areas based on occupancy.

The current crop of LLLCs are usually unconnected to the local enterprise LAN or the Internet. Effectively each fixture can be considered an “island of control” with rudimentary capability to signal neighboring nodes as to the status of the occupancy sensor. The inability of this first generation of LLLCs to communicate with the

Internet means that they lack the ability to exchange information with systems outside the building. For many installations, especially those in small buildings, it may be preferable NOT to connect the lighting to the Internet, with all the latter's attendant risks. For others, the need to future-proof the lighting will dictate the use of connected lighting.

The second generation of LLLCs were on the market in 2014, and these can be considered "IOT-ready." These new luminaires will have many added features and capabilities including energy monitoring and reporting of occupancy data, demand response integration, etc.

In the future, it is expected that LLLC lighting will compete with network-based lighting. New construction and major renovation will tend to favor network-based controls, where any additional wiring costs can be subsumed in the overall project budget. For retrofits of existing buildings, the LLLC approach will be preferred since it is a basically a "drop-in" replacement with minimal required wiring.

Measured Energy Savings from Adaptive Controls

Various studies have addressed the energy savings potential of advanced lighting controls systems, looking at the implementation of different advanced controls strategies in various commercial spaces and, in many cases, measuring energy savings of specific controls options (e.g., occupancy sensors compared to manual control) and combinations of options (e.g., occupancy sensors and daylight sensors).

To aggregate the experiences and results from the many lighting controls studies available in the published literature, a meta-analysis of lighting controls energy savings in commercial buildings was carried out in 2011 (Williams et al. 2012). Of the 88 papers examined, the majority of installations were in office buildings with some educational buildings and a small cross-section of cases across the other building types. The study evaluated the energy saving effects of occupancy sensing, daylight sensing, personal tuning, and institutional tuning. For studies in which actual energy usage was monitored over time, energy savings averaged 24 % for occupancy sensors, 28 % for daylighting controls, 31 % for personal dimming control, 36 % for institutional tuning, and 38 % when more than one of these strategies were combined.

A more recent Pacific Gas and Electric (PG&E) Emerging Technology study in a GSA building in San Francisco in 2012 found energy savings of 21 % when fluorescent troffers were replaced with LED fixtures, and an additional savings of 41 % when advanced lighting controls were added, including task tuning to 80 % power, occupancy sensors, daylight sensors, and individual dimmers (Energy Services 2012). A recent GPG program study in GSA buildings evaluated advanced wireless lighting controls retrofit on existing fluorescent fixtures in one location and advanced controls with LED fixtures in another. The study found significant energy savings resulting from the LED fixtures, the advanced controls, and the combination of both (Williams et al. 2012). Advanced controls alone saved around 32 % energy compared to a baseline of basic lighting schedules, wall switches, and some private

office occupancy sensors. The LED fixtures saved an additional 30 %. Controls savings were not uniform across the offices in each study location; however, those that already had occupancy sensors in the base case saw little energy savings, while other spaces saw larger savings, up to nearly 50 %. Also, depending on where and how the controls were installed, occupant satisfaction varied, with some concerns over implementation at one location leading to slightly negative occupant feedback on some of the controls functions. The findings underscore how important good design and commissioning of controls schemes and zones is and some of the challenges of implementing complex controls systems in the real world.

More recently, (Rubinstein); (Wei et al. 2014a) analyzed the energy savings from 18 installations of adaptive controls in offices between 2009 and 2014. These included 10 sites where the 2×4 fluorescent lighting was replaced with suspended fluorescent luminaires with embedded occupant in daylight sensors, six sites with area-based controls, and two with Luminaire Level Lighting Controls (LLLC). The annual lighting energy use intensity (EUI) for all 18 sites after installation and commissioning of the controls varied from 0.8 to 2.6 kWh/ft²/year. When compared to the National Office average EUI of 4.1 kWh/ft²/year, the relative energy savings with adaptive controls varied from 37–80 %. It should be noted that 5 of the 18 sites achieved EUIs around 1 kWh/ft²/year and 3 of these five were in fact fluorescent, not LED. This shows that 1 kWh/ft²/year is a reasonable target for lighting with adaptive controls regardless of light source.

Two Demonstrations of Luminaire Level Lighting Controls

The energy use and lighting performance from Luminaire Level Lighting Controls (LLLC) was evaluated in two Federal buildings in 2013 as part of the General Service Administration's Green Proving Ground (Wei et al. 2014b). Both sites underwent a one-for-one replacement of existing $2' \times 4'$ fluorescent fixtures with the turnkey package of LED fixtures with integrated occupancy and daylight sensors and controls to turn the fixtures on and off, and dim and brighten them according to conditions in the office.

The system allows for tuning of fixture groups to reduce fixture power from maximum output to medium or low levels, if those settings meet the lighting needs of the space. Occupancy sensors integrated on each fixture detect when the immediate surroundings are occupied and turn fixtures on to the tuned power setting in response. Fixture groups programmed during system commissioning respond to occupancy patterns such that all fixtures in the group turn on to a low background level if any fixture within the group senses occupancy. Fixtures relay occupancy readings to the group through wireless communication. Only fixtures that individually sense occupants in their immediate vicinity brighten to the full-tuned output setting. Finally, each fixture includes an integrated daylight sensor for daylight harvesting. (For the system tested at these sites, the ZigBee Pro was used to communicate occupant sensor status. Other systems use different wireless protocols (such as WiFi or BLE), while still others use IR. In the future, visible light

Table 4 Physical information on site conditions at Metcalfe and Summit buildings

Site	Description	Size (Square feet)	Number of fixtures	Fixture density (Ft ² /fixture)
Metcalfe Federal Building (Chicago, IL)	Large open office areas with assorted private offices and conference rooms	19,750	260	76
Summit Federal Building (Atlanta, GA)	Large open office areas with few private offices and conference rooms	12,900	135	94

Table 5 Annual lighting energy use intensities (EUI) pre- and postretrofit

	Metcalfe	Summit	National average lighting baseline
Preretrofit annual EUI (kWh/ft ² /year)	2.56	1.78	4.1
Postretrofit <i>annual EUI</i> (kWh/ft ² /year)	0.98	1.06	1.02 (weighted avg. of demo sites)
% Savings	62 %	40 %	75 %

communications (VLC) will be used for this purpose). Each fixture can lower its output and reduce electric lighting usage if sufficient daylight is present.

The LED fixtures with integrated sensors and controls were installed in two study areas, detailed below, with all fluorescent fixtures being replaced by LED fixtures on a one-for-one basis. At both demonstration locations, the fixtures were commissioned to the medium institutional tuning setting to provide appropriate light levels while reducing fixture wattage and increasing energy savings (Table 4).

Results

The effect of adaptive controls on lighting energy use is best determined by looking annual lighting energy use intensity (EUI). Lighting EUI, which is a metric of the lighting energy density, is expressed as kWh/ft²/year. By measuring the lighting energy consumption before and after installation, for an adequate period of time, the annual EUI of any installation can be determined. In both the projects, the EUI even before installation was significantly below the National office average of 4.1 kWh/ft²/year (Navigant 2012). Nevertheless, the LLLCs installed at both sites reduced energy use by 62 % and 40 %, at the Metcalfe and Summit sites, respectively. Compared to the National average, both sites reduced EUI by nearly 75 % (Table 5).

Light measurements before and after installation of the LLLCs (Figs. 8 and 9) indicate that the upgrades slightly increased the average illumination levels at these sites.

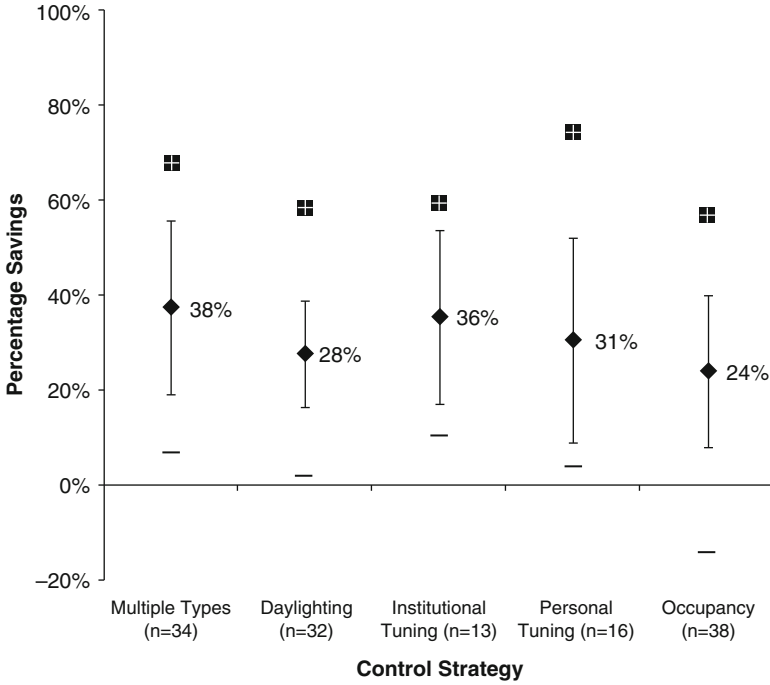


Fig. 8 Percent savings from different control strategies. The mean values are indicated by diamond; the whiskers are standard deviations. The minus symbol is the minimum value, while the square is maximum

Thus, the energy savings from the new lighting did not result in a lowering of light levels. It is remarkable that well designed and controlled LED lighting can provide 40 FC (430 lx) – typical light levels for offices and many other similar environments – with a lighting energy burden of only 1 kWh/ft²/year (Figs. 10 and 11).

By setting the LED “top end” to a value other than full ON, one can accommodate a change in target light level. This is an important benefit of adaptive controls especially for existing buildings where usage may have changed over time. By allowing facilities to “right-size” their light levels on installation, new lighting will not consume any more power than is necessary to achieve the desired target light level. In the demonstrations reported here, the default maximum level was set to 88 % of maximum during the commissioning process.

Separating the Effect of Adaptive Controls from the Light Source Change

It is important to differentiate between the energy savings due to changing the light source from fluorescent to LED and the savings due to the advanced controls features of the retrofit lighting system. The LED fixtures are a higher-efficacy,

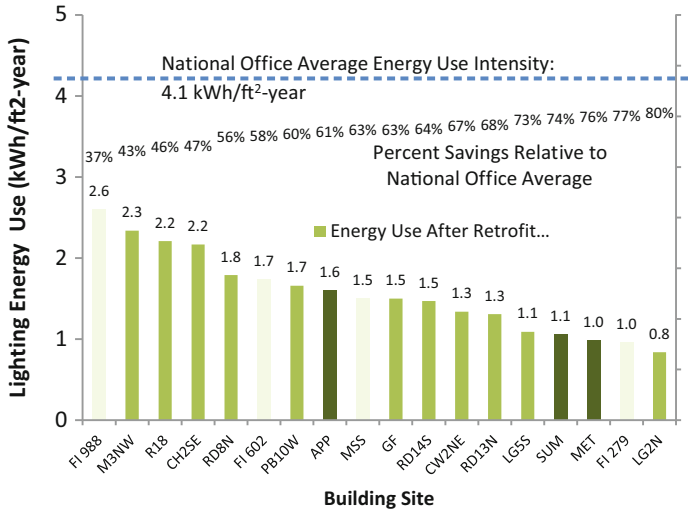
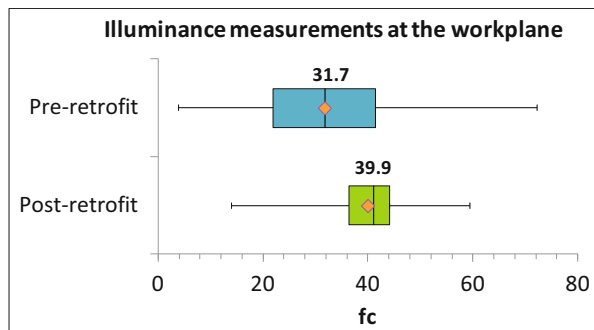


Fig. 9 Postinstallation lighting energy use intensities for 18 demonstration installations of adaptive controls in office settings. Sites where existing lighting were replaced with suspended fluorescent luminaires with Luminaire Level Lighting Controls (*medium green bars*). Sites with area-based controls are indicated by *pale green bars*. Sites with LLLCs in LED troffers are indicated by *dark green bars*. The percent energy savings are calculated relative to the National Office average of 4.1 kWh/ft²/year

Fig. 10 Illuminance measurements pre- and postretrofit at Metcalfe demonstration. Second and third quartiles are enclosed by *box*. *Whiskers* indicate minimum and maxima of data. Mean illuminance indicated with *orange diamonds*



lower-wattage light source, and the retrofit system can be tuned to a lower maximum output, depending on the needs of the space.

At Metcalfe, energy savings of 16 % were achieved by switching from fluorescents to the LED fixtures at full power (Fig. 12). With the LED fixture output commissioned to medium (88 % as described above), and the sensors and controls effecting dynamic dimming throughout the day, 46 % additional energy savings were achieved, for a total of 62 % savings for the entire system. At Summit, the

Fig. 11 Illuminance measurements pre- and postretrofit at Summit demonstration. second and third quartiles are enclosed by box. Whiskers indicate minimum and maxima of data. Mean illuminance indicated with orange diamonds

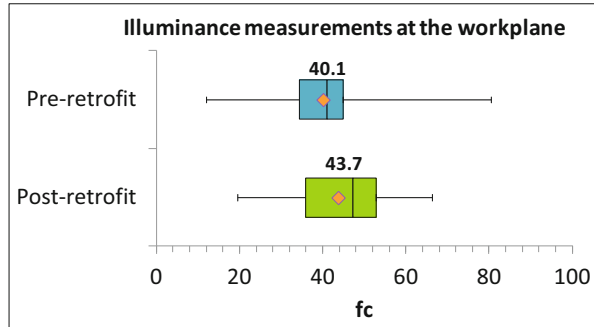
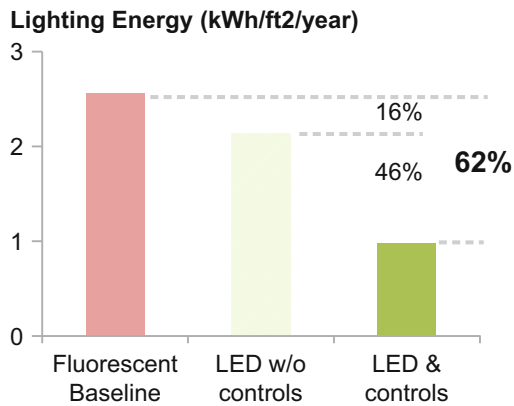


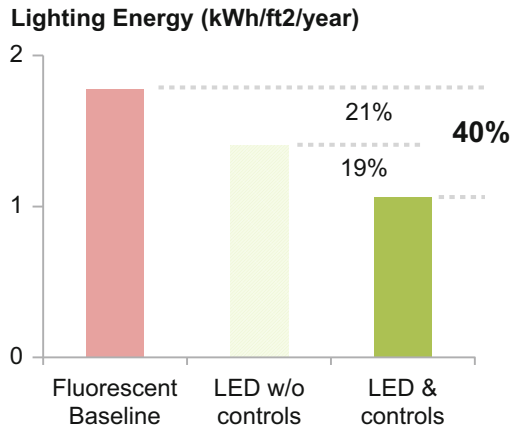
Fig. 12 Annual lighting energy use (Metcalfé bldg) for original fluorescent baseline (leftmost bar), LED replacement fixtures set to full output (middle bar) and LED replacement fixtures with adaptive controls operating (rightmost bar).



lighting operation in the demonstration space was already highly efficient, even before the lighting retrofit (the baseline lighting EUI was very low). The LED LLLC system saved the most energy simply by the change to LED fixtures, at 21 % energy savings (Fig. 13). The benefits of the institutional tuning, occupancy sensor dimming and shut-off, and daylight dimming contributed around 19 % energy savings, for a total of 40 % savings for the system.

These results show that LED lighting combined with adaptive controls applied at the individual fixture level can provide about 400 lx of illumination for a typical office application while using only about 1 kWh/ft²/year of lighting energy. With lighting energy use at only 1 kWh/ft²/year, it is feasible to implement net zero energy buildings in the near term.

Fig. 13 Annual lighting energy use (Summit building) for original fluorescent baseline (*leftmost bar*), LED replacement fixtures set to full output (*middle bar*), and LED replacement fixtures with adaptive controls operating (*rightmost bar*)



References

- DiLouie C Control wiring: a primer, Lighting Controls Association. <http://lightingcontrolsassociation.org/control-wiring-a-primer/>. Accessed 12 July 2015
- DOE Energy Information Administration (2015) Commercial building energy consumption survey (CBECS) 2012
- EMCOR Energy Services (2012) LED office lighting and advanced lighting control system (ALCS). Pacific gas and electric (PG&E) emerging technologies program, Project Number: ET11PGE3251. Project Manager: Jeff Beresini
- Illuminating Engineering Society of North America (2011) Lighting control protocols, TM-23
- Jordan Shackelford J, Robinson A, Rubinstein F LED fixtures with integrated sensors and controls, Final Report to GSA
- Navigant (2012) U.S. Lighting market characterization, volume I: national inventory and energy consumption estimate. Prepared by Navigant Consulting, Inc. for the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Washington, DC
- Navigant Consulting Luminaire level lighting controls market baseline, Report #E14-301, December 2014
- Piette MA, Ghatikar G, Kiliccote S, Watson DS, Koch E, Hennage D (2009a) Design and operation of an open, interoperable automated demand response infrastructure for commercial buildings. *J Comput Sci Inf Eng* 9(2):021004
- Piette MA, Ghatikar G, Kiliccote S, Koch E, Hennage D, Palinsky P, McParland C (2009b) Open automated demand response communications specification (version 1.0), LBNL-1779E
- Rubinstein F, Wei J, Enscoe, A Saving energy with advanced lighting controls: a study of lighting retrofits in 10 federal building offices, Final Report to GSA
- Shackelford J, Robinson A, Rubinstein F (2014) Retrofit demonstration of LED fixtures with integrated sensors and controls, Final Report to GSA
- Wei J, Rubinstein F, Robinson A, Enscoe A, Levi M (2014) Energy savings from advanced lighting control retrofits in federal office buildings. *Leukos* x.x
- Wei J, Rubinstein F, Shackelford J, Robinson A (2014) Advanced wireless lighting controls retrofit demonstration, Final Report to GSA
- Williams A, Atkinson B, Garbesi K, Page E, Rubinstein F (2012) Lighting controls in commercial buildings. *Leukos* 8(3):161–180

Ambient Light Sensor Integration

Frangiskos V. Topalis and Lambros T. Doulos

Contents

Introduction	608
Photosensor (ALS) as Part of a Daylight-Responsive System and Lumen Maintenance Control Strategy	609
ALS as Part of a Daylight-Responsive System	609
ALS as Part of Lumen Maintenance Control Strategy	611
How a Photosensor (ALS) Works with Regard to Integration	612
Main Issues of the Utilization of an ALS	612
How a Photosensor Works	613
Positioning	614
Ceiling-to-Working Plane Illuminance Ratio	616
Switching or Dimming Control Technique	617
Spatial Response	618
Spectral Response	620
Control Algorithms	622
Commissioning of Photosensor (ALS)	624
Reaction of Occupants	629
CCD Imaging Sensors as ALSs	630
CCD Sensor Integration	630
Limitations and Advantages of the Integration of CCD Sensors	631
References	632

Abstract

While daylight harvesting is one of the most energy efficient ways to minimize energy consumption in areas with adequate nature light inside buildings, Ambient Light Sensors (ALSs) are still not being widely installed. Many drawbacks on installation and commission of ALSs, are holding back their wide spread use in building sector. Most of the drawbacks are in relation with their position, field of

F.V. Topalis • L.T. Doulos (✉)

Laboratory of Lighting, National Technical University of Athens, Athens, Greece

e-mail: ftv@central.ntua.gr; ldoulos@mail.ntua.gr

view, spectral response, control algorithms, commissioning and the associated user's response. Moreover, with the advent of LED technology, lighting designers and engineers are focusing only in LED luminaires. Thus frequently ALSs are excluded from the planning, despite their even greater energy saving potential, not only in new installations but also during building retrofitting. The promising coming of CCD or CMOS image sensors can show some promise, but it is not clear yet if they can be widely incorporated in the building sector. Scope of this paragraph is the understanding of the nature of an ALS and its proper integration to buildings for their wide spread use.

Introduction

In the perimeter of buildings, part of desired illumination can often be supplied by daylight via their windows or skylights. In these areas, ambient light sensors (ALSs) or else photosensors force ballasts in luminaires to reduce power for the electric lighting in response to the amount of available daylight, and thus, the consumption of energy is decreased. For the successful application of this strategy, high levels of daylight must be present and an optimum installation, setting, and commissioning of the photosensor (ALS) must be performed by experts. Thus, there will be always sufficient illumination for the tasks. In offices, schools, and other spaces, where critical visual tasks are performed and disturbances must be minimized, continuous dimming is preferable to step or on-off techniques. ALSs except their categorization concerning their communication protocol can be categorized according to their integration, meaning the size of the lighting system they control. There are standalone ALSs for individual luminaires and sensors for larger light management systems. The difference between these two categories concerns mainly their connection circuit. Standalone ALSs are designed to be connected directly to a corresponding ballast (usually the ballast and the ALS are from the same manufacture), while the other ALSs are designed to serve as input to an individual lighting control system (ALS and other lighting control components communicate with the same protocol). However, the main commissioning principals of all ALSs remain the same (Figs. 1, 2, and 3).

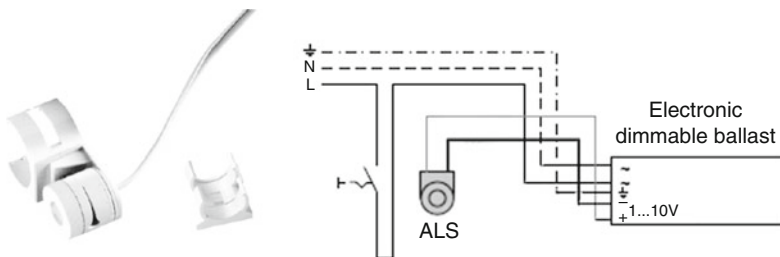


Fig. 1 Standalone ALS and diagram of connection circuit using 1–10 V signal (OSRAM)

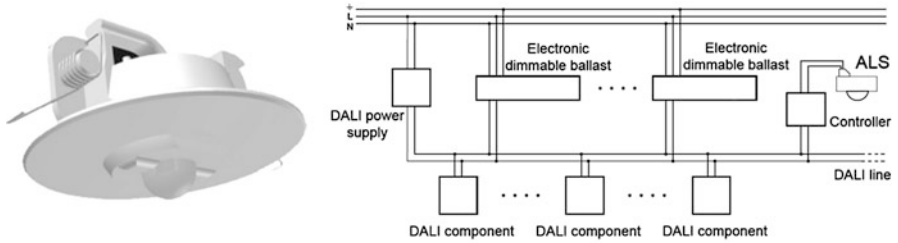


Fig. 2 ALS for larger light management systems and diagram of connection circuit using DALI signal (OSRAM)



Fig. 3 ALS connected directly to the ballast for individual luminaires (*left*) and ALS connected to lighting control system for a group of luminaires (*right*)

Photosensor (ALS) as Part of a Daylight-Responsive System and Lumen Maintenance Control Strategy

ALS as Part of a Daylight-Responsive System

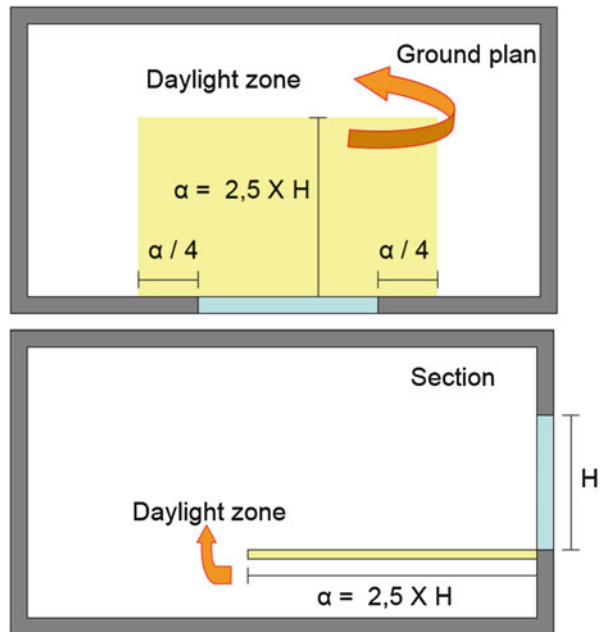
Daylight responsive dimming systems controlled by photosensor (ALS) adjust the artificial lighting level according to the amount of daylight impinging on the ALS. They consist of three basic components: sensor (ALS), lighting controller, and dimming unit. Sensor detects luminous flux and converts it to a signal that is sent to the controller. The controller processes this signal and defines the desired dimming level. After that, the controller sends a dimming control signal to the electronic ballasts forcing them to reduce power.

A typical daylighting control concept usually consists of integrated lighting control zones equipped with one or more than one ALSs. The integrated lighting control zones are areas in the building that use jointly daylight and artificial lighting to provide task area, background, or general illuminance. The size and shape of a

Table 1 Use of ALS and daylight penetration as function the daylight factor (European Standard EN 15193 2007)

Daylight factor (DF)	Access of the zone to daylight and use of ALS
$DF \geq 3\%$	Strong
$3\% > DF \geq 2\%$	Medium
$2\% > DF \geq 1\%$	Weak
$1\% > DF$	None

Fig. 4 The daylight zone with regard to a vertical external opening using European Standard EN 15193 (2007), energy performance of buildings, and energy requirements for lighting



daylight zone depend upon aperture configuration, sky condition, and solar location. These zones can be defined using either daylight factor classification (Table 1) and simulations or standards like European Standard EN 15193 (2007), energy performance of buildings, and energy requirements for lighting (Fig. 4). The daylight zones are limited on account of the rapid falloff of horizontal daylight illuminance from the window wall (Fig. 5). On the other hand, the electric lighting supplements daylight as the distance from the window is increased (Fig. 6).

In spite of the accurate definition of daylight zones, there will be no energy saving, unless an ALS is embodied to the daylight-responsive system. Furthermore, the amount of energy savings by exploiting daylight with ALSs depends on many factors such as:

Fig. 5 Distribution of daylight as the distance from the window is increased

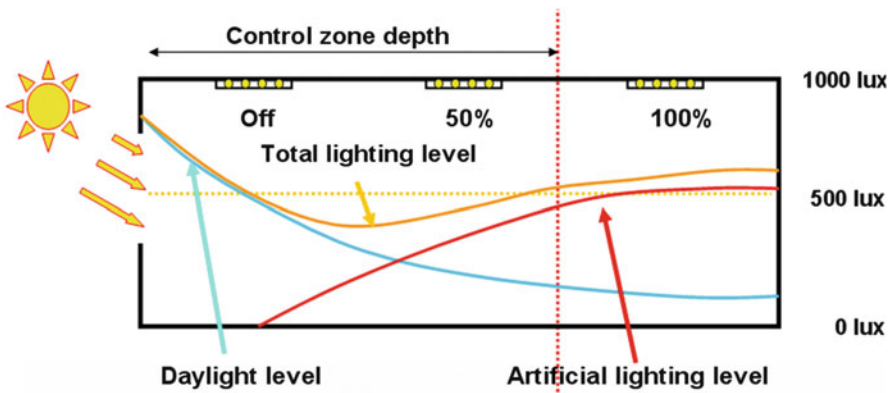


Fig. 6 Distribution of supplementary artificial lighting as the distance from the window is increased

- The climate
- The form, orientation, and design of building
- The activities taking place in the building
- The control algorithm, installation, position, calibration, and commissioning of the photosensor (ALS)

ALS as Part of Lumen Maintenance Control Strategy

Photosensors (ALSs) can also be used for the implementation of a lumen maintenance control strategy. The depreciation of the phosphors of a lamp over time (lumen depreciation of the lamp) and the accumulation of dirt on the luminaires provoke the reduction of the illumination of an area with time. This decrease of the light level can exceed a percentage of 30 % during a 2-year period. Consequently, the installed lighting system must produce an initial illumination higher than the specified level. When the illumination falls under the minimum allowed level, the luminaires must

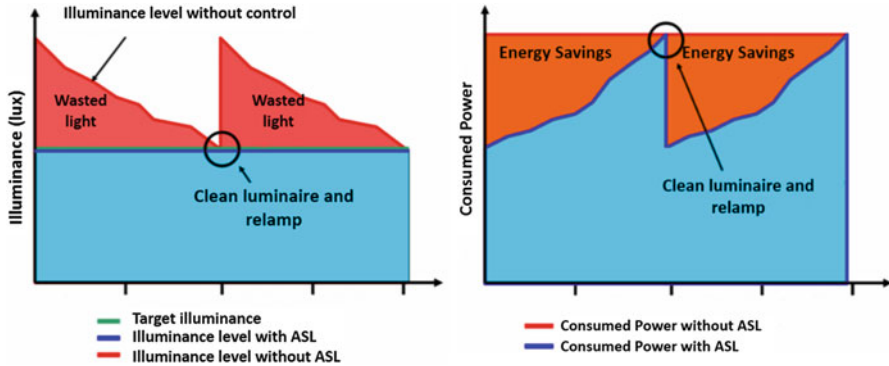


Fig. 7 Distribution of artificial lighting (*left*) and consumed power (*right*) for lumen maintenance control strategy using ALS

be cleaned and relamped. Lumen maintenance control strategy through the use of ALSs reduces the initial illumination to the designed minimum level leading to the reduction of the energy consumption. With time, the illumination is decreased due to lumen depreciation, but thanks to the lumen maintenance control strategy, the electric power is increased so that the reduction of illumination is avoided. Full power is applied only near the end of the lumen maintenance period when the luminaires are cleaned and relamped. Furthermore, lumen maintenance and daylight control strategy can be successfully implemented simultaneously with the same ALS (Rubinstein et al. 1993) (Fig. 7).

How a Photosensor (ALS) Works with Regard to Integration

Main Issues of the Utilization of an ALS

The basic operation of a photosensor (ALS) is the production of a signal that is related to the amount and the distribution of lighting in the space in which it is placed. The performance and integration of the photosensor (ALS) can be complex because it depends on a lot of variables (Fig. 8), such as:

- The ambient light level, the distribution of daylight, and the artificial lighting in the space in which it is placed
- The spectral composition of lighting
- The adjustment settings of the commissioning control
- The technical specifications of the ALS (field of view, spectral response, etc.)
- The geometry data of the room (size, external openings, location of positioning, etc.)

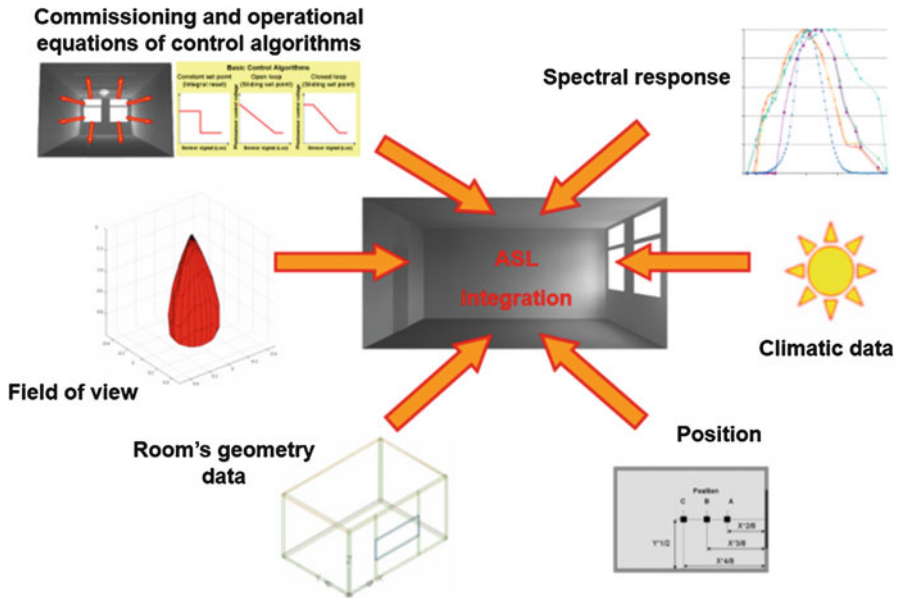


Fig. 8 Parameters affecting the performance of an ALS

The behavior of the photosensor (ALS) in response to variable lighting and blind and/or glazing conditions is significant when it comes to the design of a daylight-responsive system.

The very first steps in order to avoid erratic behavior of photosensors (ALSs) are to recognize the drawbacks and identify their causes. Some of the main parameters that can cause a problematic performance are the following (Doulos et al. 2013):

- The positioning of the photosensors (ALSs)
- The field of view (FOV) of the ALSs
- The spectral response of the ALSs
- The corresponding algorithms
- Occupants' reaction
- The lack of a common commissioning procedure

The main drawbacks due to these reasons are described in next paragraphs.

How a Photosensor Works

The photosensor is a complete control unit that contains a light-sensitive photocell, input optics, an electronic circuit needed to convert the photocell signal into an output control signal, and a housing and mounting device. The complex function of the photosensor (ALS) can be divided into smaller and more flexible subfunctions.

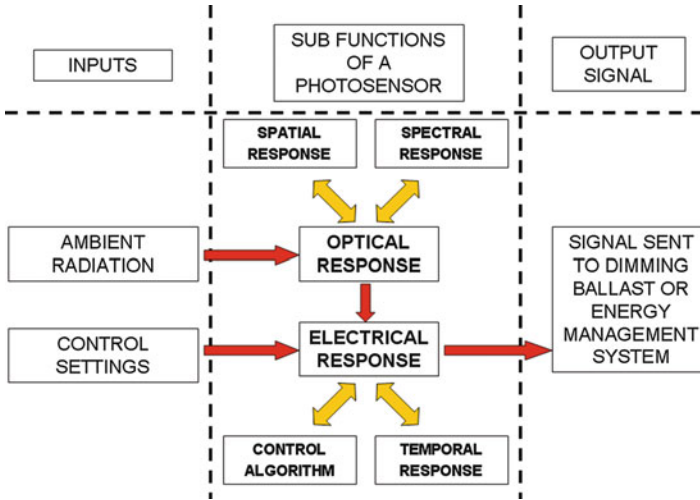


Fig. 9 System diagram of a photosensor (ALS) (Doulos et al. 2008b)

The total response of the photosensor (ALS) is the combination of the responses of each subfunction. The division of the basic subfunctions is based on the optical and electrical characteristics of the photosensor (ALS) (Fig. 9).

The incident radiation that is collected from the environment, affected by the optical subfunctions, is driven to the photocell (spatial response) and then is filtered in order to have its wavelength limited. As a result, the incident radiation can be related to photometric quantities (spectral response) and then converted into an electric signal (control algorithms) (Bierman and Conway 2000).

Positioning

After determining the daylight zones, it is crucial to select the position of the ALS which is the first step in the design process. The ALS must be away from direct sunlight and other sources of stray light; otherwise, it must be partially shielded against direct sunlight (Floyd and Parker 1995; Littlefair and Motin 2000). Furthermore, partially shielded ALS contributes to reducing illuminance fluctuation in the space especially under partly cloudy sky conditions (Kim and Kim 2007). The ALS must not be installed facing directly sunlit surfaces, and when it is positioned on a wall, it must be directed toward a wall that never receives direct sunlight. The ideal location for an ALS would be on the working plane, but such a position is not possible because it is likely for the ALS to be disturbed or shaded by activities in the room. As a result and for practical reasons, ALSs are located on the ceiling in order to minimize any interference from activities in the room, complicating at the same time the control and commissioning of the ALS (Fig. 10). The changes in the ceiling illuminance are not linearly related with the changes in the working plane

Fig. 10 Positioning of ALS on the ceiling in order for the interference from activities in the room to be minimized

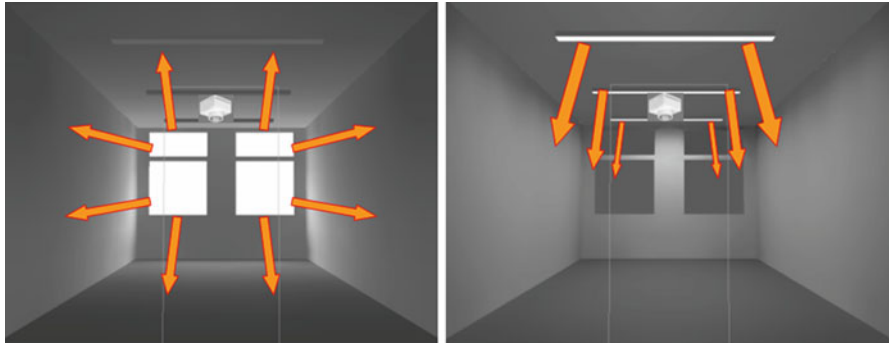


Fig. 11 Daylight (*left*) and artificial lighting (*right*) distribution within a room

illuminance mainly due to the different light distributions in a room between daylight existence conditions and existence of artificial lighting alone (Choi et al. 2005; Fig. 11). Daylight enters a room through vertical external openings in a horizontal or even an upward direction due to the window-embodied daylight systems (such as light shelves, louvers, prismatic panels, etc.) that diffuse and redirect the direct sunlight. Luminaires usually direct artificial lighting down to the working plane. Therefore, when daylight enters the room, the ceiling illuminance increases much more than the working plane illuminance. The correlation of the lighting levels between the working plane and the ceiling is strongly dependent on the position of the photosensor and its field of view (FOV) (Bierman and Conway 2000; Mistrick and Thongtipaya 1997; Mistrick and Sarkar 2005).

Every lighting control project with an ALS is unique and must be studied independently. However, the following rules concerning positioning could be used as rules of thumb (Doulos et al. 2014):

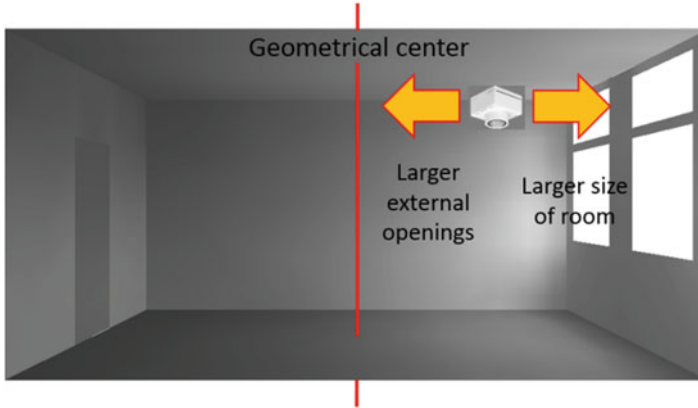
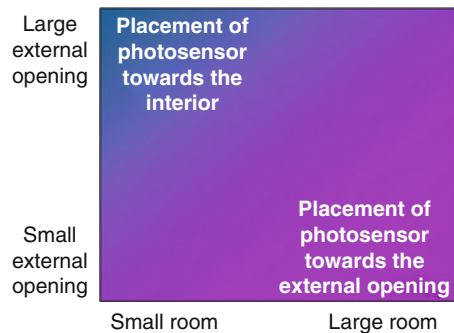


Fig. 12 Placement of an ALS toward the geometrical center or external opening depending on the size of room and the size of the external opening

Fig. 13 Position of an ALS with regard to the size of room and external opening (Doulos et al. 2014)



- In contrast to areas with small external openings and areas with large external openings, ALSs could be placed toward and up to the geometrical center of the area and not close to external openings.
- On the contrary, in areas with larger size as opposed to areas with smaller size, ALSs could be placed toward the external openings.
- The geometrical configurations (square, wide-shallow, or narrow-deep type) of an area do not affect a lot the choice of a specific location (Figs. 12 and 13).

Ceiling-to-Working Plane Illuminance Ratio

The performance of an ALS in relation to the position of the latter on the ceiling is dependent on the ceiling-to-working plane illuminance ratio which is not constant during the operation of the ALS. As already mentioned, ALSs are located for practical reasons on the ceiling of a room in an effort to minimize any interference with the activities taking place in the room. Through the use of proper control and

commissioning, the dimming system tries to maintain a predesigned illuminance on the working plane. The ceiling-placed photosensor corresponds to incident illumination and converts it to a control signal. Furthermore, the selection of proper FOV according to the geometrical characteristics of the space is quite crucial in optimizing the performance of the system and affects also the ceiling-to-working plane illuminance ratio. Different luminance patterns on the room surfaces can create the same sensor signal, resulting in the same dimming state of the lighting system although working plane illuminances differ significantly. In addition, if a shading system is used, any unwanted reflected illuminance can increase the signal of the sensor, reducing the lighting levels. Thus, the position and the FOV of the sensor should ensure a relatively constant ratio of ceiling to working plane illuminances (Choi et al. 2005; Mistrick and Thongtipaya 1997; Mistrick and Sarkar 2005) which of course is strongly dependent on the variability of daylight distribution in the room. The best correlation between these two illuminance values is achieved in areas away from exterior openings where light distribution is uniform. Nevertheless, these areas could not be considered as daylight zones where daylight can be exploited to its maximum extent (Rubinstein et al. 1989; Doulos et al. 2005). On the contrary, when placing the sensor near the windows, a greater energy saving is expected but with a poorer performance in terms of achieving necessary lighting levels and a worse correlation between ceiling and working plane illuminances. For an unshielded – ceiling-placed – ALS, it is quite difficult to track the illuminance changes on the working plane with precision, and for that reason, various shield designs have been proposed by the manufactures. Apart from ceiling-to-working plane illuminance ratio, energy savings and adequacy of lighting are both affected by the selected position of the sensor and by the FOV (Doulos et al. 2013).

Switching or Dimming Control Technique

ALS can be used in a daylight-responsive system using either a switching or a dimming control technique. Switching technique is simple and cost-effective concerning not only its installation but also the purchase of the corresponding equipment (e.g., dimmable ballasts are not necessary for switching). However, frequent switching, probably taking place under partly cloudy conditions, can be annoying to the occupants (Li 2010). For this reason, the selection of the switching technique is dependent mainly on the amount of the available daylight. Thus, this technique must be used in specific areas, where daylight levels are high for the most of daytime, and therefore, artificial lighting is off most of the time and where occupants are passers-by or their tasks are not critical (Table 2). However, there are control algorithms, such as differential switching, time delay and solar reset, which can reduce the switching off frequency (Littlefair 2001).

Using dimming control technique, artificial lighting in each daylight zone can be varied smoothly and continuously to dynamically match visual requirements in accordance with the daylight levels. Dimming components have higher costs and commissioning procedure is more elaborate. However, dimming control is best

Table 2 Recommended type of ALS control technique for different building applications

Type of area with external openings	Switching	Dimming
Circulation areas	S	M
Classrooms	M	S
Conference rooms	W	S
Corridors	S	M
Entrance halls	S	M
Exterior light	S	W
Hallways	S	W
Laboratories	W	S
Lecture halls	W	S
Libraries (bookshelves)	M	S
Libraries (reading area)	W	S
Museums	W	S
Offices	W	S
Reception	W	S
Restrooms	S	W
Retail areas	W	M
Street light	S	W
Waiting rooms	W	S
Warehouses	S	M

S strong application, *M* medium application, *W* weak or inexistent application

suitable in places with permanent occupants or occupants that perform critical tasks. When daylight levels are sufficient and artificial light output is at the lowest level, the percentage of the energy consumption of the artificial lighting system with dimming control abilities is between 15 % and 20 % of the total energy consumption of the system depending on the ballast of its manufacturer (Doulos et al. 2008a). The same lighting system with switching control abilities is more efficient when the lights are off. However, on occasions of partly cloudy sky, the dimming system is more effective than switching one concerning not only energy saving but also the occupants' satisfaction (Li et al. 2010).

Spatial Response

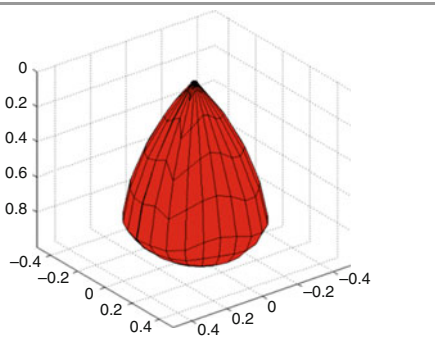
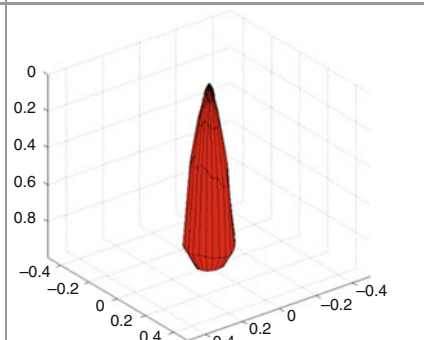
The spatial response describes the sensitivity of an ALS to incident radiation from different directions, in other words, what the photosensor “sees.” The spatial response is similar with a luminaire intensity distribution but describes sensitivity instead of light output. The field of view (FOV) and the commissioning of the photosensor can influence considerably its performance and as a consequence the operation of a daylight-responsive system (Bierman and Conway 2000).

ALSs with narrow spatial sensitivities track more efficiently the variations of illuminance in the working plane provided that the reflectance properties of the surface do not change. On the other hand, their small FOV is not representative of the

entire working plane. In the real world, the reflectance of the working plane changes depending on the activities occurring in the room. For instance, white papers on a dark desktop or the daily different clothing of the occupants can change the response of a narrow ALS. This can make a narrow ALS very sensitive causing it to function unpredictably and thus influencing negatively the occupants’ optical comfort by changing rapidly the light levels. This effect can be intensified by specular reflections from glossy mirror-like surfaces. Reflections of light directly back to the photosensor’s FOV can cause erratic performance.

The specular reflections have relatively less effect to an ALS with wide spatial response. A large FOV diminishes this kind of effects. The main advantage of ALSs with wide spatial response is that the optical signal they detect is very representative of the overall working plane or the entire room and that it is less affected by normal activity in the room. On the other hand, the illuminance of the ceiling does not correspond to the illuminance of the working plane, and therefore, this kind of photosensor needs a suitable control algorithm in order to compensate for this noncorrespondence in illuminance levels (Bierman and Conway 2000; Table 3).

Table 3 Advantages and limitations of using ALSs with wide or narrow FOV

Wide FOV		Narrow FOV	
			
Advantages	Limitations	Advantages	Limitations
Signal of ALS is representative of the overall work plane or room	Ceiling illuminance values do not correspond precisely to working plane illuminance	ALS corresponds more to luminance values of working plane	Increased sensitivity to specular reflections from shiny, mirror-like surfaces
Less sensitivity to mirror-like, specular reflections from shiny surfaces	Ceiling-to-working plane illuminance ratio differs significantly as the distribution of daylight changes	The narrower the response, the better the tracking	Reflections of light directly back to the ALS’s field of view that can lead to erratic performance
Less affected by normal activity in the room			ALS is very sensitive to changes in the reflectance properties of the aiming surface

As mentioned before, each lighting control project with ALS is unique and must be studied separately. Nevertheless, there are some rules of thumb about the FOV selection that could be used (Doulos et al. 2014):

- ALSs with wider FOV could be used in rooms with external openings facing north (for northern hemisphere). On the other hand, the selection of FOV in rooms facing south is more complicated, mainly due to daylight distribution in these rooms, and therefore, accessional attention at the design stage is demanded.
- ALSs with wider FOV could be used in rooms with large external openings.
- ALSs with wider FOV could be used in rooms with large size.
- The geometrical configurations (square, wide-shallow, or narrow-deep type) of an area do not affect much the selection of the FOV of an ALS. However, in wide-shallow areas, mostly ALSs with medium FOV could be used, while in narrow-deep areas, ALSs with narrow FOV (Table 4, Fig. 14).

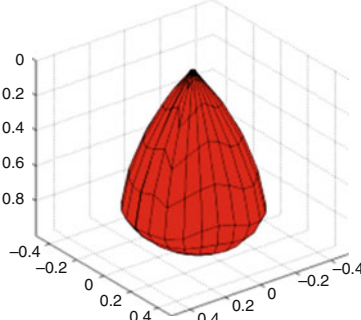
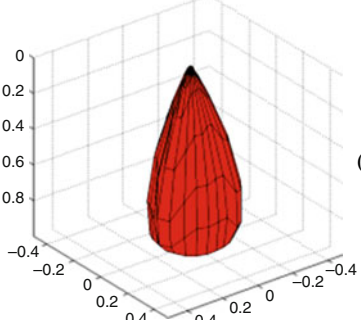
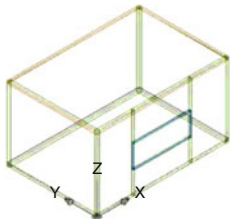
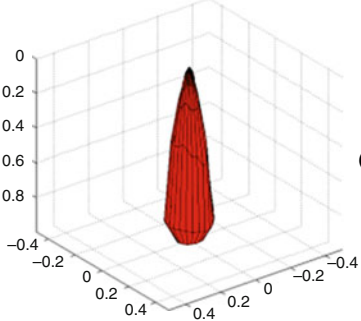
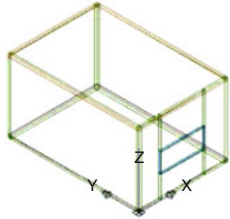
Spectral Response

The spectral response describes the sensitivity of the photosensor (ALS) to optical radiation of different wavelengths. The photocells used in ALSs are sensitive to a wider range of wavelengths than the ones the human eye sees. More specifically, photocells respond to parts of the ultraviolet (UV) and infrared (IR) spectrum as well as to the visible spectrum. Filters incorporated into the photocell housing limit the sensitivity to ultraviolet (UV) and infrared (IR) radiation (Bierman and Conway 2000).

The broader spectral response of ALSs makes them respond differently than they would if they had an ideal Commission Internationale de l'Éclairage (CIE) photopic luminous efficiency function $V_{(\lambda)}$ which actually coincides with what the human eye sees. For example, the differences in spectral composition between daylight and fluorescent lighting can affect differently the photosensor (ALS). Furthermore, the greater amount of ultraviolet and infrared radiation of daylight combined with the wider spectral response of ALSs makes them more sensitive to daylight than to artificial lighting. Greater sensitivity to daylight means that an ALS reacts as if there was more daylight than what there really is. What is more is that algorithms that use only nighttime settings (e.g., integral reset) are incapable of adjusting the lighting levels above the target illuminance for a significant part of the operation time. As a result, the ALS decreases the lighting levels of the artificial lighting system more than enough and provokes visual discomfort to the occupants.

Furthermore, different spectral sensitivities of ALSs combined with different glazing transmittance lead to different amounts of light perceived by the ALS. These differences change in a systematic way so that it is possible to be compensated by the ALS control algorithm. Both open-loop and closed-loop proportional control algorithms can perform this compensation. However, the design procedure of the lighting control system (especially when simulations are used) and the energy saving

Table 4 Selection of FOV

Type of FOV	Application
 <p>(wide FOV)</p>	<p>Rooms facing north Rooms with large external openings Rooms with large size</p>
 <p>(medium FOV)</p>	<p>Wide-shallow rooms</p> 
 <p>(narrow FOV)</p>	<p>Narrow-deep rooms</p> 

calculations require the knowledge of the exact amount of lighting that a photosensor can “see.” Although manufacturers of ALSs through their control algorithm provide a correction of this difference, in simulation tools, the $V(\lambda)$ function is used as the default spectral response for ALSs (Fig. 15).

Fig. 14 Selection of field of view (*FOV*) with regard to the size of room and the size of the external opening (Doulos et al. 2014)

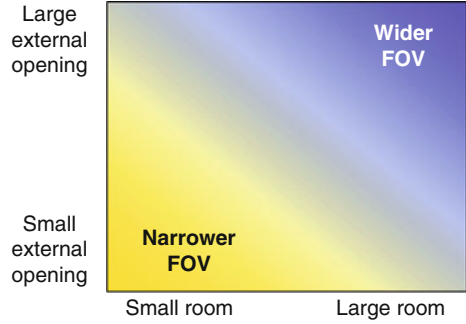
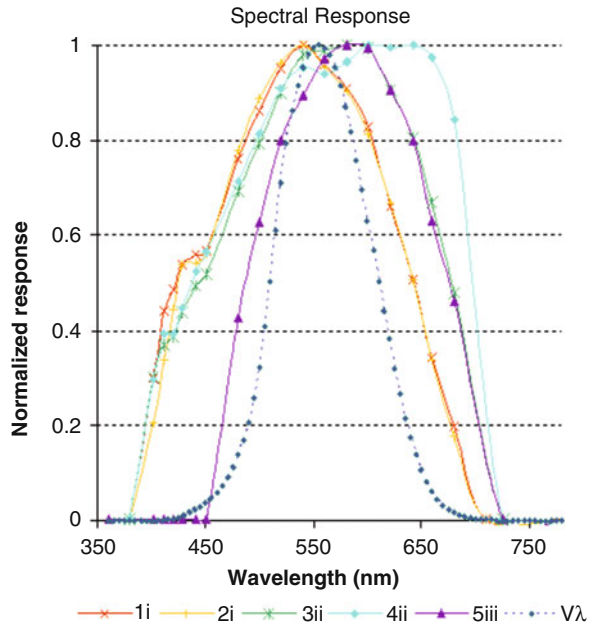


Fig. 15 Spectral response of five tested photosensors in comparison with the CIE photopic luminous efficiency function $V(\lambda)$ (Doulos et al. 2008b)



Control Algorithms

When specifying an ALS control system, the most important characteristic is the control algorithm, and it should be examined in the first place. The control algorithm describes precisely the exact output signal of the photosensor (ALS) as a function of the input data. The existing ALSs use usually one or a combination of three basic control algorithms (Rubinstein et al. 1989):

- Closed loop (for indoor ALS)
- Open loop (for outdoor ALS)
- Integral reset (for indoor ALS)

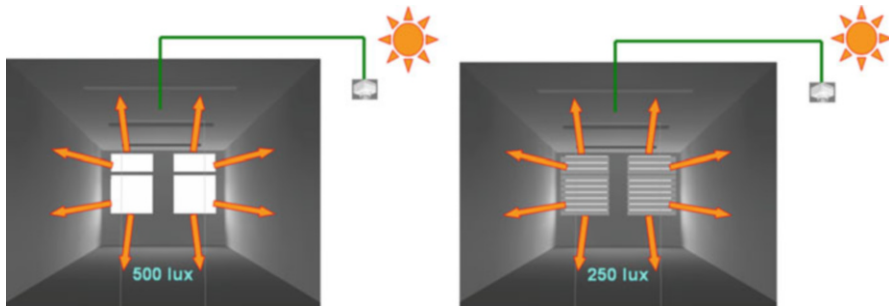


Fig. 16 A limitation of an ALS that uses open-loop algorithm is that it cannot adjust to changing light distribution, for example, due to the closing of the blinds

Usually, the control algorithm plays a dominant role in the performance of the daylight-responsive dimming system. A high-performance control algorithm could possibly, but not likely, compensate for both a nonideal spatial response and a poor spectral correction. However, the more sensitive the photosensor is to variations of the illuminance on the working plane and the better the spectral correction of the ALS, the finer the overall operation of the ALS system will be.

ALSs that use open-loop algorithm are designed to be placed in locations away from the areas they control, and they therefore do not perceive directly the corresponding lighting levels. Thus, although they are placed mostly on the outside of the building (or aim at external openings), they are able to control the interior lighting in multiple zones or in the entire building, thereby reducing the cost. However, and as a result, when various changes concerning daylight or artificial lighting take place inside the building, such as closing of blinds and manual switching, the control system cannot compensate for these actions for parts of the building or for the whole of it (Lee and Selkowitz 2006; Fig. 16).

Thus, artificial lighting still remains at the minimum light output, while daylight levels inside the building decrease, and as a result, the total lighting levels are below the target illuminance level.

The choice of the control algorithm should be based not only on the achieved energy saving but also on its ability to maintain target illuminance. The visual comfort, determined by the number of hours during which the total illuminance exceeds the target one, must be at high values. A good example for this could be the comparison between closed-loop and integral reset algorithms. The integral reset control performs poorly for a remarkable number of hours, since it dims out the lights despite the fact that the daylight levels are inadequate. The integral reset controller maintains a constant value on the sensor located on the ceiling. This results in progressively lower total illuminance levels on the working plane as daylight increases, making this type of controller unsuitable for daylight applications (Doulos et al. 2013; Fig. 17, Table 5).

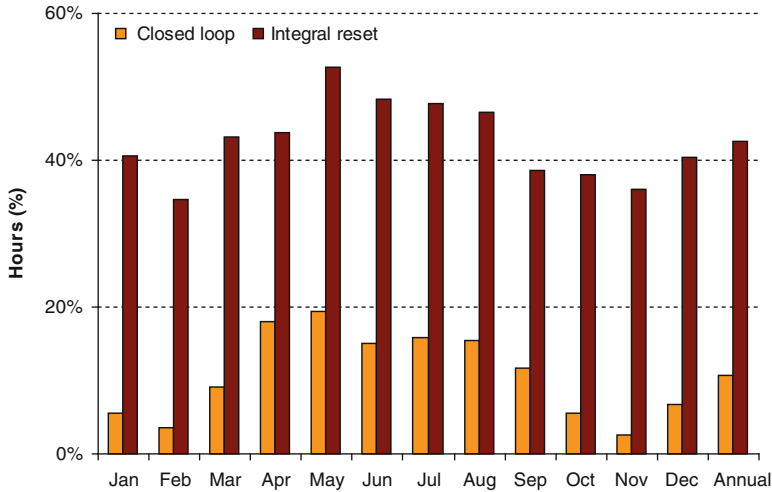


Fig. 17 Monthly percentage of operational hours (07:00–18:00) during which the achieved illuminance is lower than the target illuminance (500 lx) for two examined control algorithms (closed loop and integral reset). The light levels are below the target illuminance level for 10.74 % of the yearly operational hours for the closed-loop algorithm and for 42.60 % for the integral reset algorithm (Doulos et al. 2008a)

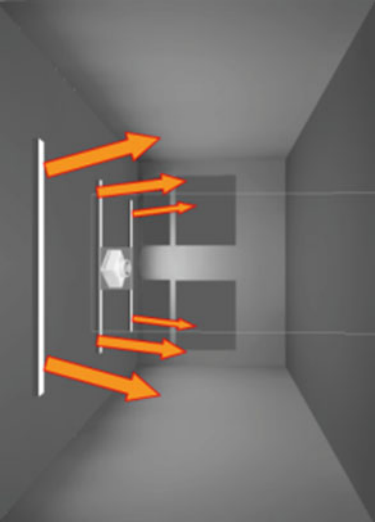
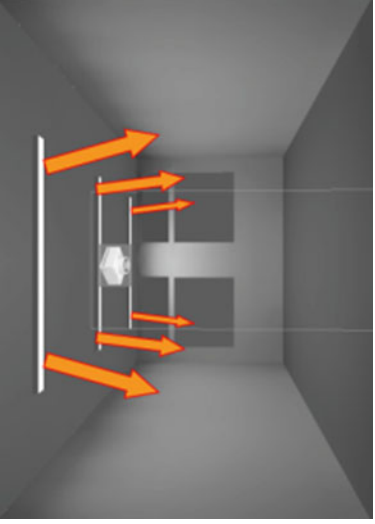
Commissioning of Photosensor (ALS)

Plenty of ALSs are currently available, but there is still lack of information concerning not only the actual performance of these systems but also their commissioning. In order to fully exploit their capabilities and implement the most energy-efficient daylight harvesting, a proper commissioning procedure and detailed guidelines must be provided mainly by the manufacturers. It is necessary for the experts of commissioning to follow the same methodologies for all ALSs placed on the ceiling.

Usually, ALSs are not properly commissioned. The FOV and the position of the sensor are significant parameters for the photosensor commissioning and the proper function of a daylight-responsive system. The optimum position and spatial response of the photosensor are directly related to the satisfaction of the following three criteria, which can improve the commissioning procedure of a daylight-responsive system. These criteria are:

- A reliable determination of the ratio between illuminance on the photosensor and working plane illuminance
- The energy savings achieved
- Illuminance values being above a specific design illuminance value (lighting adequacy)

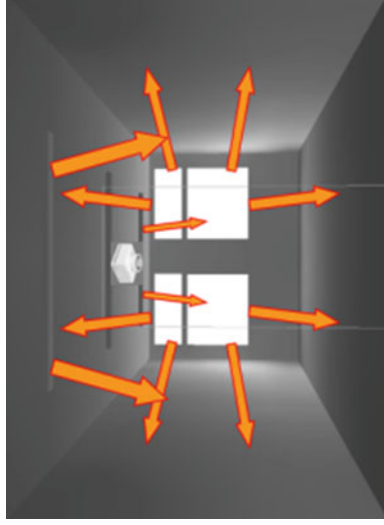
Table 5 Differences concerning the integration procedure between integral reset and closed-loop algorithms

<p>Integral reset algorithm</p>	<p>Closed-loop algorithm</p>
<p>Single parameter adjustment for the maintained photosensor signal level</p>	<p>At least two independent parameters adjustments, one for the maintained lighting level and another one for setting ratio of the photosensor optical signal (between daylight and artificial light)</p>
<p>Requires nighttime commissioning (or with opaque window blinds in order to exclude daylight)</p>	<p>Requires nighttime commissioning (or with opaque window blinds in order to exclude daylight)</p>
<p></p>	<p></p>
<p>Causes excessive dimming of the artificial lighting in the presence of daylight</p>	<p>Requires day-time commissioning</p>

(continued)

Table 5 (continued)

Integral reset algorithm	Closed-loop algorithm
Cannot compensate for an ALS's wider spectral response than the spectral response of the human eye (CIE photopic luminous efficiency function $V(\lambda)$)	Can compensate for an ALS's wider spectral response than the spectral response of the human eye (CIE photopic luminous efficiency function $V(\lambda)$)
Appropriate for lumen maintenance control	Appropriate for daylight control



Commissioning of the daylight-responsive system with ALS is a systematic procedure in order to set all system components and verify that they operate interactively and continuously according to the initial specifications. Thus, installation and commissioning of an ALS are difficult and time-consuming tasks which must be performed by qualified persons. Following the rules (the ones mentioned below) cannot substitute the work of an expert, who will carry out the commissioning of the system but could possibly improve his expertise. The factors that could affect the operation of an ALS are the ones discussed above. Empirical placement of ALS could lead to erratic performance of the control system. Proper commissioning of the daylight-responsive system is important since in a flawed system, the reactions of the occupants can lead to the disconnection of the ALS, eliminating, in this way, the opportunity for energy saving and for the payback of the initial investment for buying and installing the ALSs.

The basic commissioning procedure includes the setup of the equipment, calibration, input necessary for the setting of the corresponding control algorithms, fine tuning concerning the selection of position, FOV, daylight conditions for the system calibration, etc. This procedure may differ from system to system, but there are some simple and common steps. The first step is to verify that the right equipment is properly installed and is functioning. The following rules for commissioning can be used as rules of thumb, and they do not constitute principles for every ALS (Kim and Mistrick 2001; Lee et al. 1999; Rubinstein et al. 1997):

- A commissioning expert must be aware of the ALS technical specifications and the requirements of the manufacturers. For example, an expert must be able to understand when the sensitivity range of ALS has surpassed the specified limits (Floyd and Parker 1995). ALSs with integral reset control algorithm should be avoided in daylight-responsive systems. The expert must also be aware of the use of the area and the schedule of the occupants.
- A commissioning expert must follow the exact instructions (from the beginning of the planning of the lighting control system) concerning the optimum position of the ALS and must determine each control zone of the ALS.
- In case of a specific mounting position, a commissioning expert must select an ALS with the appropriate FOV (narrow, medium or wide) or adjust the appropriate FOV if possible.
- The parameter setting and calibration of the ALS (daytime and nighttime setting) is executed by the expert after the installation of the sensor. The expert must set the sensitive adjusters of the sensor so that he is as far as possible from the FOV and he cannot affect it. The person in charge for the adjustment should not be close to the ALS. The incident radiation to the sensor must not be prevented from the body of the person. To alleviate this problem, the expert must be at least on the opposite side of the external openings and not between the FOV and them. In some ALS systems, the control device is located away from the FOV. The problem can be avoided when the settings are electronic which means that they can be carried out remotely.

- Proper system setup requires a calibrated photometer. If possible, the cord between the recorder/meter and the measurement sensor must be long so that there is enough distance between the point at which the reading of light levels is performed and the point where the setting of the sensor is performed.
- In the first stage, the expert must adjust the illuminance level at a value that must be kept constant (night setting). Night setting is performed by adjusting the artificial lighting without the presence of daylight. This adjustment is usually carried out during the night. If the room has opaque shades that seal off the external openings (no daylight enters in the room), the adjustment of artificial lighting could be performed even during daytime.
- In the second stage, the expert must adjust the ratio between the signal of the ALS and the desired light levels to the presence of daylight and artificial lighting (daytime setting). This is accomplished by artificial lighting dimming that leads to the achievement of the desired level of the total illumination in the presence of daylight. The desired total lighting level may be equal to the illuminance level maintained as it was initially configured (only artificial lighting) or may be larger, allowing the increase of the overall lighting level as daylight enters the room.
- The ratio of daylight level on the working plane to daylight level on the sensor, ($I_D(t)/S_D(t)$, Rubinstein et al. 1989) is an important factor affecting the performance of a daylight-responsive system. The choice of the right calibration time and the proper location and configuration of the photosensor is critical in order to reduce the differences among various sky types and seasons (Choi and Mistrick 1997, 1999). The most frequent value of this ratio over a year is usually selected to be the conditional expression of the closed-loop or open-loop control algorithm ($I_D(t_{cal})/S_D(t_{cal})$) (Rubinstein et al. 1989). The daytime commissioning can be completed only when the commissioning is done on the day where both values $I_D(t)$ and $S_D(t)$ do not exceed a reference lighting level (e.g., 500 lx for office lighting planning between morning hours and 11:00 LT). In order to carry out daytime and nighttime calibration almost the same time (during day), without blocking the daylight, the signal of the ALS and the corresponding lighting levels should be received (a) with the simultaneous presence of daylight and artificial lighting and (b) only with the presence of daylight. By subtracting the two readings of the lighting levels, the signal of the ALS could be determined by the impact of artificial lighting only.
- The daytime commissioning of the ALS is dependent on the distribution of daylight at the time of the commissioning of the sensor. If these conditions are not chosen properly, the results and performance of the daylight-responsive system will not be the desirable ones (Mistrick et al. 2000). The setting must be done during the operation of the ALS and the lighting system and only when the distribution of daylight is representative. Daylight conditions must be stable (clear sunny or overcast days). There are certain values of daylight distributions that must be avoided. For example, partly cloudy conditions produce significant variations in daylight on a minute-to-minute basis.
- During the commissioning of the ALS, there are two things that must be avoided: (a) the penetration of direct solar radiation, that is, when the sun is at low solar

heights (e.g., late in the afternoon), and (b) strong direct sunlight. High values of direct solar radiation should not be selected because it is likely that this radiation is blocked by the shades of the external openings.

- If in the installation site there exist adjustable shades, the installation and commissioning of the ALS should be performed only after the occupants have selected the location of the blinds. This procedure takes place when there is no redirection of sunlight to the ALS or when the place of the blinds is not bothering for the occupants. Under no circumstances should the position of blinds impede the optimal functioning of the ALS.
- The ALS must respond to changes in the intensity of the daylight with a time lag (e.g., more than a minute in order for the lighting system to be at its maximum light output).
- The sensor should never be pointing at the external opening unless it is permitted by the specifications of the manufacturer.

Reaction of Occupants

The occupants' satisfaction must be the first priority when a daylight-responsive system is designed. When the occupants are not satisfied, they will most probably override the system. A very quick dimming of artificial lighting (usually during partly cloudy sky conditions) and inadequate lighting levels due to incorrect commissioning are the most common reasons for optical discomfort and thus dissatisfaction of the occupants. The most common action of the occupants when they wish to override the lighting control system is to tap over the photosensor (ALS). In this way, the sensor cannot perceive any light, forcing luminaires to operate in full lighting output. Furthermore, even if the occupants fail to override the sensor, there might be complaints to the building maintenance staff which will lead to the same result. Usually, the members of the building maintenance staff are not experts in lighting control, so they prefer to disconnect the sensor rather than to commission properly the system. As a result, the occupants are satisfied since they do not experience any unwanted changes in light level. However, both cases will result in energy wasting and therefore a significant loss of money and waste of the initial investment.

Of course, the occupants' reactions should under no circumstances lead to stop installing lighting control systems with ALSs. For this reason, there are many simple actions that can help either to counterbalance or to minimize their reactions (Rubinstein et al. 1997):

- Hiring commissioning experts from the earliest point of the project. It is advisable to plan properly the lighting system rather than to fix it.
- Commissioning of ALSs must be completed as initially planned. Different placements or FOV shall have unexpected results that will most probably cause dissatisfaction to the occupants.

- Building contractors must be well informed on this subject and have the knowledge, the skills, and the proper stuff required to commission the state-of-the-art ALSs.
- Education of the occupants of the building about the purpose and the benefits offered by the embodying of the ALS with the lighting system. Occupants must realize the high importance of the benefits of energy saving and, of course, be always aware of the existence of the lighting control system.
- The maintenance staff must improve their knowledge and skills for the proper recommissioning of the lighting control system. The system must be periodically fine-tuned or commissioned, if necessary, especially after major maintenance activities of the lighting system (relamping, cleaning of luminaires, etc.).

CCD Imaging Sensors as ALSs

CCD imaging sensors are quite promising (Granderson et al. 2010; Howlett et al. 2007; Newsham and Arsenault 2009; Sarkar et al. 2008; Sarkar and Mistrick 2006), in the sense that they can measure luminance patterns approximating those of the human visual system. However, their capabilities are yet rather limited due to errors associated with the evaluation of illuminance from luminance, as well as due to problems associated with their calibration procedure and commissioning. In addition, their increased cost and size can impose practical limitations during their installation, let alone the privacy issues. However, their ability to control a shading system and to be used for occupancy sensing might prove to be cost-effective (Fig. 18).

CCD Sensor Integration

CCD sensors are widely used in many systems (e.g., digital cameras, camcorders, industrial machine vision, microscopy, spectroscopy, etc.) and their characteristics are well known. The pixels they consist of are stimulated by light and produce a

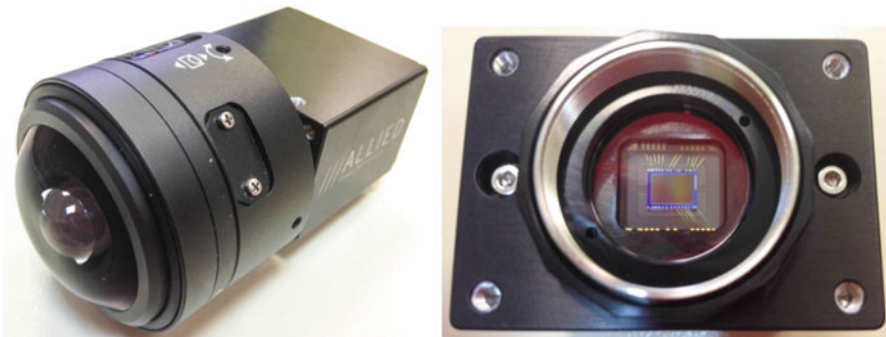


Fig. 18 CCD sensor used as ALS

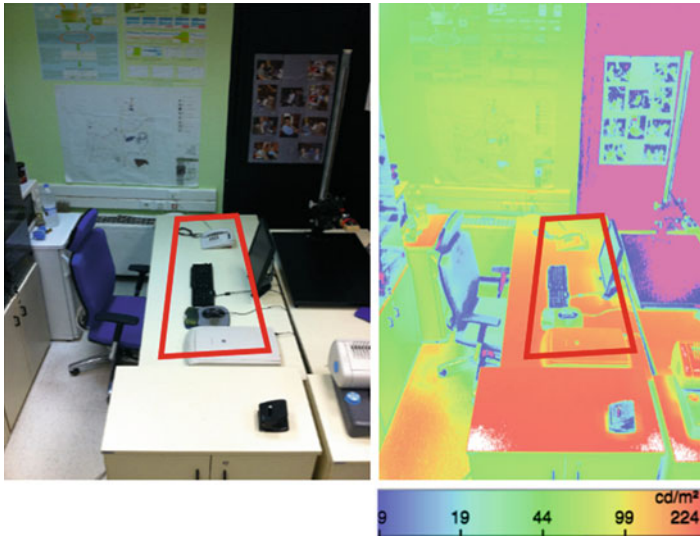


Fig. 19 The image captured by the CCD sensor (*left*) and the corresponding luminance image (*right*). The CCD sensor is able to estimate illuminance values in selected working planes (*red area*)

digital image. In detail, a charge-coupled device (CCD) transports electrically charged signals and is used as a light sensor. A CCD chip is divided into pixels. Each pixel has a potential well that collects the electrons produced by the photoelectric effect. At the end of an exposure (frame), each pixel has collected an amount of electrons (i.e., charge) proportional to the amount of light that fell onto it. The CCD is then read out by cycling the voltage values applied to the chip in a process called “clocking.” Due to the structure of the CCD, clocking transfers the charge of each pixel to the adjacent one (Fraden 2010).

Compactly, a CCD sensor captures images of the room and converts them to luminance images (Fig. 19). Then, with appropriate control, illuminance values of different working planes are calculated using the corresponding luminance maps. Signals of each luminaire individually and different dimming levels based on the results of a single image capture can be produced from the CCD sensor. This would be equal to a group of conventional ALSs, each one aiming at a different section of the room and controlling a single luminaire. As the CCD sensor’s field of view covers areas of interest, the control system will segment the image in the preset parts, calculate the lighting level on each one, and send the control signals only to luminaires that must be dimmed (Kontaxis et al. 2012).

Limitations and Advantages of the Integration of CCD Sensors

The CCD sensor can be placed almost anywhere on the ceiling, and with the use of a wide field of view lens, time series of images can be captured. Sensor shielding is not

necessary as it is in the case of ALSs. Thus, the light output fluctuation caused by outdoor lighting condition fluctuation – especially when blinds are used – is not decreased, and the control can be more accurate. By implementing an image-processing algorithm, these images can be converted to images with luminance distribution. Therefore, any part of the image can be used as input signal, so as to control the lighting system.

The current standards define the desired lighting levels in a room which are expressed in illuminance values (lux). However, CCD sensors are calibrated so that they can measure the luminance values (cd/m^2) with the highest possible accuracy. Since luminance values depend on the direction of view, a set of algorithms is necessary to calculate the lighting levels in illuminance values (lux) of any area of the room using the captured image. Using a target with known photometric properties (such as a perfect diffuser of known reflectance), illuminance can be estimated. Since in a real working environment it is possible to find a medley of surfaces with various optical properties, the conversion of luminance to illuminance values can be difficult. Nevertheless, if illuminance values are known, any comparison with a predefined set point can drive the dimming system. However, driving a control system in real time requires a very fast image-processing analysis (Doulos et al. 2013).

A single CCD sensor used as an ALS can monitor a set of working planes that require different illuminance values and adjust the artificial lighting system in response to daylight individually for each working plane. A direct view of the illuminated area of the luminaires or external openings is allowed, as long as lens scattering is minimal. Such a system can be designed to account for changes on the reflectance of the calibrated surfaces by sensing surface color and automatically shifting to neighboring pixels when needed. The CCD sensor may also be used to sense occupancy through detection of movement in the image (Sarkar and Mistrick 2006).

As mentioned in previous paragraphs, spectral response of ALSs is wider than the photopic human eye sensitivity. Thus, ALSs perceive larger quantities of light than the human eye. A Bayer filter – commonly used in color CCD sensors – filters the light intensity in RGB spectral regions. The spectral response curve of the green pattern corresponds to the photopic human eye sensitivity $V(\lambda)$, which means that luminance values can be extracted using only the green component (Doulos et al. 2013).

References

- Bierman A, Conway KM (2000) Characterizing daylight photosensor system performance to help overcome market barriers. *J Illum Eng Soc* 29(1):101–115
- Choi AS, Mistrick RG (1997) On the prediction of energy savings for a daylight dimming system. *J Illum Eng Soc* 26(2):77–90
- Choi AS, Mistrick RG (1999) Analysis of daylight responsive dimming system performance. *Build Environ* 34:231–243

- Choi AS, Song KD, Kim YS (2005) The characteristics of photosensors and electronic dimming ballasts in daylight responsive dimming systems. *Build Environ* 40(1):39–50
- Doulos L, Tsangrassoulis A, Topalis FV (2005) A critical review of simulation techniques for daylight responsive systems. In: *Proceedings DYNASTEE 2005*, Athens, 12–14 Oct
- Doulos L, Tsangrassoulis A, Topalis FV (2008a) Quantifying energy savings in daylight responsive systems: the role of dimming electronic ballasts. *Energy Build* 40:36–50
- Doulos L, Tsangrassoulis A, Topalis FV (2008b) The role of spectral response of photosensors in daylight responsive systems. *Energy Build* 40:588–599
- Doulos L, Tsangrassoulis A, Bouroussis CA, Topalis FV (2013) Reviewing drawbacks of conventional photosensors: are CCD/CMOS sensors the next generation? In: *Proceedings Lux Europa 2013*, Krakow, 17–19 Sept
- Doulos L, Tsangrassoulis A, Topalis F (2014) Multi-criteria decision analysis to select the optimum position and proper field of view of a photosensor. *Energy Convers Manag* 86:1069–1077
- European Standard EN 15193 (2007) Energy performance of buildings, energy requirements for lighting, CEN/Technical Committee 169
- Floyd D, Parker D (1995) Field commissioning of a daylight-dimming lighting system. In: *Proceedings of right light 3*, vol 1, Newcastle upon Tyne
- Fraden J (2010) *Handbook of modern sensors*, 4th edn. Springer, New York
- Granderson J, Gaddam V, DiBartolomeo D, Li X, Rubinstein F, Das F (2010) Field-measured performance evaluation of a digital daylighting system. *Leukos* 7(2):85–101
- Howlett O, Heschong L, Mchugh J (2007) Scoping Study for Daylight Metrics from Luminance Maps. *Leukos* 3(3):201–215
- Kim SY, Kim J (2007) The impact of daylight fluctuation on a daylight dimming control system in a small office. *Energy Build* 39(8):935–944
- Kim SY, Mistrick RG (2001) Recommended daylight conditions for photosensor system calibration in a small office. *J Illum Eng Soc* 30(2):176–188
- Kontaxis PA, Bouroussis CA, Doulos L, Topalis FV (2012) Applications of CCD sensors in photometry and in daylight responsive systems. In: *Proceedings Balkan light 2012*, Belgrade, 3–6 Oct
- Lee ES, Selkowitz SE (2006) The New York Times headquarters daylighting mockup: monitored performance of the daylighting control system. *Energy Build* 38:914–929
- Lee ES, DiBartolomeo DL, Selkowitz SE (1999) The effect of venetian blinds on daylight photoelectric control performance. *J Illum Eng Soc* 28(1):3–23
- Li DHW (2010) A review of daylight illuminance determinations and energy implications. *Appl Energy* 87(3):2109–2118
- Li DHW, Cheung KL, Wong SL, Lam T (2010) An analysis of energy-efficient light fittings and lighting controls. *Appl Energy* 87(2):558–567
- Littlefair PJ (2001) Photoelectric control: the effectiveness of techniques to reduce switching frequency. *Light Res Technol* 33(1):43–58
- Littlefair PJ, Motin A (2000) Lighting controls in areas with innovative daylighting systems: a study of sensor type. *Light Res Technol* 33(1):59–73
- Mistrick R, Sarkar A (2005) A study of daylight-responsive photosensor control in five daylighted classrooms. *Leukos* 3(1):51–74
- Mistrick R, Thongtipaya J (1997) Analysis of daylight photocell placement and view in a small office. *J Illum Eng Soc* 26(2):150–160
- Mistrick R, Chen C, Bierman A, Felts D (2000) A comparison of photosensor-controlled electronic dimming systems in a small office. *J Illum Eng Soc* 29(1):66–80
- Newsham G, Arsenaault C (2009) A camera as a sensor for lighting and shading control. *Light Res Technol* 41(3):143–163
- Rubinstein F, Ward G, Verderber R (1989) Improving the performance of photo-electrically controlled lighting systems. *J Illum Eng Soc* 18(1):70–94
- Rubinstein F, Siminovitich M, Verderber R (1993) 50 % energy savings with automatic lighting controls. *IEEE-IAS Trans Ind Appl* 29(4):768–773

- Rubinstein F, Avery D, Jennings J (1997) On the calibration and commissioning of lighting controls. In: Proceedings of right light 4, Copenhagen, pp 225–230
- Sarkar A, Mistrick R (2006) A novel lighting control system integrating high dynamic range imaging and DALI. *Leukos* 2(4):307–322
- Sarkar A, Fairchild M, Salvagio C (2008) Integrated daylight harvesting and occupancy detection using digital imaging. SPIE, Sensors, Cameras, and Systems for Industrial/Scientific Applications IX, San Jose, California, USA, Vol. 6816

Optical Wireless Applications

Z. Zhou and M. Kavehrad

Contents

Introduction	636
Indoor Optical Wireless Model	640
Introduction	640
Single-Input/Single-Output (SISO) Model	640
Multiple-Input-Multiple-Output (MIMO) Model	641
Typical Impulse Response Distortions	644
Summary	644
Indoor Optical Wireless Channel Model Error Analyses and Calibration	646
Introduction	646
Historical Review of Indoor Optical Wireless Channel Research and Error Analyses	646
Indoor Optical Wireless Channel Model Analyses	648
Simulation and Results	650
Discussion on the Model Error and Calibration Method	653
Summary	655
References	663

Abstract

As light-emitting diodes (LEDs) increasingly displace incandescent lighting over the next few years, general applications of optical wireless (OW) technology are expected to include wireless Internet access, broadcast from LED signage, and machine-to-machine positioning and navigation by light. This section explores several fundamental research topics of indoor optical wireless communications (IOWC). The authors develop a simulation method to generate IOWC channel

Z. Zhou • M. Kavehrad (✉)

Department of Electrical Engineering, The Center for Information and Communications Technology Research (CICTR), Pennsylvania State University, University Park, PA, USA
e-mail: mkavehrad@psu.edu

models by tracking light reflections. The method is further optimized by investigating the contribution of each order of reflections and proposing a calibration method.

Introduction

High-speed communication has found an important role in people's daily lives. The wireless home link (WHL) is becoming more and more a reality. In the next decades, wireless communication will play a significant role in electronic device interconnections.

Visible light communications (VLC) is an emerging wireless communication technology based on white light-emitting diode (LED). The LED is generally considered as the next-generation light source and may replace the universal incandescent bulb and fluorescent bulb in home and workplaces, because of its advantages such as long lifetime, low-power consumption, small size, and being environment friendly. Moreover, LED has a high response sensitivity to support high-speed communication. In VLC, LED takes both communication and illumination duties. We modulate the LED by user data to create illumination and communication dual-functional "base station light," while human can hardly sense the flickers due to high modulation speed of hundreds of megabit per second. A typical VLC system is demonstrated in Fig. 1.

Compared with conventional radio frequency (RF) wireless communication, the VLC has many advantages, such as high data rate, energy saving, secure transmission, lack of electromagnetic interference, and, most importantly, spectral regulation (Kavehrad 2010; Kavehrad and Fadlullah 2010). However, current market drivers are insatiable for bandwidth demand that is driving the consideration of VLC. Table 1 provides several examples of their capacity.

Wi-Fi revolutionized how we communicate. Now, LiFi, a hybrid system that combines a wireless network with light-emitting diodes (LEDs), may bring a similar revolution to location identification. The approach offers a way to find items or people in large stores, malls, high-rise buildings, hospitals, and museums.



Fig. 1 Typical VLC system

Table 1 Capacities of current market LED drivers

Manufacturer	Model	Date rate (Mbps)	Current (mA)
Micro Linear	ML6633	200	82
Sumitomo Electric	F0601720Q	266	100
ON Semiconductor	MC10SX1130	300	100
Semtech	NT22030	250	25
Maxim	MAX3967A	270	100

Developed by a research team at Pennsylvania State University (Lee and Kavehrad 2012), the LiFi system uses radio frequency transmitters and overhead LED lights to pinpoint an exact location in indoor and environments. This provides a much needed alternative to Global Positioning Systems (GPS) for indoor use because inside radio frequencies interfere with the GPS signal.

So how does the LiFi system work? Envision large stores or malls with overhead LED light fixtures, each with a location code. At the store or mall entrance, a computer that's accessible via keyboard or telephone contains a database of all the items available. Shortly after a query, the location of a desired item appears. The system identifies items through a photodiode and ZigBee receiver merchandise tag. Because the walls block light transmission, item locations are transmitted from the LEDs via a ZigBee multi-hop wireless network.

Even when the merchandise is moved from room to room, the location remains available because a different LED overhead light with a different location code signals the tag. Beyond malls and stores, LED-transmitted information may also find use in settings where radio frequency signals can interfere with equipment, such as hospitals. The system could identify the floor where a person is situated. In museums, navigation systems could guide people through large buildings by reading the final destination signal from a handheld photodiode device and initializing lights or other indicators to show the proper path.

To begin with, by using novel transmission techniques, VLC is able to provide 513 Mbps transmission, according to Vucic et al. (2010). It is anticipated that in the future, this rate will be increased into gigabit per second ranges. Second, light source combines illumination and communication. Third, light is confined in the room and this offers physical layer security. Fourth, VLC causes no interference in electromagnetic sensitive environments and provides an effective wireless solution to these environments, such as hospitals, aircrafts, and others. Finally, and most importantly, VLC offers sufficient bandwidth resources, free from regulation. Currently, bandwidth for wireless communication is exhausted in the microwave range and is strictly controlled by the Federal Communications Commission (FCC). Table 2 demonstrates the data rate of VLC and its RF competitors (Lee et al. 2007; <http://visiblelightcomm.com/page/3/>).

Note that very high VLC bandwidth records in the literature are science experiments, which are hardly suitable for dual use (illumination and communication). An important reason is that for the lighting LED array to be energy efficient and to produce lots of lumens, one would need to feed the array with a very high voltage

Table 2 Comparison between VLC and competing RF Techniques

	Data rate (bps)	Protocol	Frequency spectrum range (Hz)	Distance (m)	EMI
VLC	>1G	802.15.7	400 T–800 T	>20	No
WiGig	2G	802.11ad	2.4G, 5G, 60G	10	Yes
Wi-Fi	54 M – 800 M	802.11a/b/g/n/ac	2.4G, 5G	100	Yes
Bluetooth	1 M	802.15.1	2.4G	10	Yes
UWB	110 M	802.15.3a	3.1G–10.6G	10	Yes
ZigBee	250 K	802.15.4	868/915 M, 2.4G	10–100	Yes

(400 V) or a very high current, due to cascading of array elements. There is no other way. This holds for all semiconductor devices: high lumens cannot be offered along with a fast response time or a high bandwidth. In addition, most illumination LED bulbs combine blue LED chips with yellow phosphors to generate white light. The phosphors have a significantly longer response time than the blue LED chips, which substantially limits the LED modulation speed.

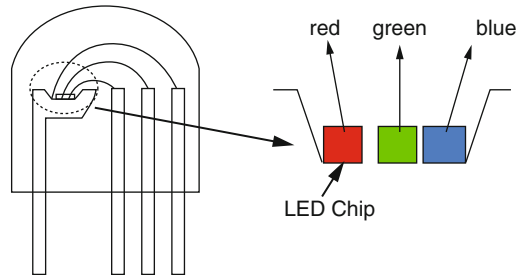
Therefore, the sensible approach is not to try to turn the 400 V at gigabit per second. What is practical is to insert in the middle of the high brightness a very high-speed LED. Then one creates data showering with that one or that section.

Recognizing the importance of VLC, there has been much research on this topic which involves numerous aspects of VLC. Yun and Kavehrad (1992) and Kahn et al. (1998) provided imaging diversity receivers for indoor optical wireless communication. Anand and Mishra (2010) proposed a novel VLC modulation scheme called pulse dual slope modulation (PDSM) which solved intra-frame and inter-frame flickers. Kim et al. (2010) designed spread spectrum code for VLC to improve system performance. Park and Lee (2011) integrated dimming control with VLC. The dimming control made users feel psychologically comfortable when they use the VLC.

Though these investigations have made substantial contribution to VLC, they mostly work on the single-source scenarios, rather than multiple sources. In order to fill the gap, this work focuses on the interference among sources when plural light sources are applied. Since most rooms install plural light sources on the ceiling, plural sources produce light coverage areas overlapping, introducing inter-symbol interference (ISI). The ISI degrades system performance. An effective technique to implement MIMO in VLC is using multi-chip White LED (WLED). A typical RGB WLED is shown in Fig. 2. It is in fact a combination of blue, green, and red LEDs. By changing the mixture ratio of these primary colors, different colors can be produced. The color-tunable lighting is being developed for market now. This technique offers new possibilities for VLC multiple-input-multiple-output (MIMO) strategies and bandwidth improvement, for instance, the wavelength-division multiplexing (WDM). Nevertheless, the market penetration of these sources might be slow.

In the author's opinion, there will be other architectures besides the separate communications channel small embedded LED approach that will work, and there

Fig. 2 The structure of RGB white LED (*WLED*)



are many more going on in LED lighting system design than in high-voltage or high-current solutions. These will make new approaches possible. A recent paper reported the method of increasing the modulation rate by sweeping out the remaining carriers (Kishi et al. 2013).

This method increases the maximum error-free bit rate by 38 %. Nevertheless, the additional current path for carrier sweep-out requires more driver power dissipation, and the authors have to apply a CMOS inverter to reduce the power dissipation.

If one uses a small visible but non-illumination grade LED, that is one approach that does divorce the chaos in the current solid-state lighting (SSL) design space from VLC – but also it frees one up to use a high-efficiency IR LED and then it is not VLC anymore. As a result, you get the advantages of communicating in the dark. This is but one possibility. In other words, we implant small lower-power LEDs in the LED array for communications only. These can then be modulated at much faster rates.

This section reports on several fundamental aspects of multiple-source VLC. Source layout is one of the most important factors that affect light footprints overlapping and thus producing ISI. It determines the pattern and extent of the lights overlapping. We will explore the VLC performance in several conventional household layouts and investigate the impact of these layouts on VLC. Multiple sources increase multipath distortion. As orthogonal frequency-division multiplexing (OFDM) is proved to be effective in reducing multipath-involved ISI, we will investigate this modulation scheme for VLC applications. Multiple-input and multiple-output (MIMO) techniques will also be included as they provide either reliability improvement or bandwidth efficiency increase. Based on these investigations, we will further explore VLC performance in real applications, such as aircraft cabin wireless communications and indoor navigations.

In this section, we describe the Lambertian emission pattern of LEDs and the diffused features in indoor environments. Note that LEDs follow the Lambertian emission pattern; however, they are usually shielded in lighting to reduce glare; therefore, diffuse source patterns are more important with secondary optical designs playing more of a role. Of course, LED technology is rewriting some of the rules of lighting fixture design. Based on the theory, we trace light pulses to establish a MIMO indoor wireless channel model on specific source layout. After that, we will investigate VLC performance in specific applications, including aircraft onboard

wireless communications and indoor navigations. These areas might be the early adopters of VLC.

We will also explain the pitfalls of MIMO systems, in particular, where there is a strong background visible light, such as sunlight shining on the photo-receivers trying to detect the information signals being carried on VLC system. This brings us to a serious problem with VLC called the near-far problem. When a receiver is trying to detect light from a far LED source and it is located near a window or a place that makes it exposed to a stronger visible light, such as sunlight.

Indoor Optical Wireless Model

Introduction

Modeling indoor wireless channel is fundamentally tracing the light pulse, which experiences multiple Lambertian reflections. A light pulse is emitted at the source and propagates to all directions, while the directional power distribution obeys Lambertian law. When a part of the light beam arrives at a point on room surface, it is reflected in a Lambertian pattern and this point works as a secondary light source. The process continues until the light reaches the receiver/user. Since the successive captured lights at the receiver come from the same pulse, but experience different optical paths, the overall receiver response demonstrates the impulse response from the source to the receiver. In most cases, there are plural sources and receivers; the indoor wireless channel is described by an impulse response matrix.

Single-Input/Single-Output (SISO) Model

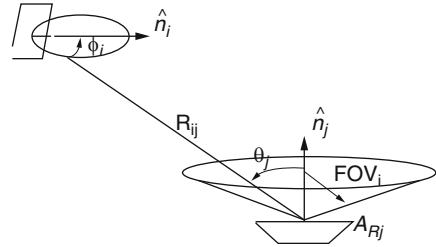
Indoor optical wireless channel characteristics by computer simulations were first presented by Barry et al. (1993). The authors set up a simulation tool to evaluate multipath impulse response of optical indoor channel.

Barry's room model divides the room surfaces including ceiling and floor into many grid elements. Each element is called a reflector and is assigned a reflection coefficient. The light pulse travels from a source to all the reflectors in a line-of-sight (LOS) Lambertian pattern. When light arrives at one of these, light intensity will be decreased by the multiplying reflection factor. Then, this element will be seen as a secondary Lambertian source emitting light pulse to all other reflectors. This procedure continues until light arrives at the system receiver.

As shown in Fig. 3, the impulse response of LOS Lambertian transmission between elements is expressed as

$$h^{(0)}(t; S, R) \approx \frac{n+1}{2\pi} \frac{\cos^n(\phi) \cos(\theta) A_R}{R^2} \text{rect}\left(\frac{\theta}{FOV}\right) \delta\left(t - \frac{R}{c}\right) \quad (1)$$

Fig. 3 Line-of-sight Lambertian transmission



where superscript zero means light travels from source to receiver directly without passing through objects (line of sight (LOS) is required); S and R represent source and receiver parameters set, respectively; n is the mode number of the radiation lobe that specifies the source directionality; ϕ is the angle between source orientation vector and the vector pointing from source to receiver; θ is the angle between receiver orientation vector and the vector pointing from receiver to source; A_R is receiver area; FOV is the field of view (FOV) of the receiver; R is the distance between the source and receiver; and c is the speed of light. When light reaches the destination reflector, the transmission continues and the destination reflector becomes the starting source of next transmission. Therefore, the total impulse response can be calculated recursively. Let $h^{(k)}(t;S,R)$ denote the response of light undergoing exactly k reflections:

$$h^{(k)}(t; S, R) = \int_S h^{(0)}(t, S, E_R) \otimes h^{(k-1)}(t, E_S, R) ds \tag{2}$$

where E_R is the parameter set of a reflector as a receiver and E_S is the parameter set of a reflector as a transmitter. The integration carries on all reflectors. As a result, the overall impulse response is the infinite sum of the impulse responses undergoing all possible number of reflections:

$$h(t; S, R) = \sum_{k=0}^{\infty} h^{(k)}(t; S, R) \tag{3}$$

Multiple-Input-Multiple-Output (MIMO) Model

In most cases, where there are plural light sources and users, it is necessary to extend an MIMO indoor wireless channel model from the SISO model.

Kavehard and Alqudah developed the MIMO indoor optical wireless channel simulation based on diffuse-transmission configuration (Alqudah and Kavehrad 2003). This method transforms impulse response into a matrix form.

In the MIMO model, considering the transmitter and receiver, the transfer function between any two points is divided into four components. Suppose there are J sources, N surface elements, and M receivers, the four components are the transfer function between a source and surface element ($F_{J \times N}$), the transfer function between surface elements ($\Phi_{N \times N}$), the transfer function between surface element and receiver ($G_{N \times M}$), and the direct transfer function between source and receiver ($D_{J \times M}$).

Source Profile ($F_{J \times N}$)

$F_{J \times N}$ is a J by N matrix:

$$F_{J \times N} = \begin{bmatrix} f_{11} & f_{12} & \cdots & \cdots & f_{1N} \\ f_{21} & f_{22} & \cdots & \cdots & f_{2N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ f_{J1} & f_{J2} & \cdots & \cdots & f_{JN} \end{bmatrix} \tag{4}$$

Each entry of the matrix f_{sk} is the transfer function between a source s and element k as

$$f_{sk} = \frac{n + 1}{2\pi} \frac{\cos^n(\phi_{sk}) \cos(\theta_{sk}) A_k}{R_{sk}^2} \delta\left(t - \frac{R_{sk}}{c}\right) u\left(\frac{\pi}{2} - \theta_{sk}\right) \tag{5}$$

where n is the Lambertian order of the source; ϕ_{sk} is the emitting angle from source s to element k ; θ_{sk} is the incident angle from source s to element k ; R_{sk} is the distance between source s and element k ; A_k is the area of the element; c is the speed of light; and $u()$ is the unit step function.

Environment Matrix ($\Phi_{N \times N}$)

$\Phi_{N \times N}$ is a N by N matrix. In matrix format, considering up to n reflections, it is expressed as

$$\Phi_{N \times N} = \begin{cases} I_{N \times N} + \Psi_{N \times N} + \Psi_{N \times N}^2 + \Psi_{N \times N}^3 + \dots + \Psi_{N \times N}^{n-1}, & n \geq 2 \\ I_{N \times N}, & n = 1 \end{cases} \tag{6}$$

where $I_{N \times N}$ is the N by N identity matrix and $\Psi_{N \times N}$ is given by

$$\Psi_{N \times N} = \begin{bmatrix} \psi_{11} & \psi_{12} & \cdots & \cdots & \psi_{1N} \\ \psi_{21} & \psi_{22} & \cdots & \cdots & \psi_{2N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \psi_{N1} & \psi_{N2} & \cdots & \cdots & \psi_{NN} \end{bmatrix} \tag{7}$$

ψ_{ik} represents the transfer function between the two elements i and k as

$$\psi_{ik} = \begin{cases} 0, & i = k \\ \frac{\rho_i \cos(\phi_{ik}) \cos(\theta_{ik}) A_k}{\pi R_{ik}^2} \delta\left(t - \frac{R_{ik}}{c}\right) u\left(\frac{\pi}{2} - \theta_{ik}\right), & i \neq k \end{cases} \quad (8)$$

where ρ_i is the reflection coefficient of element i ; ϕ_{ik} is the emitting angle from element i to element k ; θ_{ik} is the incident angle from element i to element k ; R_{ik} is the distance between element i and element k ; A_k is the area of the element; and c is the speed of light.

Receiver Profile ($G_{N \times M}$)

$G_{N \times M}$ is a N by M matrix. It is represented as

$$G_{N \times M} = \begin{bmatrix} g_{11} & g_{12} & \cdots & \cdots & g_{1M} \\ g_{21} & g_{22} & \cdots & \cdots & g_{2M} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ g_{N1} & g_{N2} & \cdots & \cdots & g_{NM} \end{bmatrix} \quad (9)$$

The entry g_{ir} is the transfer function from element i to receiver k as

$$g_{ir} = \frac{\rho_i \cos(\phi_{ir}) \cos(\theta_{ir}) A_r}{\pi R_{ir}^2} \delta\left(t - \frac{R_{ir}}{c}\right) u(FOV_r - \theta_{ir}) \quad (10)$$

where ρ_i is the reflection index of element i ; ϕ_{ir} is the emitting angle from element i to receiver r ; θ_{ir} is the incident angle from element i to receiver r ; R_{ir} is the distance between element i and receiver r ; A_r is the area of the receiver; c is the speed of light; and FOV_r is the FOV of the receiver r .

Direct Response Matrix ($D_{J \times M}$)

$D_{J \times M}$ is a J by M matrix. It is represented as

$$D_{J \times M} = \begin{bmatrix} d_{11} & d_{12} & \cdots & \cdots & d_{1M} \\ d_{21} & d_{22} & \cdots & \cdots & d_{2M} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ d_{J1} & d_{J2} & \cdots & \cdots & d_{JM} \end{bmatrix} \quad (11)$$

The entry d_{ir} is the transfer function from element s to receiver r as

$$d_{sr} = \frac{n+1}{2\pi} \frac{\cos^n(\phi_{sr}) \cos(\theta_{sr}) A_r}{R_{sr}^2} \delta\left(t - \frac{R_{sr}}{c}\right) u(FOV_r - \theta_{sr}) \quad (12)$$

where n is the Lambertian order of source s ; ϕ_{sr} is the emitting angle from source s to receiver r ; θ_{sr} is the incident angle from source s to receiver r ; R_{sr} is the distance

between element s and receiver r ; A_r is the area of the receiver; c is the speed of light; and FOV_r is the FOV of the receiver r .

Total Response ($H_{J \times M}$)

The total impulse response matrix of the MIMO system is given by

$$H = D_{J \times M} + F_{J \times N} \otimes \Phi_{N \times N} \otimes G_{N \times M} \quad (13)$$

where \otimes indicates matrix convolution.

Typical Impulse Response Distortions

In most locations, the impulse response has insignificant impulse spread. However, two kinds of locations exhibiting considerable impulse distortion are found. The first kind is the room corner (sections “[Indoor Optical Wireless Model](#),” and “[Introduction](#)”); the second kind is the overlapping area of light footprints (sections “[Single-Input/Single-Output \(SISO\) Model](#),” and “[Typical Impulse Response Distortions](#)”).

Two typical types of distorted impulse responses from the entries of H are demonstrated in Fig. 4. At the room corner, the spread of impulse response comes from multipath effect of light reflections. In the corner area, the receiver captures reflected lights, which experience different reflection paths and cause the decreasing tail in impulse response. What is worth mentioning is that the tail has much lower power compared with the peak.

In the overlapping areas, several equally high sharp peaks in the impulse response are observed. The reason for the multiple peaks is that lights from different sources enter receiver via different LOS paths, the time difference of the arrivals causes the multiple peaks. This is an important influencing factor to VLC.

Summary

In this section, the approaches to establish indoor optical wireless models are investigated. These approaches are based on tracing light reflections in the environment. The most efficient method is proposed by Alqudah and Kavehrad (2003). In their method, the entire impulse response matrix is divided into several stages. Each stage provides a group of parameters. This method substantially saves simulation time when we calculate the impulse response matrix for multiple receiver locations is calculated. Once the environmental matrix $\Phi_{N \times N}$ is obtained at one location, it is saved and reused for other locations.

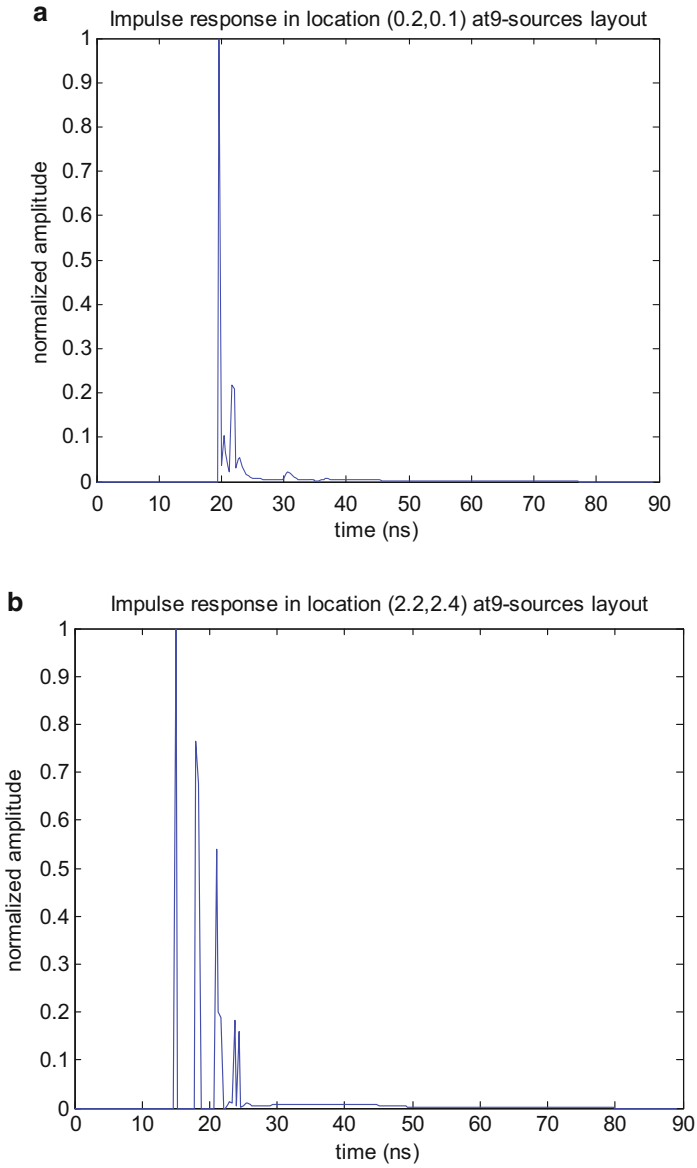


Fig. 4 (a) Multipath effect, (b) inter-source interference

Indoor Optical Wireless Channel Model Error Analyses and Calibration

Introduction

This section analyzes the impact of high-order light reflections on indoor optical wireless (IOW) channel models. Based on the authors' findings, a calibration method is proposed to reduce model errors. Channel models are generated by tracing and adding up diffuse light reflections and sequential sub-reflections along its traveling path. As computation complexity increases significantly with the number of reflection orders considered, researchers conventionally take the contribution of a first few orders, most commonly three, to represent the complete channel. Discarded high-order reflections bring no significant performance difference to low-speed systems; however, major contemporary IOW research institutions focus on high-speed Gbps communications where their impact is no longer negligible. Root-mean-square (RMS) delay spread, for instance, is severely underestimated by neglecting high-order reflections. We simulate an IOW system in an ordinary $6\text{ m} \times 6\text{ m} \times 3\text{ m}$ lab room and calculate the contributions of each order of reflections at 841 locations. It shows the RMS delay-spread estimation using first 3 orders is underestimated by 15.3 % on the average and 26.6 % as the maximum. To limit error within half a symbol period, 1 and 10 Gbps systems tolerate underestimations up to 13.7 % and 1.4 %, respectively. These must be achieved by applying first 5 and 9 orders. To keep the computation efficiency of low-order reflection models and improve their accuracies, we propose a statistical calibration method. It reduces average model error of first three reflection orders from 15.7 % to 4.3 %. After calibration, the numbers of orders required by 1 and 10 Gbps systems are individually reduced to 3 and 7.

Historical Review of Indoor Optical Wireless Channel Research and Error Analyses

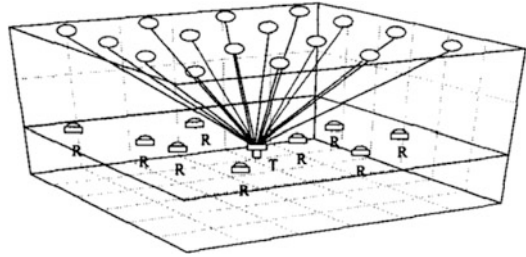
Indoor optical wireless communication (OWC) has been studied extensively in recent years. It is recognized as a strong candidate technology for the next-generation high-speed wireless networks. Compared with conventional radio frequency (RF) wireless communications, it offers significant advantages (Fadlullah and Kavehrad 2010; Barros et al. 2012). First, the visible, infrared, and ultraviolet spectral regions offer virtually unlimited bandwidth and are not regulated. Second, light is confined in rooms because it is not able to penetrate opaque barriers, such as walls, ceilings, and floors. This feature enables band reuse between neighboring rooms and provides physical layer communications security. Third, OWC neither generates nor is susceptible to electromagnetic interferences. This technology, therefore, can be widely applied to sensitive environments including hospitals, aircrafts, mines, power plants, and others.

Realizing the substantial potentials of OWC, many research institutions are being sponsored in this area by government and industry. In the United States, the Pennsylvania State University and Georgia Institute of Technology lead the NSF Center on Optical Wireless Applications (COWA). They collaborate with industrial leaders to evaluate the potentials of the interdisciplinary research center activities, in providing leadership to develop new generation of environment-friendly and extremely wideband optical wireless technology applications ([NSF Center on Optical Wireless Applications](#)). The Center for Ubiquitous Communication by Light (UC-Light) is established by the University of California, Riverside. This center focuses on white light-emitting diodes (LEDs) for wireless information sharing and retrieving ([Center for Ubiquitous Communication by Light](#)). In Europe, the Home Gigabit Access (OMEGA) project started on 2008. It will develop a user-friendly home area network capable of delivering high-bandwidth services and content at a transmission speed of one gigabit per second ([Home Gigabit Access](#)). In Japan, the Visible Light Communications Consortium (VLCC) was founded in 2003. It is aiming to publicize and standardize the visible light communication technology, which has been discussed and evaluated in various industry fields ([Visible Light Communications Consortium](#)).

In OWC, the channel model is one of the most important research subjects, since it indicates the transmission capacity and communication performance. The first channel characteristic model study for infrared was made by Gfeller and Bapst in 1979. Later, a recursive simulation method for diffusion indoor optical wireless channel was developed by Barry et al. (1993). Not long after, Alqudah and Kavehrad (2003) invented the method to simulate multiple-input and multiple-output characteristics of indoor optical wireless link. Though these methods substantially increase simulation efficiency, the computation still consumes considerable time, and the complexity significantly increases with the number of reflection orders concerned. To make sure the model can be simulated in a reasonable amount of time, researchers use channel models considering only first a few bounces, most commonly three, to represent the complete model. This approximation has been generally accepted for decades, because of the insignificant performance difference in relatively low-speed transmissions at that time. As present leading research institutions are moving to Gbps high-speed transmissions, this approximation no longer holds. Consequently, there is an urgent demand for explorations on higher-order reflections and involved model errors. This section analyzes the impact of high-order reflections on channel models. Based on our research, we summarize the general rules of model error distributions, with bounce order increase. A calibration method is developed to reduce errors and at the same time maintain computation efficiency.

The rest of the section is organized as follows: in section “[Indoor Optical Wireless Channel Model Analyses](#),” we discuss the methods to create indoor optical wireless channel model; in section “[Simulation and Results](#),” we demonstrate the simulation results and discuss the impact of high-order reflections; then, we propose a calibration method to low-order channel models in section “[Discussion on the Model Error and Calibration Method](#)” and draw summary in section “[Summary](#).”

Fig. 5 A typical indoor optical wireless communication system



Indoor Optical Wireless Channel Model Analyses

Indoor Optical Wireless Channel Overview

A typical indoor optical wireless communication system is demonstrated in Fig. 5. The transmitter is placed in the center of the room and generates multiple diffusion spots on the ceiling. Receivers spread in the room and receive data from the spots (Yun and Kavehrad 1992; Kavehrad and Jivkova 1999).

Indoor optical wireless channels are characterized by impulse responses (IRs). In this section, we consider multiple spots sending data on the ceiling as spatially diverse transmitters. The IRs are generated correspondingly when the receiver is placed at different locations in the room. Assuming there are M spots and N receiver locations, the channel characteristic profile of this communication system is represented by an IR matrix $H_{M \times N}(t)$ as

$$H_{M \times N}(t) = \begin{bmatrix} h_{11}(t) & h_{12}(t) & \dots & \dots & h_{1N}(t) \\ h_{21}(t) & h_{22}(t) & \dots & \dots & h_{2N}(t) \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ h_{M1}(t) & h_{M2}(t) & \dots & \dots & h_{MN}(t) \end{bmatrix} \quad (14)$$

The element $h_{ij}(t)$ indicates the impulse response from the i -th spot to the j -th location. As multiple spots are applied for spatial diversity, assuming equal-gain combining is applied, the total impulse response when the receiver is placed at the j -th location is

$$h_j(t) = \sum_{i=1}^M h_{ij}(t) \quad (15)$$

Channel Model Simulation

The principle of channel simulation is partitioning a room into numerous reflectors and tracing the light diffusive reflections from them. In this simulation, we place the laser transmitter in the center of the room, pointing upward. By dividing the beam, it generates multiple diffusion spots on the ceiling. The diffusions follow Lambertian pattern and light travels to all reflectors through line-of-sight (LOS) paths.

When light arrives at a reflector, after reflection loss, the reflector becomes a secondary diffusion spot and retransmits light to other reflectors in the same manner. These diffusions keep repeating in the room, with intensity decrease from propagation attenuations and surface absorptions. By collecting the lights from all diffusions at the receiver, we get the total IR through diffusion paths. It is apparent that a diffusion channel is fundamentally a combination of a large amount of scaled and delayed LOS channels.

The LOS impulse response from one transmitter i to receiver j $g_{ij}^{(0)}(t)$ is given by Barry et al. (1993):

$$g_{ij}^{(0)}(t) \approx \frac{1}{\pi} \frac{\cos(\phi_{ij}) \cos(\theta_{ij}) A_{Rj}}{R_{ij}^2} \delta\left(t - \frac{R_{ij}}{c}\right) \text{rect}\left(\frac{\theta_{ij}}{FOV_j}\right) \quad (16)$$

As shown in Fig. 3, the superscript (0) indicates that it is a LOS IR, because zero reflections are experienced from i to j ; ϕ_{ij} is the angle between the vector pointing from transmitter to receiver and the transmitter normal; θ_{ij} is the angle between the vector pointing from receiver to transmitter and the receiver normal; A_{Rj} is the effective receiver area; R_{ij} is the distance between the two; c is the speed of light; FOV_j is the field of view (FOV) of the receiver.

Previous research successfully indicates that the IR from one point i to another j experiencing exact k bounces can be calculated from the IRs experiencing exact $k - 1$ bounces and LOS IRs, as

$$\begin{aligned} g_{ij}^{(k)}(t) &\approx \sum_{l=0}^Q g_{il}^{(0)}(t) \otimes g_{lj}^{(k-1)}(t) \\ &= \frac{1}{\pi} \sum_{l=0}^Q \frac{\rho_l \cos(\phi_{il}) \cos(\theta_{il}) A_{Rl}}{R_{il}^2} \text{rect}\left(\frac{\theta_{il}}{FOV_l}\right) g_{lj}^{(k-1)}\left(t - \frac{R_{il}}{c}\right) \end{aligned} \quad (17)$$

ρ_l is the reflectivity of reflector l ; Q is the total number of reflectors in the model. By recursively using Eq. 17, we can generate the IRs experiencing arbitrary number of reflections from LOS IRs.

In this way, the accumulated IR concerning the first k orders of reflections $h_{ij}^{(k)}(t)$ is the sum of the IRs that experience exact l orders ($l = 1, 2, 3, \dots, k$):

$$h_{ij}^{(k)}(t) = \sum_{l=0}^k g_{ij}^{(l)}(t) \quad (18)$$

In Hashemi (1994), the authors show the runtime to compute $g_{ij}^{(k)}(t)$ is roughly exponential in k . At that time, it required approximately 24 h to calculate $k = 3$ bounces of IR with 2776 elements. Though in recent decades the simulation method has been optimized and the computation capacity has been upgraded, the computational complexity is still considerably high and increases significantly with reflection orders. In a practical scenario, people still have to use only first a few orders, most often three, to approximate the complete IR $h_{ij}(t)$.

Channel Features

Signal-to-noise ratio (SNR) and inter-symbol interference (ISI) are essential factors determining communication performance. The former can be estimated by received power at the receiver, while the latter can be indicated by average delay and delay spread. We therefore focus our research on power attenuation, average delay, and delay spread. For power issues, we shall keep in mind the differences between OWC systems and RF. The received optical power is proportional to the IR magnitude. Since optical power is linearly converted to the amplitude of electrical/signal current, the received electrical/signal power is proportional to the square of IR magnitude.

On the power side, to exclude the influence of ISI, we assume the transmitters are sending a unity amplitude square impulse $x(t)$ in a symbol period. For the received optical power at location j , it is calculated by

$$P_{oj} = \frac{1}{T} \int_0^T x(t) * h_{ij}(t) dt \quad (19)$$

For the received signal power at location j , it is calculated by

$$P_{sj} = \frac{1}{T} \int_0^T (x(t) * h_{ij}(t))^2 dt \quad (20)$$

On the ISI side, the average delay is the first moment of the power delay profile with respect to the first arriving path, defined as

$$\mu_j = \int_{-\infty}^{\infty} t h_{ij}^2(t) / h_{ij}^2(t) dt \quad (21)$$

The root-mean-square (RMS) delay spread is the square root of the second central moment of a power delay profile as

$$s_j = \sqrt{\int_{-\infty}^{\infty} (t - \mu_j)^2 h_{ij}^2(t) / h_{ij}^2(t) dt} \quad (22)$$

Delay spread is a good measure of multipath distortion and indicates potential ISI. Previous research shows the maximum transmission rate over a wireless channel is determined by the inverse of its delay spread, given that no diversity or equalization applied (Sexton and Pahlavan 1989; Howard and Pahlavan 1990).

Simulation and Results

We, at first, choose three typical test locations to analyze the contribution of each order of reflections to IR. The three locations are (unit: m) A (0, 0, 0.9) representing a point at the room corner, where severe diffusions are experienced; B (0, 3, 0.9)

Table 3 Room model simulation parameters

Room size, length \times width \times height (unit: m)	6 m \times 6 m \times 3 m
Laser source location (unit: m)	(3, 3, 0.5)
Diffusion transmitting spot location (unit: m)	(1.5, 1.5, 3) (1.5, 3, 3) (1.5, 4.5, 3) (3, 1.5, 3) (3, 3, 3) (3, 4.5, 3) (4.5, 1.5, 3) (4.5, 3, 3) (4.5, 4.5, 3)
Transmission power at each spot (unit: W)	1
Reflection coefficients (ceiling, wall, floor)	0.9, 0.7, 0.1
Reflection element size (unit: m \times m)	0.2 \times 0.2
Receiver FOV (unit: degrees)	60
Receiver aperture area (unit: m ²)	1e-4
Time resolution (unit: ns)	0.66

representing a point near a wall but away from corners, where medium diffusions are experienced; and C (3, 3, 0.9) representing a point at the center of the room, where weak diffusions are experienced. Next, we demonstrate the estimation accuracy distributions all over the room. The IOW system has nine transmitting spots generated from one laser source for spatially diverse purpose and the simulation parameters are given in Table 3.

High-Order Reflection's Impact to IR

Figures 6, 7, and 8 show the contributions of each order of reflections to the IRs at test locations A, B, and C. As given in Eq. 18, the total IR is the sum. In all three locations, the zeroth reflection (LOS) contributes the major impulses of the IRs, which contain most of the optical power. From the third-order reflections, the shapes of individual order IRs are similar; nevertheless, they attenuate and temporally spread out as the order increases. Compared with low-order reflections, high-order reflections contribute less amplitude but more delay to IR. Unlike received power which is only related to IR amplitude, delay spread is jointly determined by IR amplitude and delay. High-order reflections, therefore, obviously make a more significant impact to delay spread than received power. In other words, delay-spread estimation should converge slower than power with reflection orders increase.

Model Accuracy Analyses

To further explore the convergence differences, we plot the estimation accuracy of received optical power, received signal power, average delay, and RMS delay spread in Figs. 9, 10, and 11. We use the results applying the first 20 orders of reflections as references of accurate estimation. Accuracy is defined by the ratio of estimated value to referred accurate value.

The figures show that signal power and average delay converge fast and shall not be noticeably impacted by truncating them to first two or three orders; however, the received optical power and the RMS delay spread converge substantially slower. For a Gbps high-speed transmission system, delay spread deserves particular attention,

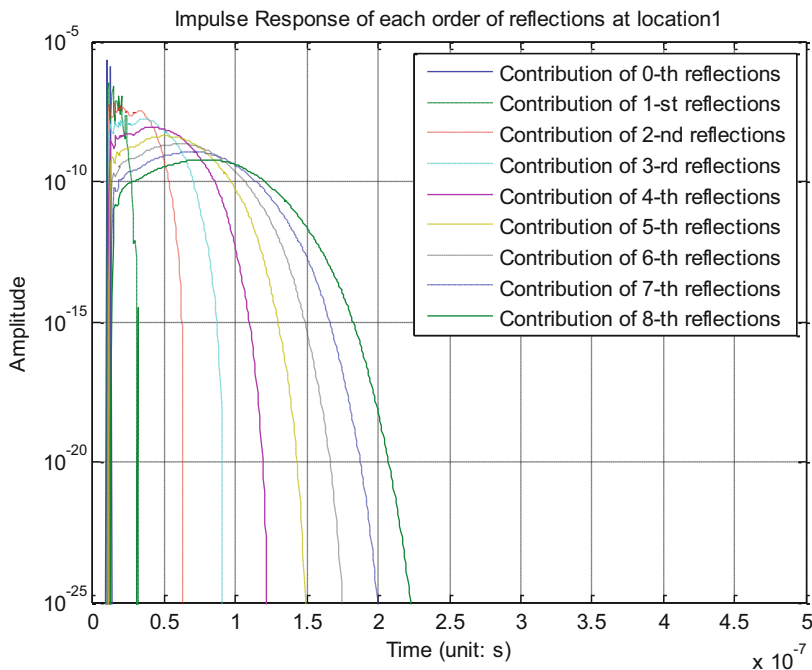


Fig. 6 Impulse response of each order of reflections at location 1

because it dominantly determines ISI. The maximum delay spread we observed all over the room is 3.66 ns. To make sure any delay-spread model error is smaller than half of a symbol period, we need the delay-spread estimation accuracy higher than $1 - \frac{1e-9}{2 \times 3.66e-9} \times 100 \% = 86.3 \%$ for 1 Gbps systems and $1 - \frac{1e-10}{2 \times 3.66e-9} \times 100 \% = 98.6 \%$ for 10 Gbps systems. As the figures show, estimation accuracies for the three locations by first three orders are only 73.4 %, 82.9 %, and 90.7 %, respectively. If we only use first three order reflections to create the model, only location C meets the accuracy needed for 1 Gbps and none of them satisfy 10 Gbps.

Delay-Spread Spatial Distributions

As IOW projects are to provide full coverage and mobility, it is necessary to extend channel model analysis from the three test points to the entire area of the room. Our research shows high-order reflections make impacts differently at different locations. The spatial distribution is explored by simulating 841 channels, representing every piece of 0.2 m × 0.2 m area of a 6 m × 6 m × 3 m room. As we demonstrated that delay spread experiences most severe impact from discarded high-order reflections, we utilize delay-spread estimation accuracy to indicate model accuracy as the worst case. Its contours for each additional order of reflections considered are shown in Figs. 12, 13, 14, 15, 16, 17, 18, and 19, respectively. Generally, the shapes of the contours are similar as the number of reflections increases: high-accuracy areas are

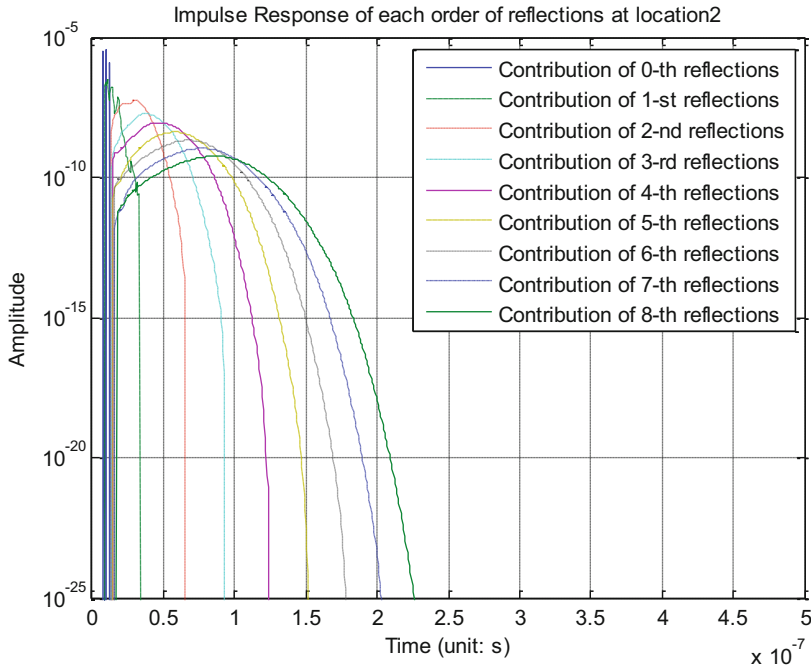


Fig. 7 Impulse response of each order of reflections at location 2

near the center of the room, and the accuracy decreases when approaching the corners of the room, where multipath effect is much richer; we also observe that dense contours exist at the corner of the room, which indicates a sharp decrease of model accuracy. We are able to get the required reflection orders for specific transmission rate from the contours. For instance, to ensure the accuracy of the entire room is above the need for 1 Gbps data rate, we need 5 orders (86.3 % above accuracy guaranteed) of reflections and 9 orders (98.6 % above accuracy guaranteed) for 10 Gbps.

Discussion on the Model Error and Calibration Method

As discussed in section “[Historical Review of Indoor Optical Wireless Channel Research and Error Analyses](#),” the computational complexity increases considerably with number of reflection orders included. It is reasonable that many researchers neglect high-order reflections to enable the model computation be conducted in a practical amount of time. As researchers are working on faster OWC systems than ever before, this sacrifice of accuracy for efficiency results in more and more significant performance errors. A reasonable and feasible solution to keep both accuracy and efficiency is applying calibration. We propose a calibration method based on the statistic data of the model accuracy curves. The model accuracy curves

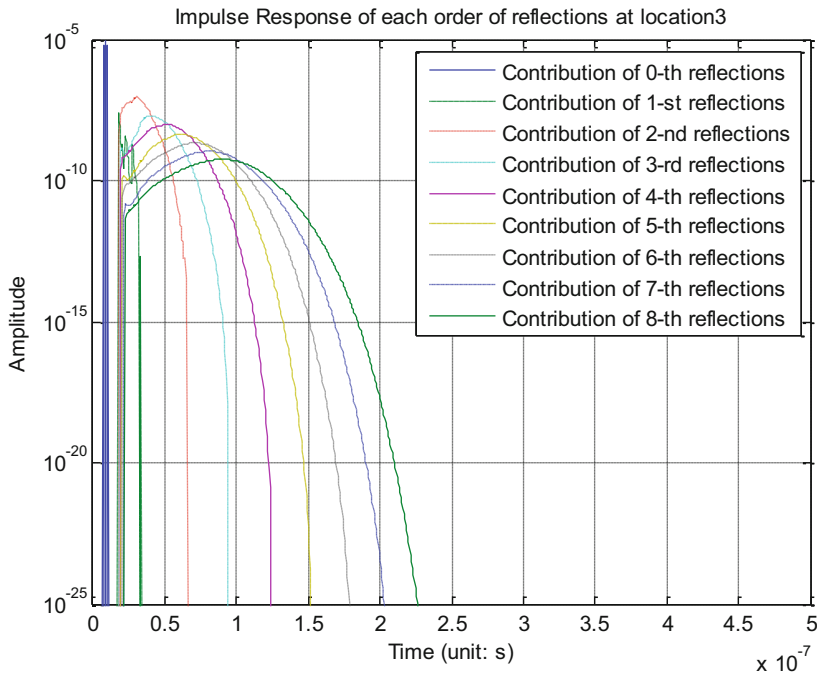


Fig. 8 Impulse response of each order of reflections at location 3

of all 841 locations are drawn in Fig. 20. By averaging them, we obtain the general calibration curve $c[k]$. It provides the correction value for each order of reflections. For the delay spread calculated from the first k orders of reflections, the value can be calibrated by

$$\tilde{s}_j^{(k)} = s_j^{(k)} / c[k] \tag{23}$$

where $\tilde{s}_j^{(k)}$ and $s_j^{(k)}$ are the post-calibrated and pre-calibrated delay spread from the first k orders of reflections, respectively (Fig. 21).

The calibration value for each order of reflections is given in Table 4.

We apply the RMS model error to compare the performance before and after calibration as in Fig. 20. These are calculated by Eqs. 24 and 25, respectively:

$$e^{(k)} = \frac{1}{N} \sum_{j=1}^N \left(\frac{s_j^{(k)} - s_j}{s_j} \right)^2 \tag{24}$$

$$\tilde{e}^{(k)} = \frac{1}{N} \sum_{j=1}^N \left(\frac{\tilde{s}_j^{(k)} - s_j}{s_j} \right)^2 \tag{25}$$

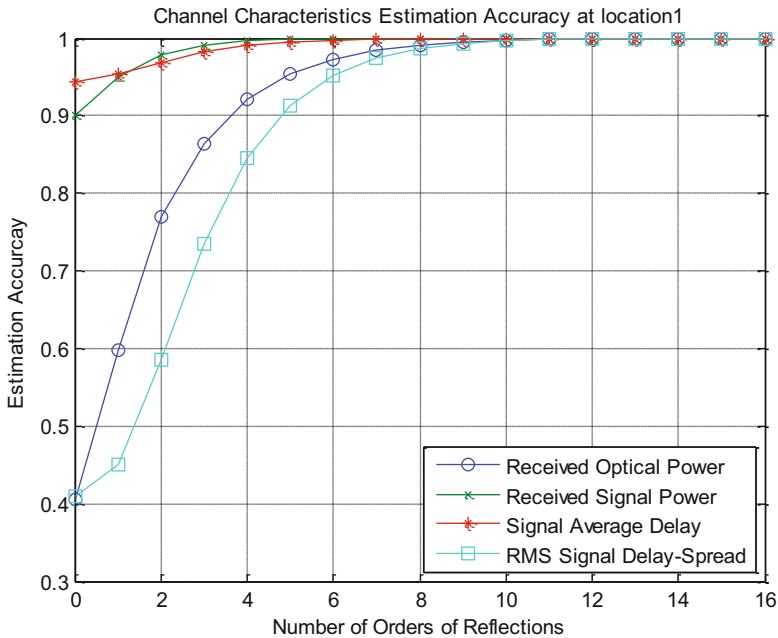


Fig. 9 Channel characteristic estimation accuracy at location 1

where N is the total number of channels tested; $e^{(k)}$ is the pre-calibrated estimation error; $\tilde{e}^{(k)}$ is the post-calibrated estimation error; and s_j is the reference of accurate estimation. As we can see, there is a substantial decrease in delay-spread model error after calibration. In the model applying the first three orders of reflections, the average RMS error drops from 15.7 % to 4.3 %. We draw the contour for the calibrated model accuracy as in Figs. 22 and 23. It can be discovered that the order number of reflections needed for 1 Gbps systems reduces from 5 to 3 and from 9 to 7 for 10 Gbps systems.

Summary

This section extensively discusses the impact of high-order light reflections on IOW channel model error. The results indicate that conventional channel models based on a first few, most commonly three, orders of reflections are not able to provide sufficient model accuracy for contemporary research over Gbps high-speed IOW systems. The spatial distribution of model error and the impact of each additional order are explored in details. Based on their findings, the authors develop a calibration approach. It successfully mitigates RMS delay-spread error and keeps efficiency simultaneously.

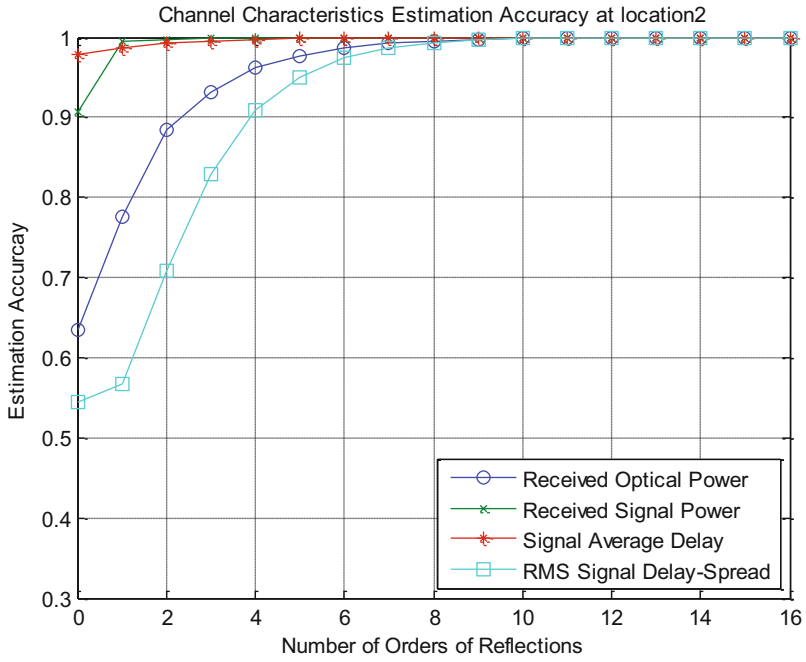


Fig. 10 Channel characteristic estimation accuracy at location 2

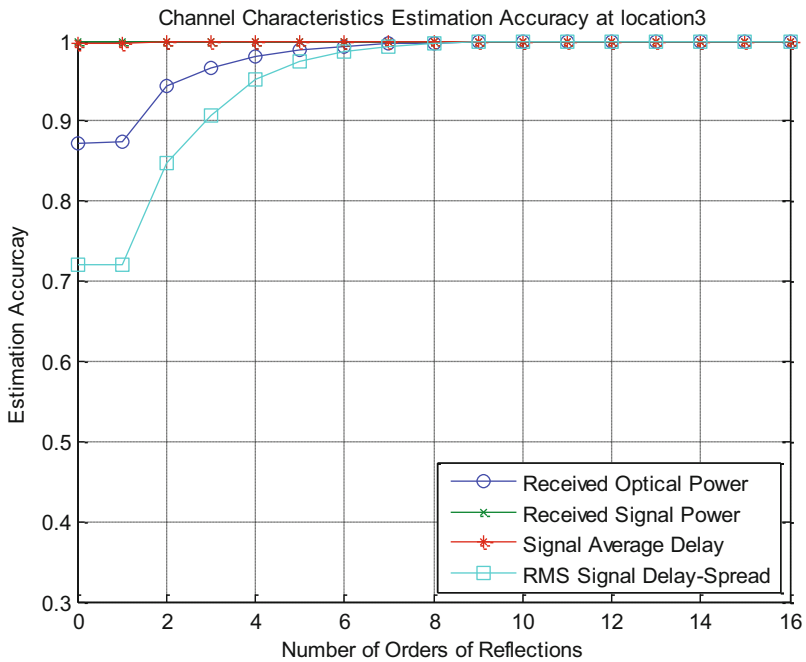


Fig. 11 Channel characteristic estimation accuracy at location 3

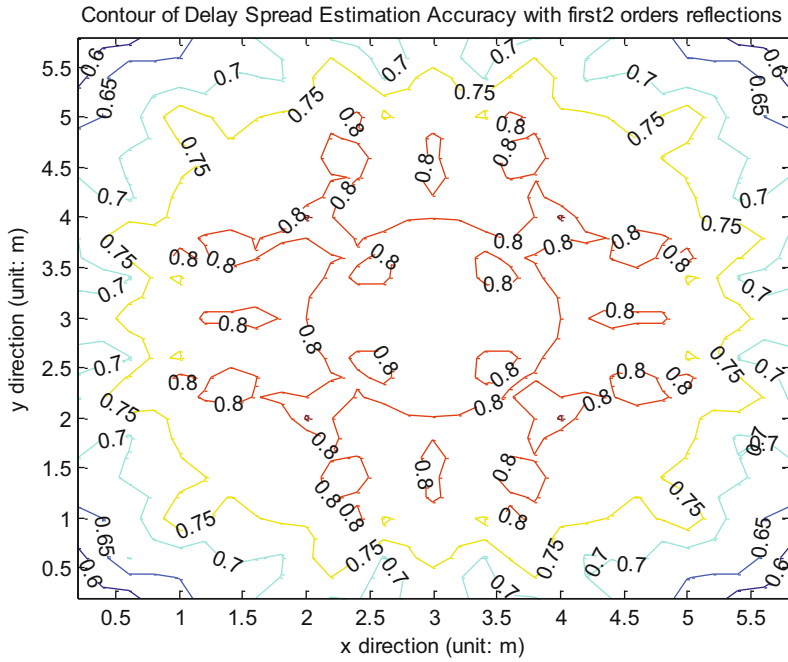


Fig. 12 Delay-spread model accuracy contour considering first 2 orders of reflections (%)

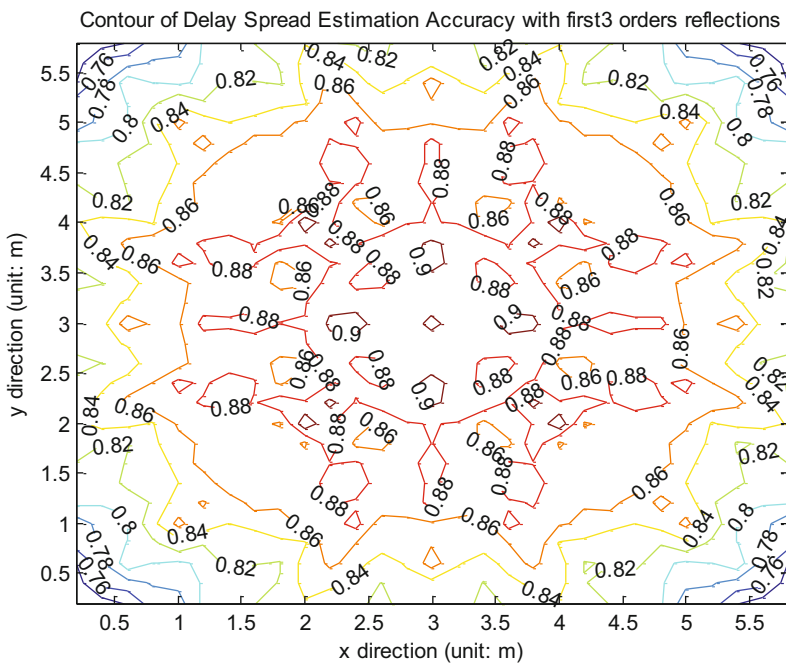


Fig. 13 Delay-spread model accuracy contour considering first 3 orders of reflections (%)

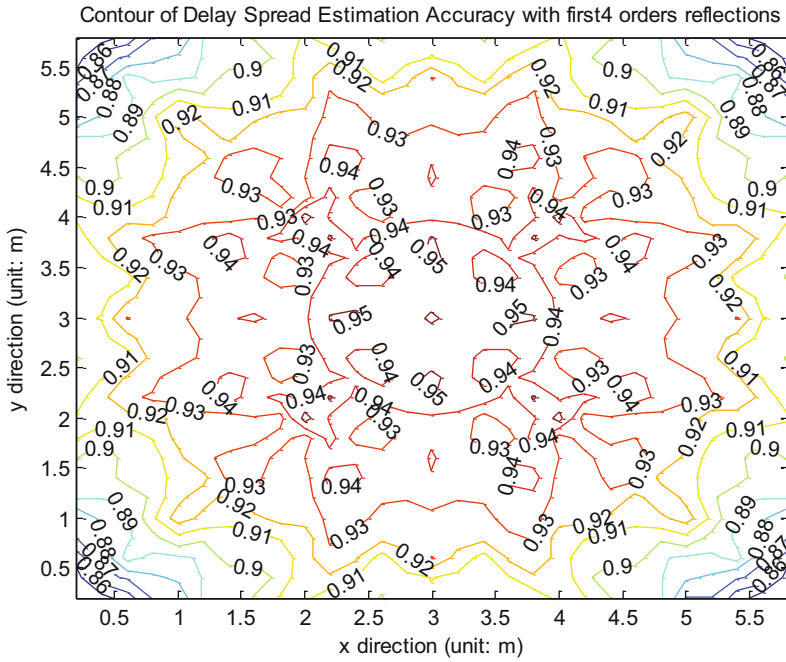


Fig. 14 Delay-spread model accuracy contour considering first 4 orders of reflections (%)

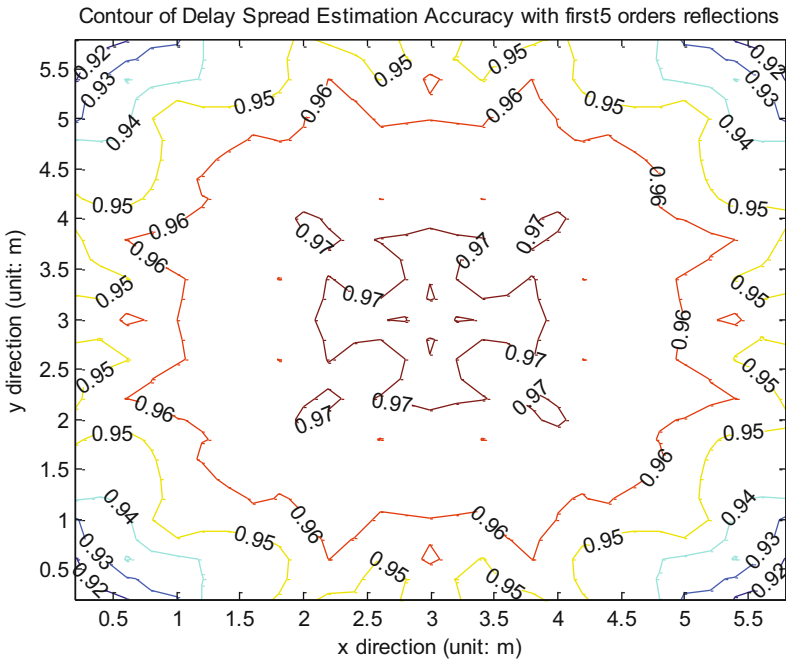


Fig. 15 Delay-spread model accuracy contour considering first 5 orders of reflections (%)

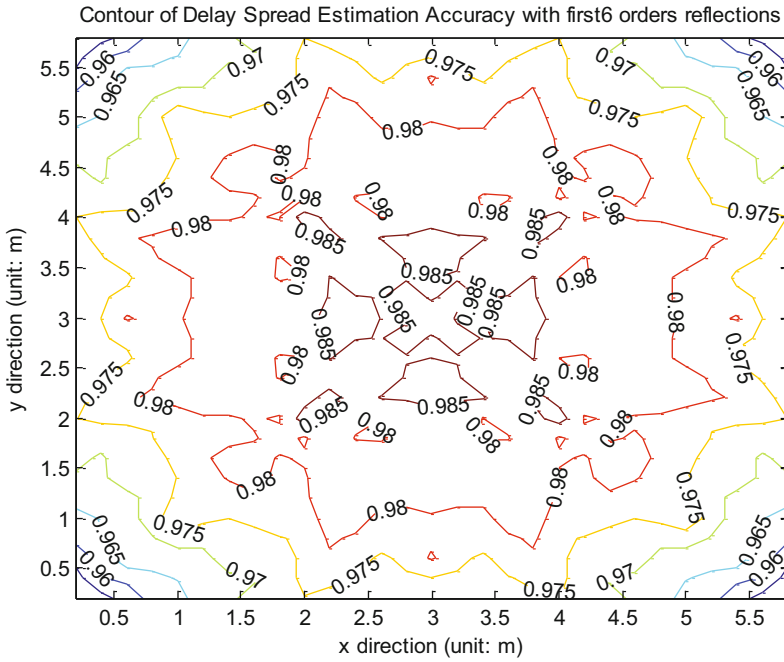


Fig. 16 Delay-spread model accuracy contour considering first 6 orders of reflections (%)

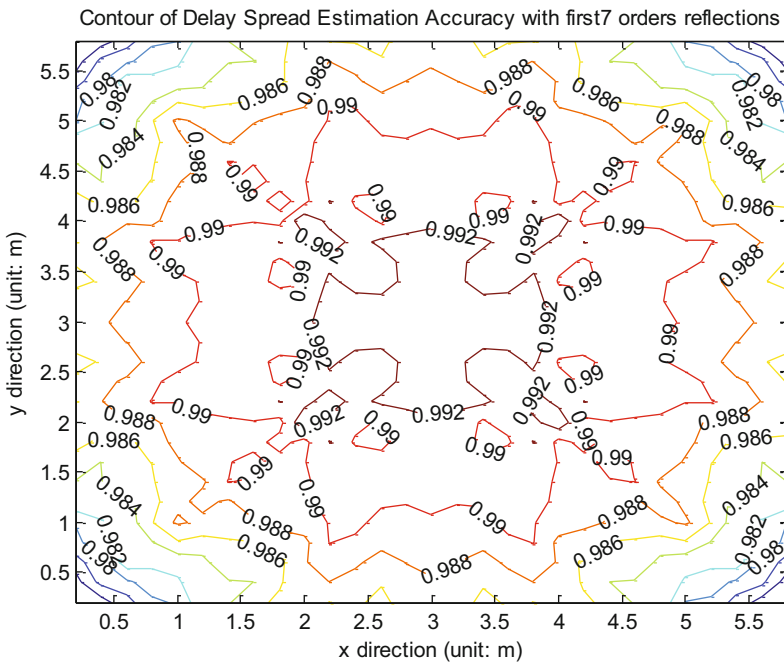


Fig. 17 Delay-spread model accuracy contour considering first 7 orders of reflections (%)

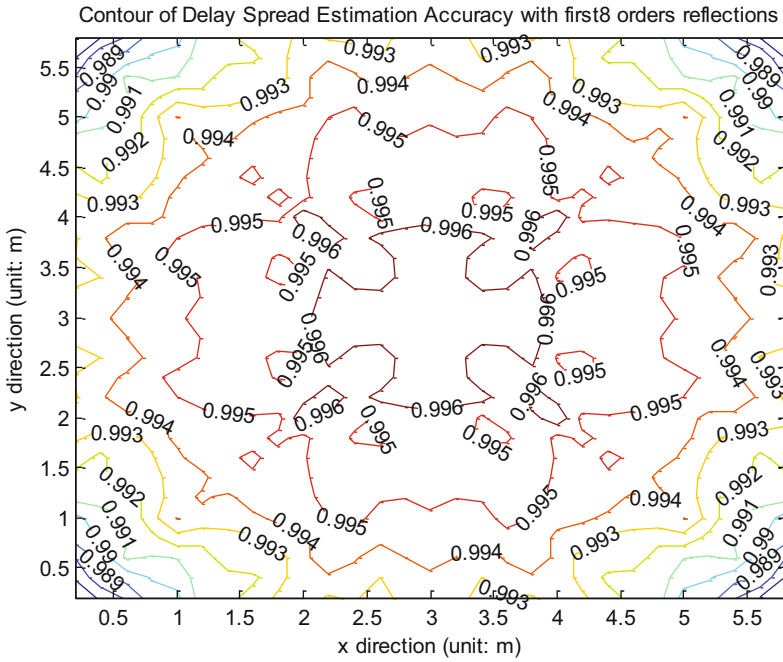


Fig. 18 Delay-spread model accuracy contour considering first 8 orders of reflections (%)

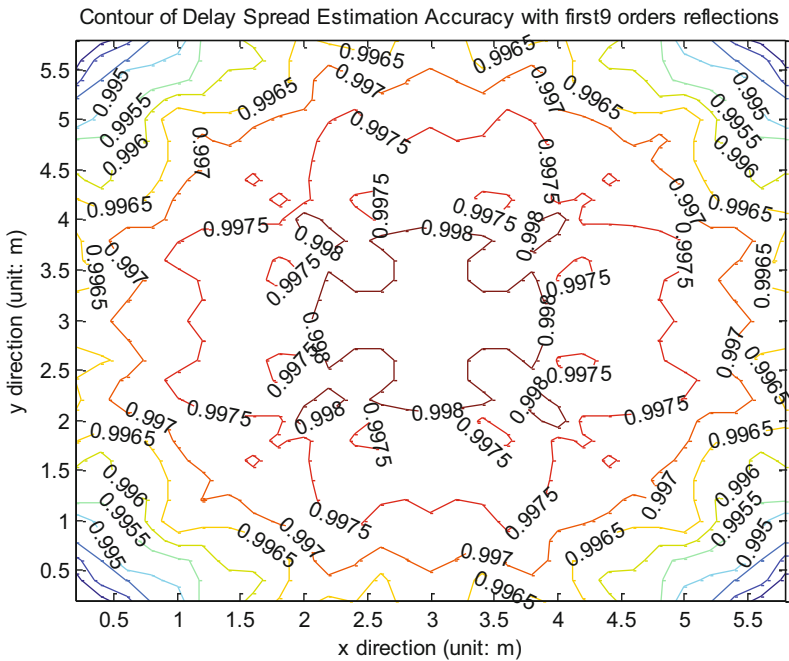


Fig. 19 Delay-spread model accuracy contour considering first 9 orders of reflections (%)

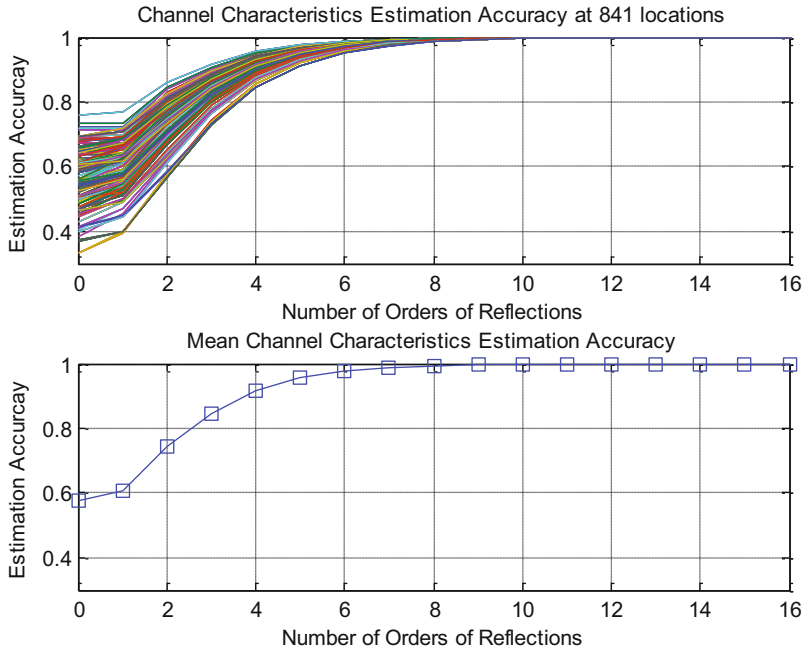


Fig. 20 Channel estimation accuracies for all 841 locations and average

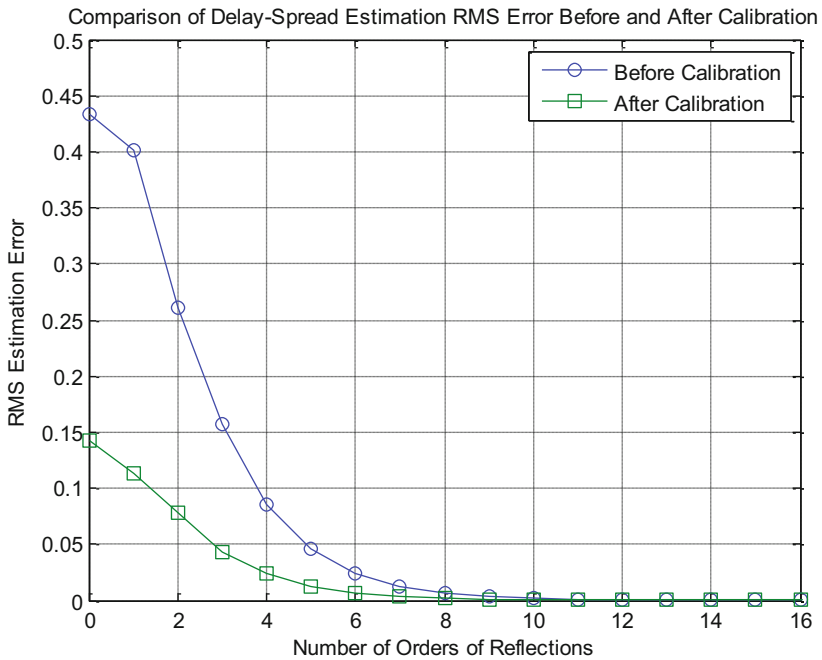


Fig. 21 Comparison of average RMS model error before and after calibration

Table 4 Calibration values for channel model

Orders	1	2	3	4	5	6
Calibration (%)	60.53	74.53	84.71	91.77	95.52	97.67
Orders	7	8	9	10	11	12
Calibration (%)	98.79	99.38	99.69	99.84	99.92	99.96

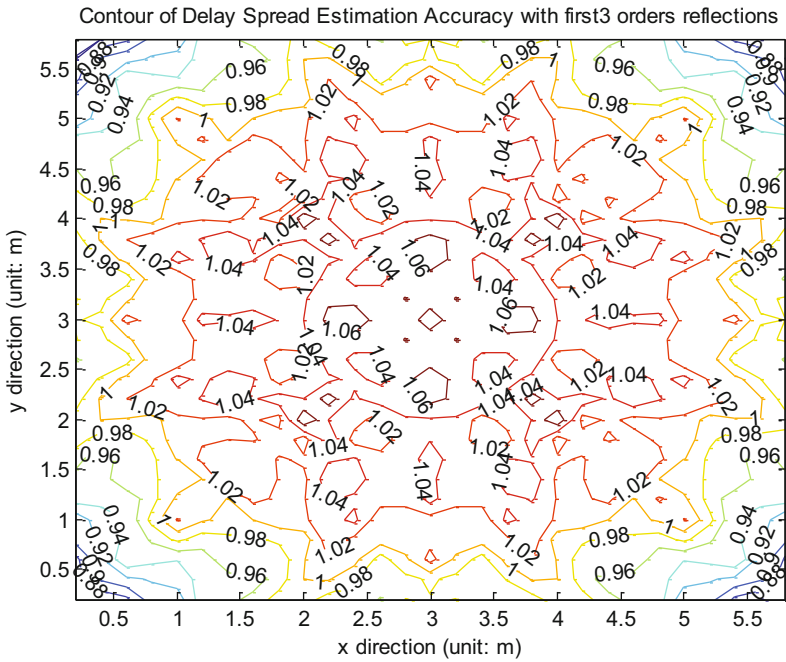


Fig. 22 Calibrated delay-spread model accuracy contour considering first 3 orders of reflections (%)

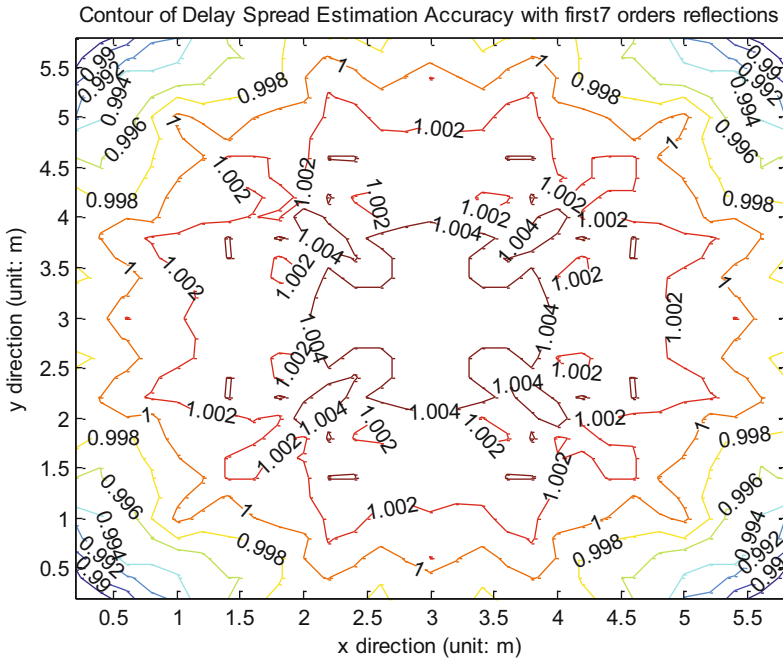


Fig. 23 Calibrated delay-spread model accuracy contour considering first 7 orders of reflections (%)

References

- Alqudah Y, Kavehrad M (2003) MIMO characterization of indoor wireless optical link using a diffuse-transmission configuration. *IEEE Trans Commun* 51(9):1554–1560
- Anand M, Mishra P (2010) A novel modulation scheme for visible light communication. In: *Proceedings of India Conference (INDICON)*. pp 1–3
- Barros DJF, Wilson SK, Kahn JM (2012) Comparison of orthogonal frequency-division multiplexing and pulse-amplitude modulation in indoor optical wireless links. *IEEE Trans Commun* 60(1):153
- Barry J, Kahn J et al (1993) Simulation of multipath impulse response for indoor wireless optical channels. *IEEE J Sel Areas Commun* 11(3):367–379
- Center for Ubiquitous Communication by Light <http://www.uclight.ucr.edu/>

- Cui K, Cheng G, Xu Z (2010) Line-of-sight visible light communication system design and demonstration. In: 7th communication systems networks and digital signal processing
- Fadlullah J, Kavehrad M (2010) Indoor high-bandwidth optical wireless links for sensor networks. *J Lightwave Technol* 28(21):3086–3094
- Gfeller FR, Bapst UH (1979) Wireless in-house data communication via diffuse infrared radiation. *Proc IEEE* 67(11):1474–1486
- Grubor J, Randel S et al (2008) Broadband information broadcasting using led-based interior lighting. *J Lightwave Technol* 26(24):3883–3892
- Hashemi H (1994) Statistical modeling and simulation of the rms delay spread of indoor radio propagation channels. *IEEE Trans Veh Technol* 43(1):110–120
- Home Gigabit Access <http://www.ict-omega.eu/>
- Howard SJ, Pahlavan K (1990) Performance of a DFE modem evaluated from measured indoor radio multipath profiles. In: *ICC '90 communications*, vol 4. pp 1341–1345
- Kahn JM, You R et al (1998) Imaging diversity receivers for high-speed infrared wireless communication. *IEEE Commun Mag* 36(12):88
- Kavehrad M (2010) Sustainable energy-efficient wireless applications using light. *IEEE Commun Mag* 48:66–73
- Kavehrad M, Fadlullah J (2010) Wideband optical propagation measurement system for characterization of indoor optical wireless channels. *Proc SPIE* 7620:7620 0E
- Kavehrad M, Jivkova S (1999) Indoor wireless infrared local access, multi-spot diffusing with computer generated holographic beam-splitters. *IEEE Int Conf Commun* 1:604–608
- Kim J, Lee D, Kim K, Park Y (2010) Performance improvement in visible light communication by using spread spectrum coding. In: *OptoElectronics and communication conference (OECC) 15th*. p 278
- Kishi T et al (2013) A high-speed LED driver that sweeps out the remaining carriers for visible light communications. *J Lightwave Technol* 32(2):239–249
- Komine T, Nakagawa M (2004) Fundamental analysis for visible-light communications system using LED lights. *J IEEE Trans Consum Electron* 50(1):100
- Lee YU, Kavehrad M (2012) Two hybrid positioning system design techniques with lighting LEDs and ad-hoc wireless network. *IEEE Trans Consum Electron* 58:1176
- Lee JS, Su YW, Shen CC (2007) A comparative study of wireless protocols: bluetooth, UWB, ZigBee, and Wi-Fi. In: *Industrial Electronics Society, 2007. IECON 2007. 33rd annual conference of the IEEE*
- NSF Center on Optical Wireless Applications <http://cowa.psu.edu/>
- Park H, Lee K (2011) Modulations for visible light communications with dimming control. *J IEEE Photon Technol Lett* (99)
- Sexton TA, Pahlavan K (1989) Channel modeling and adaptive equalization of indoor radio channels. *IEEE J Sel Areas Commun* 7(1):114–121
- Visible Light Communications Consortium <http://www.vlcc.net/>
- Vucic J, Kottke C et al (2010) 513 Mbit/s visible light communications link based on DMT-modulation of a white LED. *J Lightwave Technol* 28(24):3512
- Xu Z, Sadler BM (2008) Ultraviolet communications: potential and state-of-the-art. *IEEE Commun Mag* 46(5):67–73
- Yun G, Kavehrad M (1992) Spot diffusing and fly-eye receivers for indoor infrared wireless communications. In: *Proceedings of IEEE wireless communications conference, Vancouver, Canada*
- Zhou Z, Kavehrad M, Deng P (2012) Energy efficient lighting and communications. In: *Proceedings of the SPIE 8282, broadband access communication technologies VI, San Francisco, CA*

Indoor Localization and Applications

Shinichiro Haruyama

Contents

Introduction	666
Indoor Localization Technologies	666
Indoor Localization Using Visible Light Communication	667
Existence Checking Method	668
RSSI Method	669
PDOA Method	670
TDOA Method	671
AOA Method Using Image Sensor	672
Applications of Indoor Localization Using Visible Light Communication	675
Indoor Navigation Using Visible Light Communication	675
Location-Based Advertising Using Visible Light Communication	676
Robot Control Using Visible Light Communication	678
Accurate Measurement of Architectural Structure Using Visible Light Communication	678
Standardization	680
Future Directions	681
References	682

Abstract

Visible light communication can be used for indoor localization by sending location information from LED lights. Several methods of such indoor localization have been proposed. The indoor localization using visible light communication will make it possible to realize many new applications including indoor

S. Haruyama (✉)

Graduate School of System Design and Management, Keio University, Kohoku-ku, Yokohama, Japan

e-mail: haruyama@sdm.keio.ac.jp

navigation, location-based advertising, robot control, and accurate measurement of architectural structure.

Introduction

LED lights have recently been used widely as efficient light sources replacing incandescent light bulbs and fluorescent lamps. Because of the widespread use of current LED lights all over the world, LED lights can also be used as ubiquitous visible light communication transmitter. Furthermore, if the data transmitted from LED lights is position information, various indoor localization services will become a reality.

Visible light communication has properties that are both advantageous and disadvantageous compared with radio-wave wireless communication. Some of its advantages are freedom from communication regulation and health safety. It is free from communication regulation, because the frequency above 3 THz is not currently regulated by the radio regulation law. Besides, the visible light usually poses no health hazards to human body and eyes. Disadvantages, on the other hand, include short communication distance and relatively slow data rate. The communication distance using visible light communication is typically between 1 and 100 m. This distance is short compared with radio-wave communication, due to the fact that visible light communication is basically line-of-sight communication, which means that communication is interrupted when there is an object between a transmitter and a receiver. There is another disadvantage of visible light communication, which is data rate. Its data rate is typically up to 10 megabits per second, although there have been active researches going on to reach the speed of gigabits per second (Kottke et al. 2012). The bottle neck of the data rate is mainly caused by the slow reaction time of the yellow phosphor used for white LEDs. The above disadvantageous properties of visible light communication may limit its use for some applications. However, these seemingly disadvantageous properties are indeed useful for some applications by taking advantage of line-of-sight property. Its property of visible light communication results in the short communication distance, which means that a transmitter and a receiver are in the close vicinity. A representative application using its line-of-sight property is indoor localization. Indoor localization measures the position of a user or an object, and the localization technology makes it possible to offer various location-based services such as navigation, advertising, etc.

Indoor Localization Technologies

There are several wireless technologies that have been developed for indoor localization applications. The wireless methods used for indoor localization technologies include radio wave, sound wave, and optical signal (Gu et al. 2009). These wireless methods will be described in more details below.

Representative indoor localization technologies using radio wave are WLAN (wireless local area network)-based positioning, UWB (ultra wide band) positioning, and bluetooth positioning. WLAN-based positioning senses the received signal strength indication (RSSI) values of the transmitted WLAN signals. The sensing is done either at the access points or at a user's terminal. This method is inexpensive and flexible because almost all the terminals such as smart phones have WLAN transceivers, and there is no need to install additional equipment. The WLAN-based positioning, however, suffers from the multi-path problem of radio wave where radio signals reflected by walls distort the original radio signals. The accuracy of WLAN positioning is in the range of several meters to tens of meters, which depends on the environment where WLAN access points are installed. UWB positioning uses short radio wave pulses. Its pulses have a short duration of less than 1 ns, which makes it possible to filter the reflected signals from the original signal. Because of this capability, the accuracy of UWB positioning is in the range of tens of centimeters, which is more accurate than that of WLAN-based positioning. Bluetooth positioning is similar to WLAN-based positioning in a sense that the bluetooth devices are embedded in many personal equipments such as mobile phone, smart phone, laptop, and desktop PC. The bluetooth chipsets are also low cost which will make it possible to construct a low cost positioning system. The accuracy of bluetooth positioning is in the range of several meters. "iBeacon" of Apple Inc. uses bluetooth low energy (BLE) technology for positioning, which allows its device to run for months or years with a battery.

The indoor localization technology of sound wave uses ultrasound which cannot be heard by humans. Typical ultrasound positioning uses a short pulse of ultrasound. A transmitter emits the ultrasound pulse which is received by a microphone of a receiver. Either RSSI method or time of arrival (TOA) method is used. The RSSI method measures the strength of received sound signal, and the TOA method measures how much time is needed for the ultrasound pulse to arrive at the receiver. The ultrasound localization technology suffers from reflected ultrasound signals from walls and other sound noise sources similar to WLAN-based positioning method. The accuracy of ultrasound positioning is in the range of several centimeters to several meters.

Indoor Localization Using Visible Light Communication

The indoor localization technology of optical signal uses either infrared or visible light. The difference between infrared and visible light is that infrared light cannot be noticed by a user, while visible light is literally visible to human. Since LED lights are becoming less and less expensive and the percentage of LED lights in public spaces, offices, and homes is increasing, LED lights can be more readily used as an infrastructure for indoor localization. Thus visible light communication is more suitable for ubiquitous indoor localization than infrared communication.

Figure 1 shows a representative use of visible light communication for indoor localization, where an LED light is used as a data transmitter and a terminal with visible light sensor is used as a data receiver. A user is able to know not only the

room location where he/she is in but also which position in the room he/she is at. This application is especially useful indoor because GPS receivers do not work well indoors even though they work well outdoors.

Several indoor localization methods have been proposed: existence checking, RSSI, PDOA, TDOA, and AOA. Each of these methods will be explained in the following sections.

Existence Checking Method

Existence checking method is the simplest localization method among the methods described in this section. The existence checking method checks if there exists a signal from visible light transmitter as shown in Fig. 1. In other words, a terminal of a user checks if it receives a signal or an ID from an LED light above a user and if it finds a signal, the user is located within the radiation range of the LED light. This method of existence checking is simple and straightforward, because it does not need any coordination or synchronization among LED lights. The disadvantages of existence checking method is that it is difficult to calculate the accurate distance from the LED light whose signal reached the terminal. This method is useful for applications where rough estimation of a user's location is good enough. The accuracy of position using the existence checking depends on the radiation pattern of an LED light and also on the height of the LED light from the floor, but it ranges from tens of centimeters to several meters. This method was used by Nakajima and Haruyama (2012) for the navigation of the visually impaired.

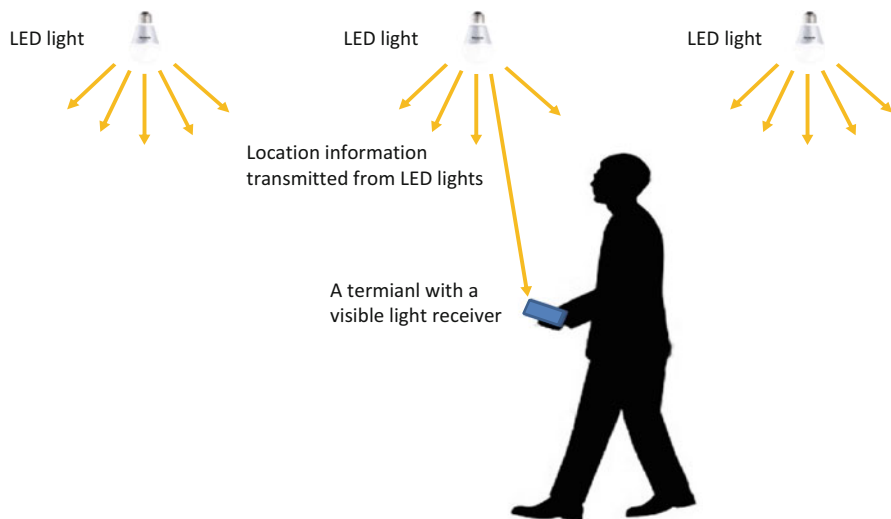


Fig. 1 Representative use of visible light communication for indoor localization

RSSI Method

RSSI (received signal strength indicator) indicates the power received by a receiver. The RSSI is typically used for radio communication, but it can also be used as an indicator for the optical power received by an optical receiver. The RSSI can be used to estimate the distance between a transmitter and a receiver (Zhou et al. 2012), because in general the longer the distance is, the smaller the RSSI is. A typical model of a light source and a receiver is shown in Fig. 2, where a light source has a generalized Lambertian radiation pattern and a receiver has a flat photo detector without a lens (Barry 1994).

The generalized Lambertian radiation pattern of a light source $R(\varphi)$ is expressed as:

$$R(\varphi) = \frac{(m + 1) \cos^m(\varphi)}{2\pi}$$

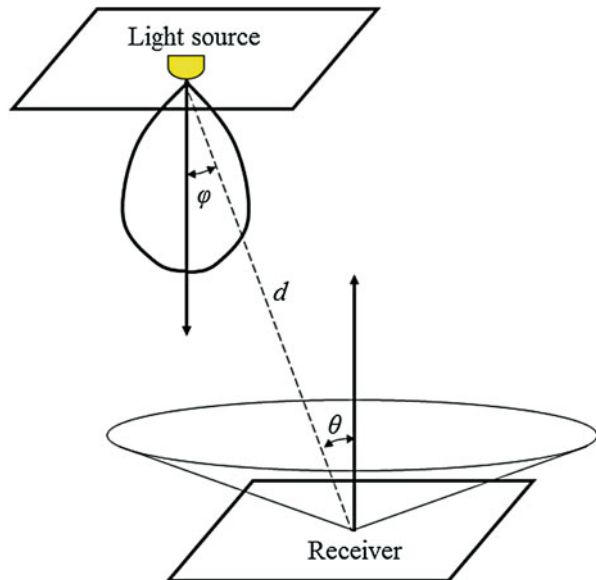
where φ is the angle of radiation from a light source and m is the order of Lambertian emission.

The line-of-sight optical power P received by a photo detector is:

$$P = R(\varphi) \frac{A}{d^2} \cos(\theta) = \frac{(m + 1) \cos^m(\varphi) A}{2\pi d^2} \cos(\theta)$$

where A is the area of a photo detector, d is the distance between a light source and a receiver, and θ is the angle of incidence.

Fig. 2 Model of a light source with a generalized Lambertian radiation pattern and a photo detector without a lens



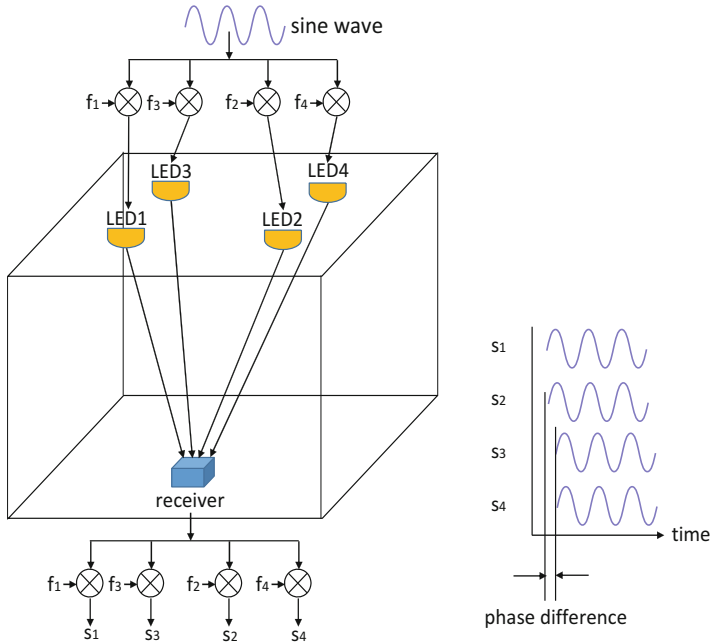


Fig. 3 Phase difference of arrival method

The problem of applying the RSSI for free-space optical communication is that the RSSI depends not only on the distance d between a transmitter (a light source) and a receiver but also on the radiation pattern $R(\varphi)$ of LED source, the angle of radiation from a light source φ , as well as the angle of incidence θ of a receiver. If non-line-of-sight reflection is considered, the calculation of the power received by a photo detector will become even more complex. Because of these problems, the estimation of the distance between a transmitter and a receiver is not so accurate.

PDOA Method

PDOA (phase difference of arrival) method detects the phase difference between two LED transmitters, each of which sends sinusoidal signals from different locations (Jung et al. 2011; Nah et al. 2013). Figure 3 shows the basic concept of the phase difference of arrival method. A sine wave is multiplied by multiple carrier frequencies in order to generate multiple carrier signals for each LED light. The multiplied signal is intensity modulated that results in the modulated optical signal. A receiver receives signals from multiple LED lights and does direct detection of the incoming optical signal. After the signal is multiplied by multiple carrier frequencies, the original sine wave for each LED light is independently recovered. The phase difference between any two of the sine waves indicates the time difference of arrival

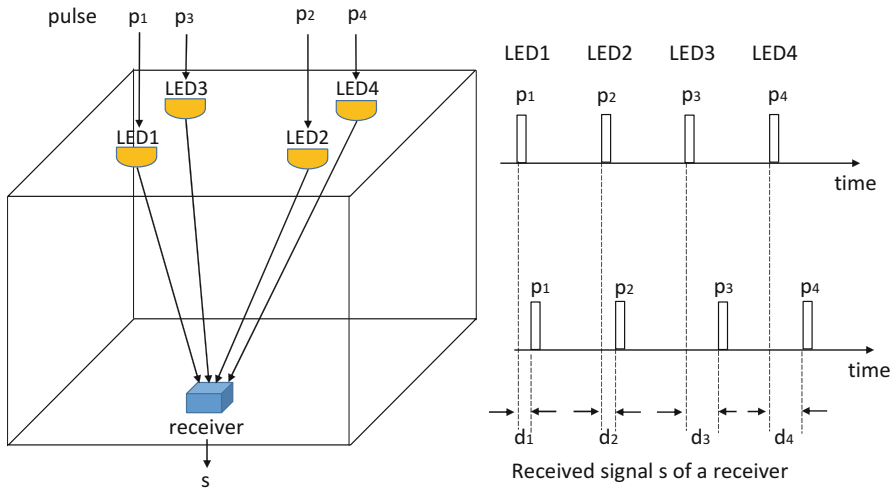


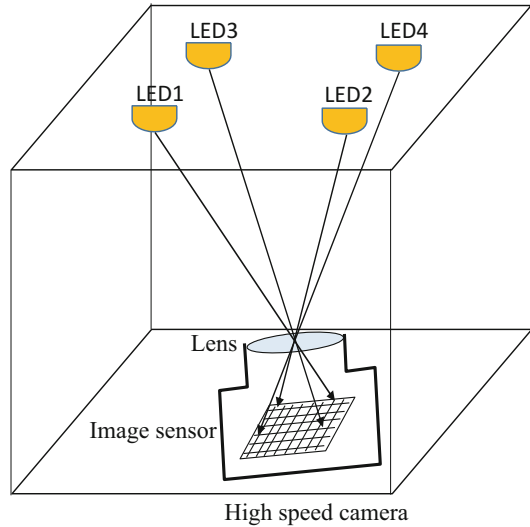
Fig. 4 Time difference of arrival method

from the two different LED lights. By multiplying it with the speed of light, the three-dimensional distance difference between the two LED lights can be calculated. This method requires that all optical signals from the LED lights have to be synchronized with the same sine wave, which will be accomplished by wiring them with the same clock signal. The PDOA method needs the up-conversion using different frequencies for different LED lights and the down-conversion using different frequencies at the receiver, in order to distinguish multiple LED lights.

TDOA Method

The TDOA (time difference of arrival) method is simpler than the PDOA method, because TDOA do not need complicated up-conversion and down-conversion processes. Figure 4 shows the basic concept of the time difference of arrival method (Do et al. 2013). A pulse signal is used to generate optical signal for each LED light. The pulses of LED lights 1, 2, 3, and 4 are generated one by one in this sequence. Each LED light must generate an optical pulse with a constant interval. A receiver receives signals from multiple LED lights and does direct detection of the incoming optical signal. The multiple pulses from multiple LED lights are received by a receiver. The intervals between multiple pulses are observed. If the distance between a receiver and each LED light is the same, then the intervals of pulse sequence will be the same. However, if the distance between a receiver and each LED light is different, the intervals will be different. Thus, by analyzing the differences of time intervals among the pulses, it is possible to calculate the distance differences between the receiver and LED lights. This TDOA method requires that all optical signals from the LED lights have to generate pulses at different timing with a constant interval, which

Fig. 5 Angle of arrival method



will be accomplished by wiring them with the same clock signal, just like PDOA. However, the TDOA does not need complicated up-conversion and down-conversion and can be more easily implemented with a simpler circuit.

AOA Method Using Image Sensor

The AOA (angle of arrival) method detects not only the data transmitted from a LED light but also the three-dimensional vector from the LED light to a receiver. In order to obtain the accurate angle of arrival, an image sensor is typically used as shown in Fig. 5.

An image sensor is able to do simultaneous image acquisition and data reception, which conventional photo diode cannot do. Another advantage of an image sensor as receiver over a photo diode is that even if there is a strong interfering light along with a desired signal, the interfering light will be focused onto a pixel which is different from a pixel onto which a desired signal is focused. This implies that image sensor reception is much more robust against interference than photo diode reception.

Image sensors used for digital cameras or video cameras usually have frame rate of tens of frames per second. If a visible light signal from a visible light LED is received at a pixel of such an image sensor, the data rate is on the order of only several bits per second. However, using a high-speed image sensor, it is possible to achieve data rate on the order of kilo bits or even mega bits per second (Takai et al. 2013). Recently there are some proposals to increase the data rate of image sensor reception even if its frame rate is on the order of tens of frames per second. Ryan et al. (2013) uses the property of a rolling shutter of a digital camera. A rolling shutter is a mechanism to read out each scanline serially from the top to the bottom,

Method	Minimum number of reference points	Restrictions
Direct Linear Transformation	6	Points must not be on the same plane
Planar Homography	4	Points must be on the same plane
Analytical Perspective-n-Point	3	Uniqueness of the solution is not guaranteed
Linear Registration with gravity sensor	2 or 3	2 points if they are on co-horizontal plane and 3 points if they are not co-horizontal plane

Fig. 6 Comparison of single view pose estimation methods

which means that it is possible to obtain temporal change of the light by observing the data at different scanlines. Since the sampling speed of the rolling shutter is faster than the frame rate of the image sensor, it is possible to receive data at the speed much faster than the frame rate.

The image sensor is thus able to not only receive data but also detect the position of the pixel onto which a LED light is projected. After the position of the pixel is identified, it is possible to calculate the accurate incoming vector from the LED light to a receiver. In addition, it is possible to calculate the pose of a receiver with computer vision techniques (Szeliski 2011).

Single view pose estimation is performed in order to solve the pose of a camera using only one image. There are several methods of single view pose estimation which are listed in Fig. 6.

Direct linear transformation (Hartley and Zisserman 2004) is a representative algorithm to find the relation between three-dimensional reference points and their projection onto the image plane of a pinhole camera. At least six reference points are needed to estimate the pose of a camera for direct linear transformation. Planar homography (Hartley and Zisserman 2004) is an algorithm to find the relation between reference points on a plane and their projection onto the image plane of a pinhole camera. As the name suggests, reference points must be on the same plane, and under that restriction, minimum of four reference points are needed to estimate the pose of a camera. Analytical Perspective-n-Point (Haralick et al. 1994) is an algorithm to obtain a camera pose by analytically solving nonlinear equations of a Perspective-n-Point problem. The solution of Analytical Perspective-n-Point is not unique when the number of the reference points is three because of the nonlinearity of equations. Linear registration with gravity sensor (Kotake et al. 2005) is an algorithm to combine direct linear transformation with gravity sensor such as an accelerometer. The algorithm of registration with gravity sensor is based on linear operations and uses an accelerometer to find the vector of gravity of the earth. The registration with gravity sensor needs only two reference points if the two reference points are on co-horizontal plane, which means that the height of the two reference points are the same. Tanaka and Haruyama (2009) used this method with the same height for all the LED lights, so that only two LED lights are needed to calculate the

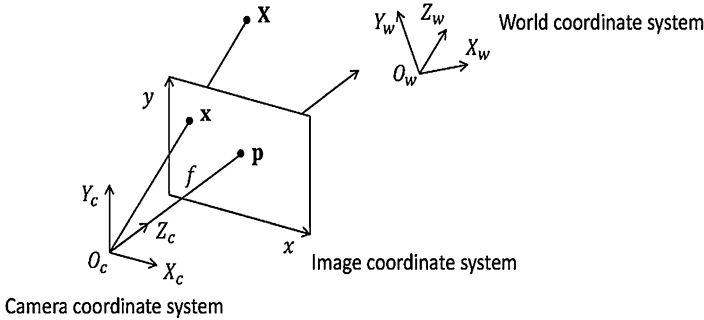


Fig. 7 Single view geometry

pose of a camera. If the height of reference points is not the same, then the registration with gravity sensor needs three reference points. Details of the direct linear transformation, which is one of the representative single view pose estimation method, will be described in more detail below.

As shown in Fig. 7, single view geometry is described with three coordinate systems: 3D world coordinate system, 2D image coordinate system, and 3D camera coordinate system. A camera pose is normally defined as the position and orientation of the camera coordinate system relative to the world coordinate system. The mathematical representation of the position and orientation is equivalent to the parameters of geometric transformation from the world coordinate system to the camera coordinate system as

$$\tilde{\mathbf{X}}_c = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \tilde{\mathbf{X}}_w$$

where $\tilde{\mathbf{X}}_c = (X_c, Y_c, Z_c, 1)^T$ is a homogeneous camera coordinate, $\tilde{\mathbf{X}}_w = (X_w, Y_w, Z_w, 1)^T$ is a homogeneous world coordinate, \mathbf{R} is a 3×3 rotation matrix (orientation), \mathbf{t} is a 3×1 translation vector (position). Therefore, a camera pose is equivalent to $[\mathbf{R}|\mathbf{t}]$.

The camera coordinate system is defined as the system in which its origin is located at the camera center and the direction of Z_c is perpendicular to the image plane from the camera center. The intersection of the image plane and Z_c axis is especially called principal point $\mathbf{p} = (p_x, p_y)$. In the pinhole camera model, a 3D point $\mathbf{X}_c = (X_c, Y_c, Z_c)^T$ in the camera coordinate system is projected onto a 2D point $\mathbf{x} = (x, y)^T$ in the image coordinate system as

$$(x, y)^T = \left(f \frac{X_c}{Z_c} + p_x, f \frac{Y_c}{Z_c} + p_y \right)^T$$

where f is the focal length of a lens. By making a camera calibration matrix \mathbf{A} as

$$\mathbf{A} = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix}$$

the projection of a 3D point in the world coordinate system onto a 2D point in the image coordinate system is finally described as

$$\tilde{\mathbf{x}} \sim \mathbf{A}[\mathbf{R}|\mathbf{t}]\tilde{\mathbf{X}}_w$$

where $\tilde{\mathbf{x}} = (x, y, 1)$ is a homogenous image coordinate. This equation is also simplified as

$$\tilde{\mathbf{x}} \sim \mathbf{P}\tilde{\mathbf{X}}_w$$

$$\mathbf{P} = \mathbf{A}[\mathbf{R}|\mathbf{t}]$$

where \mathbf{P} is a 3×4 perspective projection matrix that also represents a camera pose. In order to estimate a camera pose by solving the above equations, getting multiple sets of $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{X}}_w$ is mandatory. For example, \mathbf{P} is linearly computed from six sets because there are 12 unknown parameters in \mathbf{P} and two equations are prepared from one set. Using above operations, the camera pose can be calculated.

Applications of Indoor Localization Using Visible Light Communication

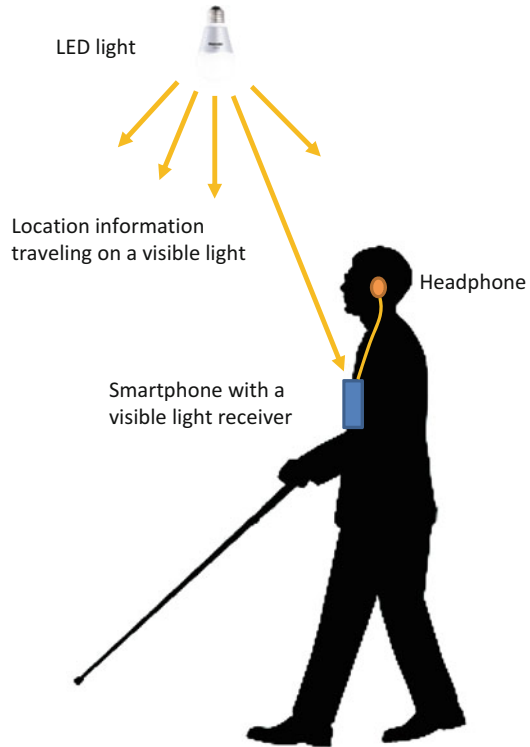
Indoor localization technology using visible light communication has advantages of the positioning accuracy and the ubiquitous communication capability due to the widespread use of LED lights. The applications include indoor navigation, location-based advertising, robot control, and accurate measurement of architectural structure. Details of each application will be described in the following sections.

Indoor Navigation Using Visible Light Communication

A user is able to get indoor navigation service if a terminal side or a server side runs a navigation software. Indoor navigation is thus convenient for everyone, but the people who need such a service most are the visually impaired. Nakajima and Haruyama (2012) is an example of such indoor navigation for the visually impaired as shown in Fig. 8. LED lights emit visible light with location data and a smartphone with a visible light receiver receives the data. The smartphone calculates the optimal path to a designation and speaks to the visually impaired through a headphone.

Its navigation prototype was made as shown in Fig. 9. The visually impaired in the middle of the photo was able to walk with the audio guidance of the indoor navigation system.

Fig. 8 Indoor navigation system for the visually impaired using visible light communication



Location-Based Advertising Using Visible Light Communication

The location-based mobile advertising is gaining popularity recently as more and more people began using smartphones (Müller et al. 2011). In the initial phase of location-based advertising, shops sent coupons to a smartphone of a consumer who came near the shop. This method is called “geofencing.” Recently, as the technology improved, a more accurate targeting of customers is being tried by using more information about customers such as customer's current needs for the retailer's products or services, a user's past purchase history, etc. Furthermore, if a customer's detailed location in a shop can be measured, not just the information that he/she is near the shop, a more accurate advertising will be possible, and location-based advertising by sending data from LED lights will fulfill such a requirement.

One example of location-based advertising using visible light communication is shown in Fig. 10, where digital signage of advertisement information was sent from the LED backlight. In this application, the advertisement information from the LED backlight is received by a user's terminal using a PIN photo diode. This prototype was made by Tamura Corporation, Tokyo.

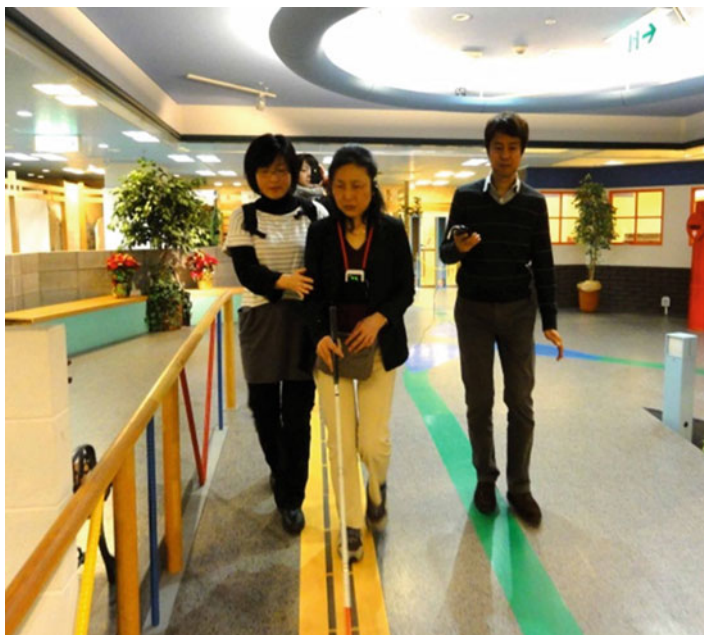
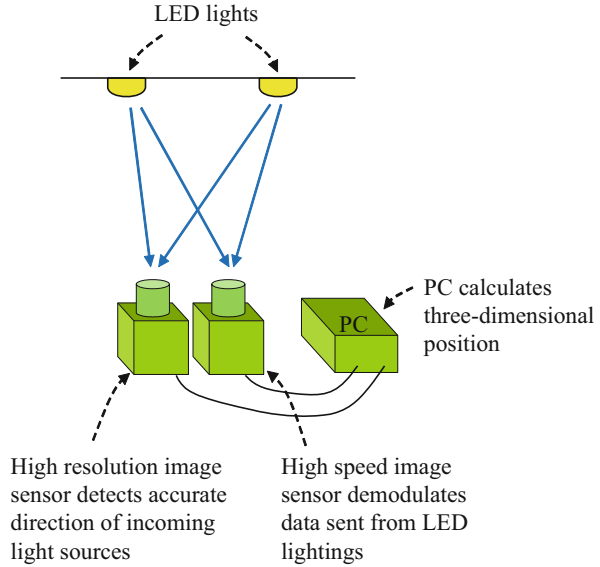


Fig. 9 Indoor navigation prototype for the visually impaired using visible light communication

Fig. 10 Digital signage of advertisement information sent from LED backlight



Fig. 11 Robot control system using visible light communication



Robot Control Using Visible Light Communication

An image sensor is able to calculate a very accurate camera pose by doing simultaneous image acquisition and data reception using visible light communication. If such a camera is embedded in a robot, it will be possible to control its motion very accurately (Tanaka and Haruyama 2009; Arai et al. 2010; Nakazawa et al. 2012).

Figure 11 shows its robot control system using visible light communication. LED lights on the ceiling send location data. A robot has two image sensors: one image sensor obtains high-resolution image to detect an accurate direction of incoming light and the other image sensor obtains high frame rate images in order to demodulate the incoming data on a visible light. In this application, the algorithm of registration with gravity sensor was used, so the minimum of two LED lights at a co-horizontal plane are needed to calculate the three-dimensional position of a robot.

Figure 12 shows a picture of its robot control prototype. For this prototype, the accuracy of the pose calculation was measured and a robot was able to detect its position with an accuracy on the order of centimeter (Arai et al. 2010). Nakazawa et al. (2012) used a fish-eye lens in order to obtain a wide view of the ceiling.

Accurate Measurement of Architectural Structure Using Visible Light Communication

The calculation of the camera pose using visible light communication is able to measure the position of objects very accurately, and one of such applications that

Fig. 12 Robot control prototype with centimeter accuracy



need high accuracy is the measurement of architectural structure using visible light communication.

The direct linear transformation is done to detect the pose of a camera at multiple camera positions as shown in Fig. 13.

In an example shown in Fig. 14, the positions of LEDs attached to a water tank were measured (Uchiyama et al. 2008; Mikami et al. 2011). The accuracy of position using photogrammetric method and visible light communication was about several millimeters at a distance of about 50 m away from an image sensor. This accuracy of position is comparable to that of a typical surveying device called a total station. This system has another advantage of continuous monitoring of positions over time. The positions of LEDs attached to a water tank in Fig. 14 were monitored for 24 h. The roof of the water tank expands in the daytime and shrinks at night due to the heat from the sunshine. The visible light communication photogrammetric system was able to detect the position displacement of several millimeters with an accuracy of a millimeter at the distance of 40 m.

Fig. 13 Measurement of architectural structure using visible light communication

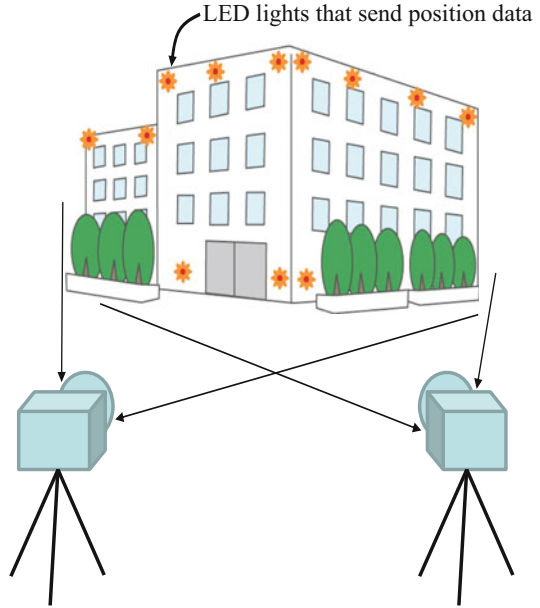


Fig. 14 Survey measurement using image sensors as receivers



Standardization

Several standards for visible light communication have been proposed. The first visible light standard was the CP-1221 of JEITA (Japan Electronics and Information Technology Industries Association) published in 2007. The CP-1221 defines an indicator minimum in order to prevent the interference between different optical communication equipment and also a minimum necessary requirement in various visible light communication applications.

At the IEEE 802.15 Working Group for WPAN (wireless personal area network), the IEEE 802.15.7 Visible Light Communication Task Group has completed a PHY and MAC standard for visible light communications (VLC) in 2011 (IEEE 2011). The IEEE 802.15.7 standard specifies various modulation methods such as OOK (on-off keying) and VPPM (variable pulse position modulation) and a wide range of data rates from 11.67 kb/s to 96 Mb/s and support for dimming.

For indoor localization, a fixed identification signal needs to be transmitted from LED lights and the JEITA CP-1223 standard called “Visible light beacon system” was published in May 2013 to send such an ID from LED lights as a beacon signal (JEITA 2013). The JEITA CP-1223 was a simplified version of the JEITA CP-1222 published in 2007, and it has such features as the data rate of 4.8 kb/s and a modulation method of 4-PPM and a 128 bit data frame content of either ID (fixed data) or arbitrary data (nonfixed). The key portion of the CP-1223 was proposed as “Visible light beacon system for multimedia application” at TC-100 of IEC (International Electrotechnical Commission) in 2013.

Future Directions

Among various indoor localization technologies, the indoor localization method using visible light communication is able to measure the position of a user or an object very accurately. Especially when an image sensor is used to receive position data, its accuracy is on the order of millimeters. Using this technology, various new applications will be possible such as indoor navigation, location-based advertising, robot control, and accurate measurement of architectural structure. There are, however, several problems to be solved in order to make this technology widely available for general users. Some problems are associated with a receiver and some are with an LED light transmitter and infrastructure. The major problem of a receiver that hinders its widespread use is that there are currently no appropriate devices for visible light communication embedded in smartphones. An illuminance sensor that measure the surrounding brightness may be used as an optical receiver, but the software to use it as a receiver is not currently available. A digital camera embedded in a smartphone may also be used as an image sensor receiver, but its speed is not fast enough for many applications. There are also problems associated with an LED light transmitter and infrastructure. LED lights are being widely used now, but adding communication function to LED lights will require additional cost. There will have to be a business model in which some stakeholder is willing to pay for the additional cost. Besides, there has to be a new infrastructure of indoor map database that is required for such applications as indoor navigation, but as of now, there is no universal standard of indoor map data. Nevertheless, if useful services of indoor localization using visible light communication are introduced, many users will ask for them, and many of the above-mentioned technological problems will be solved in the future.

References

- Arai M, Tanaka T, Suzuki S, Haruyama S, Nakagawa M (2010) Visible light positioning system with high speed/low speed image sensor. Technical report of IEICE (Institute of Electronics, Information and Communication Engineers), PN2009-102, pp 91–96
- Barry JR (1994) Wireless infrared communications. Kluwer, Boston
- Do T-H, Hwang J, Yoo M (2013) TDoA based indoor visible light positioning systems. In: 2013 fifth international conference on ubiquitous and future networks (ICUFN), Vietnam, pp 456–458
- Gu Y, Lo A, Niemegeers I (2009) A survey of indoor positioning systems for wireless personal networks. *IEEE Commun Surv Tutor* 11(1):13–32, First quarter
- Haralick RM, Chung-Nan L, Karsten O, Michael N (1994) Review and analysis of solutions of the three point perspective pose estimation problem. *Int J Comput Vis* 13(3):331–356, Kluwer
- Hartley R, Zisserman A (2004) Multiple view geometry in computer vision, 2nd edn. Cambridge University Press, Cambridge
- IEEE Standards Association (2011) 802.15.7-2011 – IEEE standard for local and metropolitan area networks – part 15.7: short-range wireless optical communication using visible light. pp 1–309
- JEITA (2013) CP-1223 standard <http://www.jeita.or.jp/japanese/standard/book/CP-1223/>
- Jung S-Y, Hann S, Park C-S (2011) TDOA-based optical wireless indoor localization using LED ceiling lamps. *IEEE Trans Consum Electron* 57(4):1592–1597
- Kotake D, Sato K, Uchiyama S, Yamamoto H (2005) A hybrid and linear registration method utilizing inclination constraint. In: Proceedings of the 4th IEEE/ACM international symposium on mixed and augmented reality (ISMAR), Vienna, pp 140–149
- Kottke C, Hilt J, Habel K, Vucic J, Langer K-D (2012) 1.25 Gbit/s visible light WDM link based on DMT modulation of a single RGB LED luminary. In: Proceedings of European conference on optical communications (ECOC 2012), Amsterdam
- Mikami H et al (2011) Reports of technical research and development of Sumitomo Mitsui Construction Co., Ltd., no 9, pp 79–84
- Müller J, Alt F, Michelis D (2011) Pervasive advertising. Springer, London
- Nah JHY, Parthiban R, Jaward MH (2013) Visible light communications localization using TDOA-based coherent heterodyne detection. In: 2013 I.E. 4th international conference on photonics (ICP), pp 247–249
- Nakajima M, Haruyama S (2012) Indoor navigation system for visually impaired people using visible light communication and compensated geomagnetic sensing. In: 2012 1st IEEE international conference on communications in China (ICCC 2012). pp 524–529
- Nakazawa Y, Makino H, Nishimori K, Wakatsuki D, Komagata H (2012) Indoor positioning using visible light communication and a high-speed camera equipped with fish-eye lens. In: International conference on indoor positioning and indoor navigation (IPIN), Sydney, Australia
- Richard Szeliski R (2011) Computer vision: algorithm and applications. Springer-Verlag London Limited
- Ryan D, Staats P, Sumner R (2013) Method and system for digital pulse recognition demodulation. US Patent 8,416,290 B2, Apr 2013
- Takai I, Ito S, Yasutomi K, Kagawa K, Andoh M, Kawahito S (2013) LED and CMOS image sensor based optical wireless communication system for automotive applications. *IEEE Photonics J* 5(5), article # 6801418
- Tanaka T, Haruyama S (2009) New position detection method using image sensor and visible light LEDs. In: IEEE second international conference on machine vision (ICMV), Dubai. pp 150–153
- Uchiyama H, Yoshino M, Saito H, Nakagawa M, Haruyama S, Kakehashi T, Nagamoto N (2008) Photogrammetric system using visible light communication. In: 34th annual conference of IEEE industrial electronics society (IECON), pp 1771–1776, Orlando, Florida, USA
- Zhou Z, Kavehrad M, Peng D (2012) Indoor positioning algorithm using light-emitting diode visible light communications. *Opt Eng* 51(8):085009, pp 1–6

Integration of RF and VLC Systems

Michael B. Rahaim and Thomas D. C. Little

Contents

Introduction	684
Trends in Wireless Communication	685
WLANs and Cellular Networks	686
Asymmetric Traffic Distribution	686
Small Cells	686
Directional Wireless	687
Heterogeneous Wireless Networks	688
Visible Light Communication (VLC)	689
Channel Model	689
System-Level Constraints	690
System Integration	691
System Layout	691
Backhaul Network	693
Physical Channel	693
Network Topology	694
Handshaking	695
Traffic Distribution	695
Handover	696
Conclusions	698
Future Directions	698
References	699

Abstract

As the lighting industry moves toward long-lasting solid-state luminaires, advanced systems will begin to integrate novel use cases into the lighting infrastructure. The proliferation of wireless devices and the demand for wireless access in indoor environments create a synergy between the wireless

M.B. Rahaim (✉) • T.D.C. Little
Boston University, Boston, MA, USA
e-mail: mrahaim@bu.edu

communications and indoor lighting industries. Since wireless traffic demand is at its highest in areas where artificial lighting is already in place, it makes perfect sense to incorporate novel wireless access technologies into the lighting infrastructure. This chapter focuses on the integration of visible light communication (VLC) with radio-frequency (RF) networks in order to provide additional wireless capacity in areas where RF is challenged with meeting the growing demand. We review current trends in wireless network access, provide an overview of VLC, and detail the requirements for implementation of such an integrated system.

Introduction

The introduction of solid-state lighting and the trend toward replacement of incandescent and fluorescent light sources with LED-based luminaires have been driven primarily by the promise of improved efficiency. The high-speed switching capability of solid-state lighting is another marketable trait that allows visible light communication (VLC), or the use of the visible spectrum for wireless data transfer, to be implemented with illumination-grade LEDs. This creates a potential for dual-use luminaires that provide lighting and wireless data transmission (Komine and Nakagawa 2004); (O'Brien 2011).

The mobile communication industry is in the midst of a drastic boom in data traffic and is challenged with meeting a potential $1000\times$ growth (Qualcomm 2013). As the number of wireless devices and applications increases, the associated growth in data traffic will push the limits of current wireless network capacity. This is further impacted by a diversification of the types of wireless devices and an increase in typical application complexity. The industry accounts for approximately 1 % of the total US GDP (\$146.2 billion in 2011), and estimations show that every 10 MHz of additional licensed spectrum leads to a \$263 million increase in wireless application and content sales (Entner 2012); therefore, limiting the wireless capacity of end users can impede the development of novel wireless applications and stall technological advancement.

The biggest gains in aggregate wireless capacity stem from (a) increasing available spectrum and (b) increasing bandwidth density (b/s/m^2) by increasing the number of cells and decreasing the per-cell coverage area (Chandrasekhar et al. 2008). Primary motivations for VLC are that it (a) uses the vast, unused, and unregulated visible spectrum and (b) is directional, leading to small coverage area and allowing for an increased density of Access Point (AP), 2 or Base Station (BS), 2. In addition, most data traffic occurs indoors, and the lighting infrastructure is designed such that luminaires are placed according to the expected distribution of users and User Device (UD), 2. Placing APs at the luminaires adds wireless capacity where it is needed most and allows traffic to flow through the available backhaul network of the lighting infrastructure.

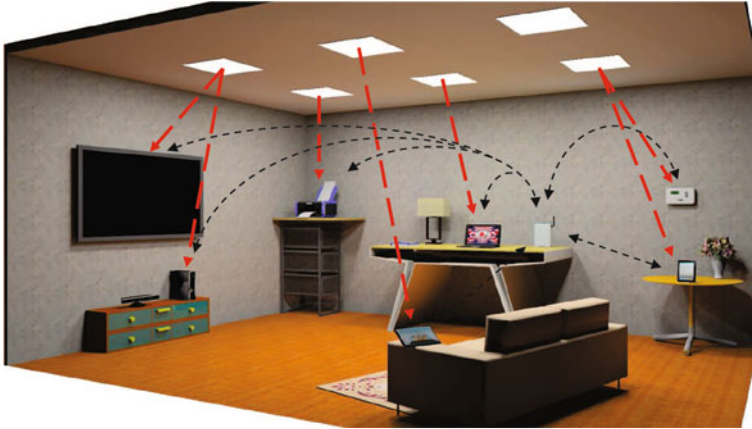


Fig. 1 Heterogeneous network integrating RF and VLC. High data rate devices are offloaded to a VLC channel in order to minimize congestion on the RF channel (*Image courtesy of Yuting Zhang*)

While VLC is a viable technology for downlink connectivity in many cases, constraints of the optical channel limit its potential as a stand-alone wireless medium. The optical channel is susceptible to blocking conditions, and a VLC uplink is not ideal since a light coming from a UD can be intrusive. In addition, RF communications currently dominate the market, and most UDs already incorporate antennas for transmitting and receiving RF signals. This motivates heterogeneous wireless networks that integrate RF and VLC as shown in Fig. 1. In an integrated system, VLC downlinks can alleviate congestion from the RF channel caused by heavily asymmetric traffic (e.g., audio/video streaming), while the RF channel provides a nonintrusive uplink and an alternative link in the event of a blocked VLC signal (Rahaim et al. 2011). Such an integration can also increase VLC market acceptance since the VLC channel is primarily supplemental, allowing UDs that are not VLC enabled to access the network as normal.

In this chapter, we provide an overview of the current trends in wireless communications, define the components of a VLC link, discuss system-level considerations for implementing a practical VLC network, and discuss the necessary features of an integrated system combining RF and VLC. We then conclude with future directions of research related to heterogeneous RF/VLC networks.

Trends in Wireless Communication

In this section, we provide a broad overview of the current trends in wireless communications. More specifically, those associated with cellular networks and wireless local area networks (WLANs). This includes trends relating to use cases, layout, and physical media, as well as the integration of various wireless networks.

WLANs and Cellular Networks

The most widespread wireless networks today are WLANs and cellular networks. The majority of WLANs are based on the IEEE 802.11 standard or Wi-Fi. Both Wi-Fi WLANs and cellular networks implement RF communications; however, there are differences that primarily stem from their original use. WLANs operate at a range on the order of 100 ft and are intended to provide high data rate connections to UDs in the home or work place. The range of a macrocell cellular BS is measured in miles and was originally intended for voice traffic to mobile UDs indoors or outdoors. Since voice traffic is now digitized and many mobile devices now accommodate data traffic, cellular networks have started deploying smaller cells (e.g., microcells, femtocells, etc.), and many UDs are capable of accessing various wireless networks.

Regarding the RF spectrum, Wi-Fi WLANs use various ISM bands with unlicensed spectrum, whereas cellular networks use licensed spectrum owned by the service provider. This implies that Wi-Fi contends with other devices and protocols using the spectrum, whereas cellular networks can allocate dedicated resources to specific UDs. Wi-Fi WLANs implement contention-based multiple access, specifically carrier sense multiple access (CSMA), whereas cellular networks implement resource allocation methods, such as frequency-division or code-division multiple access (FDMA, CDMA).

Asymmetric Traffic Distribution

IP data traffic has evolved from early dominance of asymmetric web surfing applications, to relatively symmetric peer-to-peer networking, to the current dominance of highly asymmetric applications such as IP TV, online gaming, and streaming music. The asymmetry in these applications leads to a need for added capacity in the wireless downlink. In cellular networks, supplemental downlink has been enabled in recent releases of the LTE Advanced standard. This combines unpaired spectrum with the primary paired spectrum in order to provide additional capacity from BS to UD. The idea of adding downlink capacity via an alternative medium can also be applied to WLANs; however, the implementation will vary for networks where contention-based multiple access methods are used rather than resource allocation methods.

Small Cells

Over the past 50⁺ years, the majority of wireless capacity gains have come by means of increased cell density and reduced transmit distance, in turn, providing similar aggregate coverage while reducing per-cell coverage area. As cells become smaller and cell density increases, wireless network capacity improves due to spatial reuse and increased area spectral efficiency (b/s/Hz/m²) (Nakamura et al. 2013).

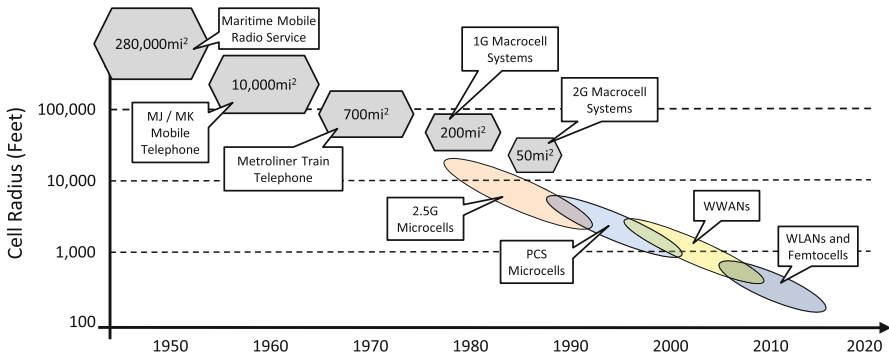


Fig. 2 Historical trend of cell coverage (Image modified from (ZTE Corporation 2012))

Figure 2 depicts the small cell trend leading to WLANs and femtocells (small, low-power, in-home cellular BSs that connect to a service provider’s network through an IP network). Although Wi-Fi WLANs are not part of the cellular network, they are often considered small cells.

Macrocells and femtocells are both *controlled* by a global entity (i.e., service provider); however, femtocell BSs are *owned* by a local entity (e.g., home owner), similar to WLAN APs. Femtocell BSs and WLAN APs are therefore placed in an ad hoc manner as opposed to the planned layout of macrocell BSs. This ad hoc layout can affect the quality of service (QoS) of UDs in the network because of interference between neighboring WLANs or femtocells operating in the same band (Fig. 3a, b). Femtocell BSs and UDs also interfere with overshadowing macrocells and UDs assigned to the macrocell (Fig. 3c, d). The converse is also true (Fig. 3e, f). This interference is compounded by the ad hoc placement of small cell APs or BSs as opposed to the structured layout of larger cells in cellular networks. As the number of small cells continues to grow, there is interest in a paradigm shift from the typical macrocell coverage model to an “Inside-Out” model where small cells cover the indoor space and also provide access to UDs in the vicinity outdoors. The problem that arises in current small cell implementations is that typical femtocell owners prefer to utilize closed access in order to provide the best coverage to their UDs rather than open access which has been shown to provide better aggregate network performance (Andrews et al. 2012).

Directional Wireless

Continuing the small cell trend beyond a single AP or BS per home raises questions regarding the appropriate infrastructure. The omnidirectional emission pattern of most WLAN APs and femtocell BSs limits the minimum effective coverage area since reducing cell size is associated with reducing the maximum distance between the AP or BS and a UD. Directional wireless media, such as VLC, infrared (IR), and mm-wave communications, can emit a narrow signal beam – providing a small

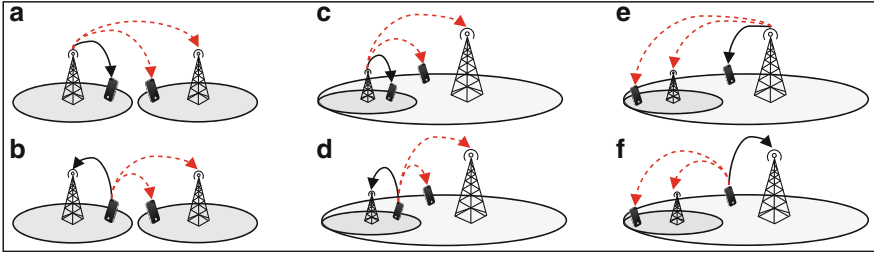


Fig. 3 Interference in small cell networks. *Solid lines* indicate signal and *dashed lines* indicate interference. Interference occurs between small cells (**a, b**) and between tiers using similar resources (**c–f**)

coverage area at the working surface without the same distance requirement. In this case, emission pattern and transmit power jointly affect the coverage area.

The IEEE 802.11 ad standard (IEEE 2012) offers multi-gigabit throughput in the 60 GHz mm-wave ISM band. The standard defines four channels, each with 2.16 GHz channel bandwidth as opposed to the maximum 160 MHz channel bandwidth defined in the IEEE 802.11 ac standard which operates in the 5 GHz ISM band. Higher bandwidth promises higher potential throughput, and the directionality of the medium allows the signal to be dynamically steered or directed toward the UD; however, higher-frequency signals suffer from very high attenuation or complete blocking when physical barriers obstruct the line-of-sight (LOS) path.

Heterogeneous Wireless Networks

Various wireless networks have been designed to operate optimally under specific conditions; however, many of today's UDs are not designed for a specific purpose. This is leading the wireless communication industry toward mobile convergence where wireless networks of different sizes and access technologies work together in a heterogeneous network (HetNet) (Andrews 2013). Consider that a smart phone can be used for voice traffic while riding in your car, web surfing while walking around your home, and video streaming while sitting on your couch. Each of these instances has an increasing amount of traffic and a decreasing requirement for mobility; therefore it would make sense for your phone to transfer between a macrocell in the cellular network, a femtocell or WLAN, and a highly localized directional channel such as VLC or 60 GHz.

The objective of HetNets and mobile converged networks is to incorporate a framework that intelligently distributes UDs among the various cells or networks in order to opportunistically exploit characteristics of the channel that best fits the current mode of operation. This distribution can be across tiers in a multi-tier HetNet where spectrally similar APs or BSs are distinguished by density and transmit power or across heterogeneous access technologies (e.g., Wi-Fi offloading from cellular networks). Figure 4 and the smartphone example both observe a distribution of

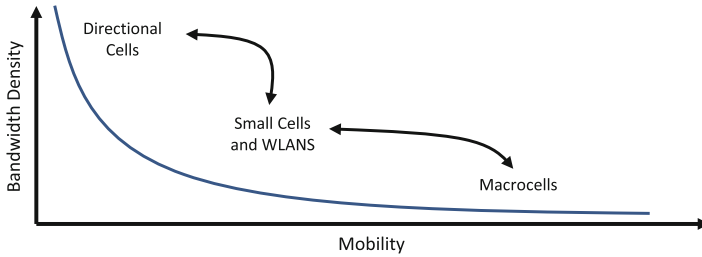


Fig. 4 UDs should be opportunisticly distributed across various networks

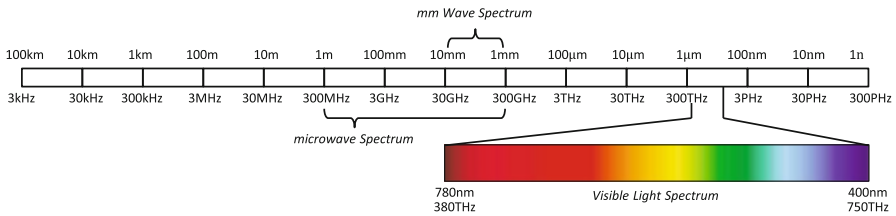


Fig. 5 Electromagnetic spectrum

UDs based on mobility and rate requirements; however, other traits can be incorporated in the decision process.

Visible Light Communication (VLC)

VLC is an optical wireless communication (OWC) technique that utilizes the visible spectrum for data transmission (Elgala et al. 2011). VLC implements intensity modulation with direct detection (IM/DD), implying that the optical intensity is modulated by the desired signal which is recovered via direct conversion of the optical signal to an electrical current. Although the visible light spectrum, shown in Fig. 5, offers a vast range from 400 to 780THz, the use of IM/DD limits the single channel bandwidth to the switching speed of the LED. The immense spectrum can be exploited though wavelength division multiplexing (WDM) techniques that allow many parallel channels to be transmitted using various frequency bands in the optical range. Multi-gigabit per second VLC has been demonstrated with WDM using three colors (red, green, and blue), and there is potential for much higher parallelism given narrowband emitters and receivers (Kottke et al. 2012).

Channel Model

In an IM/DD OW channel, $X(t)$ is defined as the instantaneous optical *signal* power, or intensity [W], of the light source; therefore $\min(X(t)) = 0$ and the constraint

$X(t) \geq 0$ holds for all t . The transmitter generates an average signal power, P_t , and an optical receiver produces an instantaneous received signal current, $y(t)$, such that

$$P_t = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T X(t) dt \quad (1)$$

$$y(t) = R(X(t) * h(t)) + n(t) \quad (2)$$

where $h(t)$ is the channel response, R is the responsivity of the photosensor [A/W], and $n(t)$ is the electrical noise (Kahn and Barry 1997).

Given the presence of a LOS path, the approximation $h(t) = V H \delta(t - \tau)$ is typically very good. The visibility of the link, V , is 1 when the LOS path is unobstructed and 0 otherwise. The propagation delay, τ , represents the time difference between the transmitted and received signal, and the LOS DC channel gain, H , is dependent on both the angle of emission, ϕ , and the angle of arrival, ν . For a VLC receiver with concentrator optics, H is defined as

$$H = \begin{cases} \frac{A}{d^2} R_o(\phi) T_s(\nu) g(\nu) \cos(\nu) & 0 \leq \nu \leq \Psi_c \\ 0, & \nu \geq \Psi_c \end{cases} \quad (3)$$

where A is the area of the photodiode, d is the distance between transmitter and receiver, Ψ_c is the concentrator FOV, and $R_o(\phi)$, $T_s(\nu)$, and $g(\nu)$ are the angle-dependent radiation pattern, filter transmission, and concentrator gain, respectively. The average received signal current, \bar{y} , is proportional to the average transmitted optical signal power such that $\bar{y} = RHP_t$.

Note that OWC requirements, such as illumination in dual-use VLC luminaires, place a constraint on the average transmitted optical power as opposed to the constraint on average electrical transmit power in RF communications. In order to compare performance of various modulation techniques under similar optical power constraints, the OWC signal to noise ratio is defined relative to P_t ,

$$SNR = \frac{(RHP_t)^2}{\sigma^2} \quad (4)$$

where σ^2 is the total noise variance in the electrical domain. Error rate equations for modulation techniques also differ from convention when analyzed with this value of SNR. Since illumination requirements can vary in scenarios where dimming or color-tunable luminaires offer dynamic lighting environments, the SNR and achievable data rate will also vary with the lighting (Gancarz et al. 2013).

System-Level Constraints

Given the potential for VLC as a high data rate point-to-point wireless access technology, the next step in the development of a practical VLC implementation is to determine the potential use cases within an environment incorporating mobile

UDs and dynamic signal conditions. In order to meet the demands of most UD, the wireless link should be bidirectional and promise a high probability of connectivity.

In a dual-use scenario, the optical power emitted in the downlink direction achieves both lighting and communication requirements; however visible light generated from a UD does not assist in meeting room lighting requirements and is likely to be considered intrusive to the user. Because of this, VLC links between a dual-use luminaire and a UD are typically proposed as an asymmetric link implementing an infrared (IR) or RF uplink (O'Brien et al. 2008). In order to maximize coverage in a wireless network, AP layout should be configured with appropriate overlap so that a UD moving through the environment maintains connectivity between cells and doesn't lose network connectivity as it moves out of range of an AP. In addition to out-of-range signal loss, VLC suffers from signal loss due to blocking conditions where the LOS path is obstructed. In the next section, we discuss how heterogeneous network integration can mitigate the effect of these constraints.

System Integration

Integration of VLC in the higher layers of a wireless communication network provides a directional medium that can be opportunistically utilized for high data rate traffic to UD operating in a low-mobility condition. This adds localized downlink channels, allowing intelligent networks to distribute UD across the curve in Fig. 4. It also allows the network layout to be planned without a requirement of overlapping VLC cells since the regions between cells are covered by the overshadowing RF channel. In addition, VLC-enabled UD are capable of (a) using the nonintrusive RF uplink for handshaking and bidirectional traffic and (b) switching to a symmetric RF link in cases where the VLC signal is lost. While RF wireless networks would benefit from the additional capacity of supplemental VLC downlinks, the following considerations must be accounted for when developing and optimizing such integrated systems.

System Layout

The structure of wireless networks is in the process of moving from planned macrocells, maintained by a global entity, to a high-density ad hoc placement of small cells, each maintained by a local entity. Since these small cells don't have any centralized organization, they typically provide some level of self-organizing capability. This is required because the emission pattern of a WLAN or femtocell is typically wide enough to generate interference in an area operated by an unassociated entity (e.g., neighboring apartments each with a Wi-Fi WLAN set to the same channel). On the other hand, VLC cells have a relatively contained emission pattern and can be planned locally because intercell interference will be negligible outside of the area responsible to the local entity. The lighting

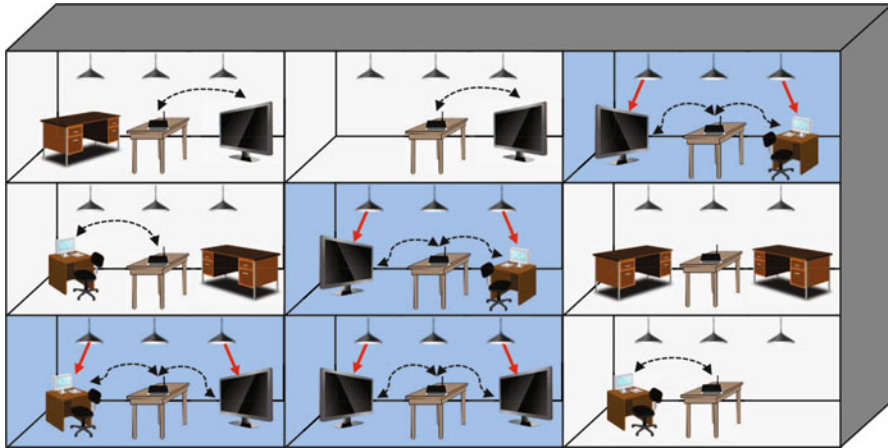


Fig. 6 Apartment complex with various entities. All apartments have interfering RF small cells, and the blue apartments utilize VLC APs to offload wireless traffic from the RF channel

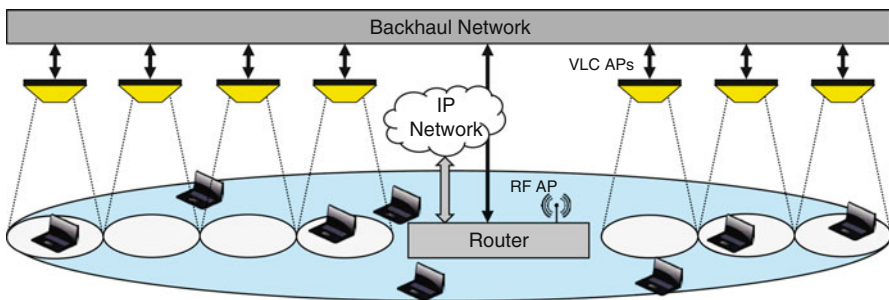


Fig. 7 Basic service set for an RF small cell with VLC integration

infrastructure provides a locally planned distribution of luminaires that maps well to the distribution of UD in many environments.

Figure 6 shows a hypothetical environment where VLC cells supplement RF small cells within an extended service set encompassing many other small cells. When traffic is offloaded to a VLC AP, it removes congestion from the associated small cell and interference from neighboring cells. The basic service set for the integrated system, shown in Fig. 7, consists of multiple UD, one or more VLC APs, and a single central AP consisting of an RF AP, a router, and a gateway to external networks. All UD are equipped with RF transceivers in order to maintain backward compatibility on the RF channel, and VLC-enabled UD are also equipped with a VLC receiver. Connections to the public network pass through the gateway at the central AP, and the backhaul network connects VLC APs and the central AP.

For a given UD associated with a VLC AP within the coverage area of the RF small cell, the physical links available can be implemented in the three

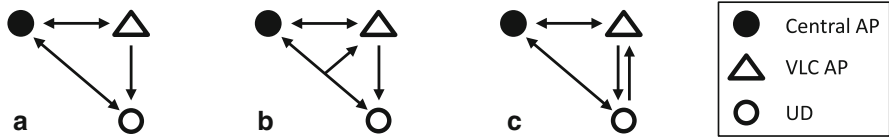


Fig. 8 Potential connections for a UD in an integrated system

configurations shown in Fig. 8. Figure 8a depicts the basic system with a VLC downlink and bidirectional link to the RF AP. Figure 8b assumes the VLC AP has an RF transceiver such that the UD has an additional bidirectional link to the VLC AP; however this channel is in contention with the central AP and other UDs or VLC APs using the same shared RF channel. Figure 8c assumes the UD is enabled with an additional transmitter and the VLC AP has a corresponding receiver such that a non-interfering uplink channel (e.g., IR) is available. In case (a), traffic between the UD and external networks can either flow from UD to central AP, central AP to UD, or central AP to VLC AP to UD. In the scenarios where an uplink from UD to VLC AP is available, there is the additional path from UD to VLC AP to central AP. Within the local network, traffic may also flow from UD to central AP to VLC AP.

Backhaul Network

In order to observe the desired spatial reuse of the optical channel, VLC APs should be spatially distributed throughout the environment. Each of these VLC APs requires network connectivity in order to relay network data to associated UDs; therefore data packets must be able to pass between the central AP and each of the VLC APs. The backhaul network allows data traffic and additional overhead to flow. There are various options for implementing the backhaul network in regard to both the physical channel and the network topology connecting the central AP and the set of VLC APs.

Physical Channel

In networking, a layer model is commonly used to separate the various components of a communication system. The physical channel is typically the fundamental layer that consists of the basic hardware transmission technologies. As the lighting industry moves toward intelligent systems capable of dynamically adapting to user preferences, a key component in the design is the physical channel providing connectivity between devices. Currently, controllable luminaires are often connected with copper wire (e.g., DALI, DMX) or RF mesh networks (e.g., Zigbee) (IES TM-23-11 2011). These techniques provide low data rate throughput – on the order of 100 kbps – which is appropriate for control; however, they are not intended for

high data rate traffic. Two technologies that provide promise for high throughput are power line communication (PLC) and Ethernet – specifically power over Ethernet (PoE). PLC and PoE provide both communication and power, minimizing installation overhead. In home PLC is capable of operating on the order of 100 Mbps, and PoE can be utilized with gigabit Ethernet links. The IEEE PoE+ standard provides up to 25.5 W of power, and some vendors provide PoE+ products offering up to 51 W of power.

Network Topology

A network's topology describes how the various components of the network are connected. The networks *physical topology* defines how the components are physically connected, whereas the *logical topology* defines how traffic can flow in the network. Figure 9 depicts some of the potential topologies to be considered when defining the backhaul network. A *star* topology consists of a central hub that connects to each of the other nodes in the network. In the case of the backhaul network, the central AP would connect to each VLC AP with a unique link. The *tree* topology observes a hierarchy of nodes, beginning with a root node. The root node connects to one or more other nodes with unique links. Each of these nodes may connect to additional nodes that are not already part of the network. The *line* topology is a specific type of tree where each node only has a single child node. In the backhaul network, the central AP is the root node and each VLC AP may route traffic to additional nodes. The *mesh* topology can have a link between any pair of nodes, as long as a path exists between any two nodes in the network. Each node acts as an independent router, allowing nodes to connect to each other in various multi-hop paths. The *bus* topology connects all nodes to a single shared channel. This allows all nodes to have a direct link with any other node; however they must contend for use of the channel.

If the channel used for backhaul connectivity is shared between multiple VLC APs, as in the bus topology, it can become a system bottleneck. This is also the case when a link is the only path connecting the central AP to a subset of VLC APs, as in the tree topology, since all traffic to the subset will need to be routed through the link. The system does not need to operate under the requirement that all VLC APs are

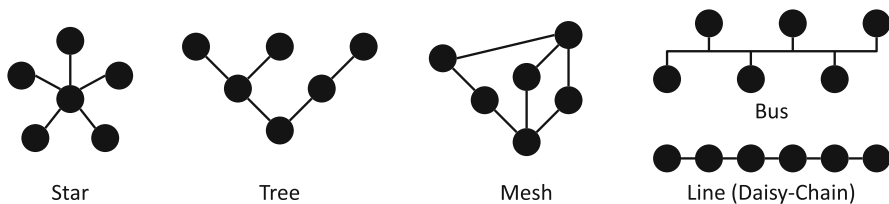


Fig. 9 Some of the various potential network topologies

capable of operating at full capacity simultaneously. For example, assume a system where VLC APs are either unused or at max capacity of X b/s with $P(\text{max}) = 0.5$. If four VLC APs are connected by a bus with capacity $3X$ b/s, the backhaul is a bottleneck when all VLC APs have an associated UD. Since this occurs $100(P(\text{max}))^4 = 6.25\%$ of the time, requirements are satisfied 93.75% of the time. If $P(\text{max}) = 0.2$, the backhaul satisfies requirements 99.84% of the time. Beyond a certain point, additional backhaul capacity provides diminishing gains in the probability of being overloaded.

Handshaking

Handshaking is the process of negotiation between two nodes that allows both to determine that the transmitted data was received. This can occur between the source and destination nodes as well as between two nodes that share a link. When a UD is receiving packets from a VLC AP, an acknowledgment (ACK) should be sent to the VLC AP in order to provide reliability of packet delivery at the physical layer. Given the available paths between a UD and the VLC AP, an ACK can either be sent directly to the AP in case (b) and (c) from Fig. 8 or routed through the central AP in case (a). Since the addition of VLC should allow the RF and VLC media to operate simultaneously, the uplink channel is not reserved, and the ACK may take an indefinite time to reach the VLC AP. This implies that the VLC AP should not wait for the ACK before sending the next packet. One potential handshaking method is a negative acknowledgment (NAK) protocol. In a NAK protocol, the VLC AP would maintain a recent history of previous packets and place a sequential label on each packet sent to the UD. Rather than sending an ACK for every packet, the UD would instead reply with a NAK when it notices a packet error or a missing packet in the sequence. When the VLC AP receives a NAK for a specified packet, it retransmits the packet if it is still in the packet history. Some higher-layer streaming protocols also allow for lost packets, in which case a VLC channel without any physical layer acknowledgments is acceptable since packets with errors can be disregarded.

Traffic Distribution

In an environment with many UDs and multiple VLC APs, the system should be able to intelligently determine how to distribute the UDs across the various APs. In a static environment, the simplest form of distribution is to offload a UD to a localized VLC channel whenever one is available; however having multiple UDs associated with a single VLC AP can potentially saturate the VLC channel. In addition, for certain backhaul topologies, it is possible that a link in the backhaul network is saturated. In either case, associating some of the VLC-enabled UDs with an underutilized RF channel will allow the network to operate with better performance.

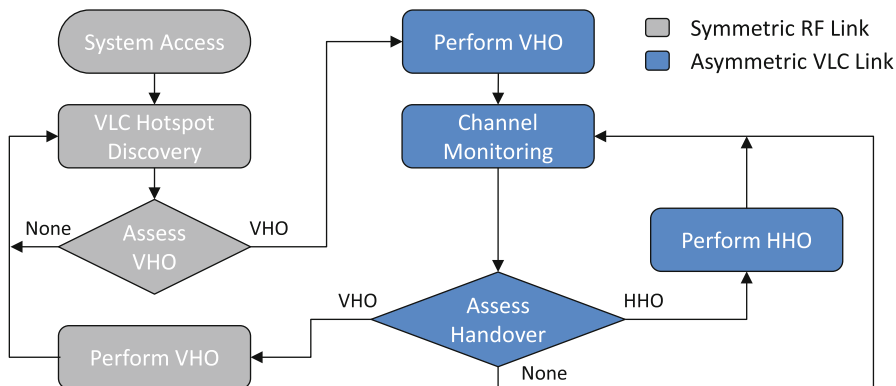


Fig. 10 High-level handover flow diagram

For a specific instant in time within a dynamic environment, highly mobile UD or UD with a high probability of VLC signal loss are better suited for the overshadowing RF channel. The highly localized signal of a VLC channel is best suited for quasi-static UDS – devices that are mobile in nature but typically used in a stationary manner. This includes devices like laptops or tablets that are seldom used in motion. Since switching between channels requires overhead, the network may perform better when a highly mobile UD is associated with the RF channel rather than the VLC AP.

Handover

In a dynamic environment, traffic flow must be rerouted when a signal is lost or an AP becomes overloaded. The process of rerouting traffic to a different AP is called handover (Pollini 1996; Nasser et al. 2006). When switching between two APs of the same type, as in transferring from one VLC AP to another, a *horizontal handover* (HHO) occurs. When switching between APs of different types, as in a transfer from Wi-Fi to VLC, a *vertical handover* (VHO) occurs. Figure 10 shows a high-level flow diagram of a single UD’s actions in an integrated system incorporating a single RF link and multiple asymmetric VLC links. The UD first accesses the network via the RF channel. Downlink traffic is rerouted through the appropriate VLC AP if the UD discovers one to be available and the assessment determines that the handover should be initiated.

In the process of making a handover decision, the UD must first determine whether any VLC APs are available. The UD observes the VLC channel while VLC APs send an intermittent beacon with a unique identifier. If a VLC AP is found and determined to be accessible, handover can be initiated. The assessment of whether the specific UD should perform a handover can either be done (a) in a distributed *user-controlled* manner where each UD makes a decision specifically on

the knowledge of the available channels; (b) in a centralized *network-controlled* manner where the intelligent network maintains knowledge of the UDs within the network and their potential AP associations, then accordingly distributes the UDs such that they are well divided among APs; or (c) in a *mobile-assisted* manner where UDs relay knowledge of their current status (e.g., traffic pattern, mobility pattern, etc.) to the centralized network coordinator such that traffic is distributed among APs in a more optimal way.

A handover request can either be *optional* or *mandatory*. When the RF channel is in use and a VLC AP is located, the VHO is optional since a UD can remain on the RF channel. When a UD is using a VLC link and the signal is lost, VHO is mandatory since the VLC downlink no longer exists. When multiple channels are available, the assessment process should determine if it's better to remain on the current link or initiate a handover. A utility function evaluates parameter values of the various channels, and a handover is initiated when the utility of the new channel meets the requirements for handover. Immediately switching to the highest utility can lead to the ping-pong effect where multiple handovers occur while transitioning between channels; therefore the requirements can include an absolute threshold that the utility of the current channel must drop below or a hysteresis margin, h , such that the utility of a new channel must be greater than the utility of the current channel plus h .

In the case of user-controlled assessment, UD-centric parameters such as channel reliability, signal strength, or required UD power consumption are used. For network-controlled assessment, network-centric parameters such as channel usage are used. With mobile-assisted assessment, a combination of both parameter sets is used. The utility function observes a desired set of parameters for the network, p_1 through p_n , and a set of weights for the network or the specific UD, ω_{p_1} through ω_{p_n} :

$$U = f(\omega_{p_1}, p_1, \omega_{p_2}, p_2, \dots, \omega_{p_n}, p_n) \quad (5)$$

Given the two signal loss conditions for an OWC channel, the type of VHO may differ. An *immediate* handover occurs as soon as the primary signal is lost, whereas a *delayed* handover dwells for a specified time to see if the channel returns before initiating the handover. If an out-of-range signal loss occurs, the signal is usually lost for an extended period – implying that the handover should be made immediately. When a blocking condition occurs, it's likely that something is passing through the LOS path and will return soon – implying that the device should delay before handover initiation in the likelihood that the signal will return (Hou and O'Brien 2006).

As an example, consider a system with an R_V b/s VLC link, R_W b/s Wi-Fi link, X second handover delay, and Y second VLC outage time. After T seconds,

$$D = R_V(T - Y) \quad (6)$$

$$I = R_W(Y - X) + R_V(T - Y - X) \quad (7)$$

where D is the throughput of a UD that waited for the VLC to return and I is the throughput of a UD that immediately switched to the Wi-Fi link when the VLC signal was lost and switched back when it returned. Comparing these two values,

$$I - D = R_W Y - (R_V + R_W) X \quad (8)$$

we find that immediate handover performs better when $Y > \frac{R_V + R_W}{R_W} X$ and delayed handover is optimal when $Y < \frac{R_V + R_W}{R_W} X$. Since the system doesn't know a priori when the VLC link will return, predictive techniques can observe past tendencies, UD motion or rate of signal loss in order to increase the probability that an appropriate decision is made (Rahaim et al. 2012).

Once the network or UD determines that a handover should be initiated, both ends need to coordinate the handover. In a simple case, this implies that the router updates where incoming traffic is routed and the UD changes its expectation of where the downlink traffic is coming from. If a UD is switching to resource allocation channel that is in use by multiple UDs, this coordination includes the definition of the allocated resources for the UD. For example, if a UD is joining a VLC AP using orthogonal frequency-division multiple access, the UD must know which frequency bins to observe.

Conclusions

In summary, integration of VLC luminaires with an RF network provides much needed capacity to keep up with the requirements for the next generation of wireless devices. The initial planning of the system requires an appropriate layout with enough backhaul capacity to satisfy the wireless requirements with high probability. Traffic routing for a given UD should be determined by weighting the additional costs of hardware with the routing complexities for simplified systems. When a system has multiple UDs, distribution of traffic should be such that any individual channel has a low probability of being overloaded. In a dynamic system, appropriate decisions should be made in real time such that the distribution remains satisfactory as network conditions change.

Future Directions

Moving forward, there are many directions for the future of research in heterogeneous VLC and RF networks. The use of multiple-input-multiple-output (MIMO) VLC is an area drawing a great deal of attention due to the parallelism leading to increased link capacity (Butala et al. 2013). From the integrated system view, MIMO VLC fits into all of the concepts described in the chapter since a VLC "cell" can incorporate a set of VLC luminaires. In addition, MIMO allows for dynamic cells since UDs can select the set of luminaires to associate with. This adds some

complexity to the traffic distribution and handover decisions, but provides additional flexibility for the system.

The RF uplink and network connectivity also adds potential for dynamic rate adaptation and lighting control. As UDs move around the environment, the quality of the VLC channel will vary. If the UD can send feedback directly to the VLC AP, the system can incorporate a control loop for illumination and modulation. Dynamic rate adaptation occurs when a UD observes a change in the signal quality and accordingly requests an increase or decrease in the modulation scheme or modulation order.

Another area of potential research is in the scope of large-scale networks and small cell implementation. Given that adding UDs to small cells generates additional traffic and can degrade performance of other UDs, many small cell owners are resistant to the idea of open access. However, given localized VLC channels that are contained within the premises of the local entity and provide the required capacity for UDs belonging to the small cell owner, it is possible that owners would be more willing to open access to the RF channel for secondary UDs.

Finally, integration of VLC luminaires into traditional RF networks creates a wireless link between the lighting infrastructure and devices in the environment. The directionality of the VLC channel provides localization capabilities and additional information that can improve the systems knowledge of light field. Such integration can improve lighting functionality while also allowing the lighting industry to tap into the wireless broadband market and provide additional wireless capacity to meet the growing demand for ubiquitous high-speed wireless network connectivity.

References

- Andrews J (2013) Seven ways that hetnets are a cellular paradigm shift. *IEEE Commun Mag* 51 (3):136–144. doi:[10.1109/MCOM.2013.6476878](https://doi.org/10.1109/MCOM.2013.6476878)
- Andrews J, Claussen H, Dohler M, Rangan S, Reed M (2012) Femtocells: past, present, and future. *IEEE J Sel Areas Commun* 30(3):497–508. doi:[10.1109/JSAC.2012.120401](https://doi.org/10.1109/JSAC.2012.120401)
- Butala P, Elgala H, Little T (2013) SVD-VLC: a novel capacity maximizing VLC mimo system architecture under illumination constraints. In: GLOBECOM workshops (GC Wkshps), 2013 IEEE, <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6825137>
- Chandrasekhar V, Andrews J, Gatherer A (2008) Femtocell networks: a survey. *IEEE Commun Mag* 46(9):59–67. doi:[10.1109/MCOM.2008.4623708](https://doi.org/10.1109/MCOM.2008.4623708)
- Elgala H, Mesleh R, Haas H (2011) Indoor optical wireless communication: potential and state-of-the-art. *IEEE Commun Mag* 49(9):56–62. doi:[10.1109/MCOM.2011.6011734](https://doi.org/10.1109/MCOM.2011.6011734)
- Entner R (2012) The wireless industry: the essential engine of us economic growth. Technical report, Recon analytics. <http://reconanalytics.com/2012/04/essential-engine-of-us-economic-growth/>
- Gancarz J, Elgala H, Little T (2013) Impact of lighting requirements on VLC systems. *IEEE Commun Mag* 51(12):34–41. doi:[10.1109/MCOM.2013.6685755](https://doi.org/10.1109/MCOM.2013.6685755)
- Hou J, O'Brien D (2006) Vertical handover-decision-making algorithm using fuzzy logic for the integrated Radio-and-OW system. *IEEE Trans Wirel Commun* 5(1):176–185. doi:[10.1109/TWC.2006.1576541](https://doi.org/10.1109/TWC.2006.1576541)
- IEEE (2012) IEEE std 802.11ad-2012 (amendment to IEEE std 802.11-2012, as amended by IEEE std 802.11ae-2012 and IEEE std 802.11aa-2012)

- IES TM-23-11 (2011) Lighting control protocols. Technical report. Illuminating Engineering Society
- Kahn J, Barry J (1997) Wireless infrared communications. *Proc IEEE* 85(2):265–298. doi:[10.1109/5.554222](https://doi.org/10.1109/5.554222)
- Komine T, Nakagawa M (2004) Fundamental analysis for visible-light communication system using LED lights. *IEEE Trans Consum Electron* 50(1):100–107. doi:[10.1109/TCE.2004.1277847](https://doi.org/10.1109/TCE.2004.1277847), http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4654267&tag=1
- Kottke C, Hilt J, Habel K, Vučić J, Langer KD (2012) 1.25 gbit/s visible light wdm link based on dmt modulation of a single rgb led luminary. In: European conference and exhibition on optical communication, Optical Society of America, p We.3.B.4, doi:[10.1364/ECEOC.2012.We.3.B.4](https://doi.org/10.1364/ECEOC.2012.We.3.B.4), <http://www.opticsinfobase.org/abstract.cfm?URI=ECEOC-2012-We.3.B.4>
- Nakamura T, Nagata S, Benjebbour A, Kishiyama Y, Hai T, Xiaodong S, Ning Y, Nan L (2013) Trends in small cell enhancements in lte advanced. *IEEE Commun Mag* 51(2):98–105. doi:[10.1109/MCOM.2013.6461192](https://doi.org/10.1109/MCOM.2013.6461192)
- Nasser N, Hasswa A, Hassanein H (2006) Handoffs in fourth generation heterogeneous networks. *IEEE Commun Mag* 44(10):96–103. doi:[10.1109/MCOM.2006.1710420](https://doi.org/10.1109/MCOM.2006.1710420)
- O'Brien D (2011) Visible light communications: challenges and potential. In: Photonics conference (PHO), 2011 IEEE, pp 365–366. doi:[10.1109/PHO.2011.6110579](https://doi.org/10.1109/PHO.2011.6110579)
- O'Brien D, Zeng L, Le-Minh H, Faulkner G, Walewski J, Randel S (2008) Visible light communications: challenges and possibilities. In: Personal, indoor and mobile radio communications, 2008. PIMRC 2008. IEEE 19th international symposium on, pp 1–5. doi:[10.1109/PIMRC.2008.4699964](https://doi.org/10.1109/PIMRC.2008.4699964)
- Pollini G (1996) Trends in handover design. *IEEE Commun Mag* 34(3):82–90
- Qualcomm (2013) The 1000x data challenge. <http://www.qualcomm.com/solutions/wireless-net-works/technologies/1000x-data>, [Online]. Accessed 6 Mar 2014
- Rahaim M, Vegni A, Little TDC (2011) A hybrid radio frequency and broadcast visible light communication system. In: GLOBECOM workshops (GC Wkshps), 2011 IEEE, pp 792–796. doi:[10.1109/GLOCOMW.2011.6162563](https://doi.org/10.1109/GLOCOMW.2011.6162563)
- Rahaim M, Prince G, Little T (2012) State estimation and motion tracking for spatially diverse VLC networks. In: Globecom workshops (GC Wkshps), 2012 IEEE, pp 1249–1253. doi:[10.1109/GLOCOMW.2012.6477760](https://doi.org/10.1109/GLOCOMW.2012.6477760)
- ZTE Corporation (2012) Evolution of microwave radio for modern communication networks. http://www.zte.com.cn/endata/magazine/ztechnologies/2012/no5/articles/201209/t20120912_343888.html, [Online]. Accessed 24 Mar 2014

Part V
Applications

Agricultural and Horticultural Lighting

Paulo Pinho and Liisa Halonen

Contents

Introduction	703
Influence of Light on Plant Growth and Development	704
Quantification and Characterization of Horticultural Lighting	707
Horticultural Light Sources	710
Energy, Environment, and Market Aspects	714
Final Remarks and Future Aspects	717
References	718

Abstract

One of the earliest reports on the usage of artificial lighting in horticulture dates from before the invention of the incandescent lamp. Horticulture in controlled and closed environments is one of the most energy-intensive cultivation systems in agriculture. Artificial lighting is an important part of it. Its use allows for year-round production of quality vegetable, fruit, and ornamental crops independent of weather conditions and geographic location. This chapter reviews the main aspects involving the usage of horticultural lighting considering the challenges faced by the horticultural industry in the context of a food- and energy-hungry world.

Introduction

One of the first written reports on experiments involving the use of artificial lighting to support plant growth and development in horticulture dates from 1861 (Mangon 1861). However, it was just in the first half of the twentieth century that the use of

P. Pinho (✉) • L. Halonen
Department of Electrical Engineering and Automation, Lighting Unit, School of Electrical Engineering, Aalto University, Espoo, Finland
e-mail: paulo.pinho@aalto.fi; liisa.halonen@aalto.fi

electric lighting was introduced for cultivation of horticultural crops in greenhouses (Muijzenberg 1980). Originally, the main objective of the use of artificial lighting in plant cultivation was to supplement daylight during the periods of low availability. Horticulture is typically related to intensive plant cultivation for human use. Horticultural lighting is intended to sustain and promote healthy plant growth and development. Year-round horticultural products are commonly cultivated in protected environments such as in greenhouses, glasshouses, phytotrons, growth rooms, and chambers. Globally, a wide variety of horticultural edible (e.g., tomato, cucumber, pepper, eggplant, lettuces, and herbs) and ornamental (e.g., chrysanthemum, begonia, roses, young bedding plants, and foliage plants) crops are cultivated in these facilities.

As a niche application, the market value of horticultural lighting is therefore relatively small in comparison to other lighting applications more related to human vision and visual performance. Nevertheless, its potential to contribute to address important global challenges such as food security should not be underestimated (Pinho et al. 2012). The fast growth of world population and effects of climate change contribute directly to the continuous increased demand for energy, food, and respective prices globally. It is forecasted that by the year of 2050, the world's population will reach 9.1 billion people (Godfray et al. 2010). The Food Agricultural Organization (FAO) of the United Nations estimates that in order to feed this larger population and achieve food security, the food production must increase by 70 %. This will require that the area of arable land be expanded by approximately 5 % while preserving the scarce fresh water supplies which are unevenly distributed at global scale. In order to address these challenges, it is urgent to find new and environmentally sustainable solutions. Technology plays a key role to facilitate the development of these solutions. Nevertheless, it will be necessary to develop solutions with high energy-saving potential in relation to conventional approaches in order to effectively address these important global challenges.

Horticultural lighting has the technological potential to positively influence sustainable food production, food habits, and human well-being independently of the geographic location, weather conditions, and time of the year. Horticultural lighting offers the opportunity to develop novel cultivation systems and facilities to be deployed inside urban areas where approximately 50 % of the global population is expected to be located by the year of 2050. To promote urban horticulture (i.e., viable intensive plant cultivation in urban spaces) with efficient use of land area, appropriate horticultural lighting technologies and systems will be required.

Influence of Light on Plant Growth and Development

There are several factors influencing plant growth and development, such as water, soil, nutrients, temperature, insects, air humidity, carbon dioxide (CO₂) concentration, and light. However, light is crucial for the healthy growth of the majority of plant species. It is well known that the quantity, quality, and periodicity of light influence the growth and development of plants. The influence of lighting on plant

growth can be seen throughout the analyses of several specific physiological responses such as flowering, stem elongation, fresh weight accumulation rate, organ orientation, stomatal opening, germination, leaf expansion, root growth, sleeping, and phototropic movements. Specialized photoreceptors, usually located in plant leaves, use the captured energy of light to mediate these responses.

The photoreceptors in plants can be grouped into at least three known photosystems, named as photosynthetic, phytochrome, and cryptochrome. The main photoreceptors in the photosynthetic system are chlorophylls and carotenoids. The activity of those photoreceptors is mainly related to light harvesting where light quantity plays an important role. The two main absorption peaks of chlorophylls are located in the red and blue regions from 625 to 675 nm and from 425 to 475 nm, respectively. A new chlorophyll photoreceptor (chl f) has been recently discovered (Chen et al. 2010). The optical absorption spectrum of chl f has a red-shifted absorption peak at 706 nm. This finding suggests that photosynthesis extends further into the infrared region than previously thought. The previously known chlorophylls included chl a, chl b, chl c, and chl d. Carotenoids such as xanthophylls and carotenes also belong to the group of photoreceptors composing the photosynthetic photosystem of plants. Carotenoids absorb mainly in the blue region and are known as auxiliary photoreceptors of chlorophylls.

The phytochrome photosystem includes two interconvertible forms of phytochrome (Pr and Pfr), which have their maximum absorption peaks located in the red and far-red regions of the electromagnetic spectrum around 660 and 730 nm, respectively. Plant responses mediated by phytochromes are usually related to the sensing of the light quality through the red (R) to far-red (FR) ratio (R/FR). There are five types of phytochromes (phyA, phyB, phyC, phyD, and phyE) currently identified in *Arabidopsis* (Fankhauser 2001). Phytochromes are involved in the biological processes controlling several plant responses to light such as leaf expansion, shade avoidance, stem elongation, seed germination, and flowering induction. The red to far-red (R/FR) ratio of light is known to play a key role in shade-avoidance response in plants throughout the mediation of phytochromes. However, there are evidences that blue light and light intensity are also involved in the adaptive morphological responses such as shade avoidance (Christophe et al. 2006).

The cryptochrome photosystem is composed by photoreceptors absorbing light in the blue and UV-A (ultraviolet A) region of the electromagnetic spectrum. Photoreceptors such as cryptochromes (cry1, cry2) and phototropins (phot1, phot2) are known to be involved in monitoring the quality, quantity, direction, and periodicity of the light. These photoreceptors are involved in the mediation of a wide variety of biological processes in plants such as control of gene expression, transition to flowering, leaf expansion, stem elongation, stomatal opening, regulation of pigment content, and the positioning of photosynthetic organs and organelles in order to optimize the light harvest (Spalding and Folta 2005). It is interesting to notice that cryptochromes have also been shown to be part of the circadian clock in mammals and small insects (Cashmore et al. 1999; Christie and Briggs 2001). In spite of the harmful effects of UV radiation on the DNA structure of cells in living organisms, there are recent proofs of evidence of the existence of UV photoreceptors in plants

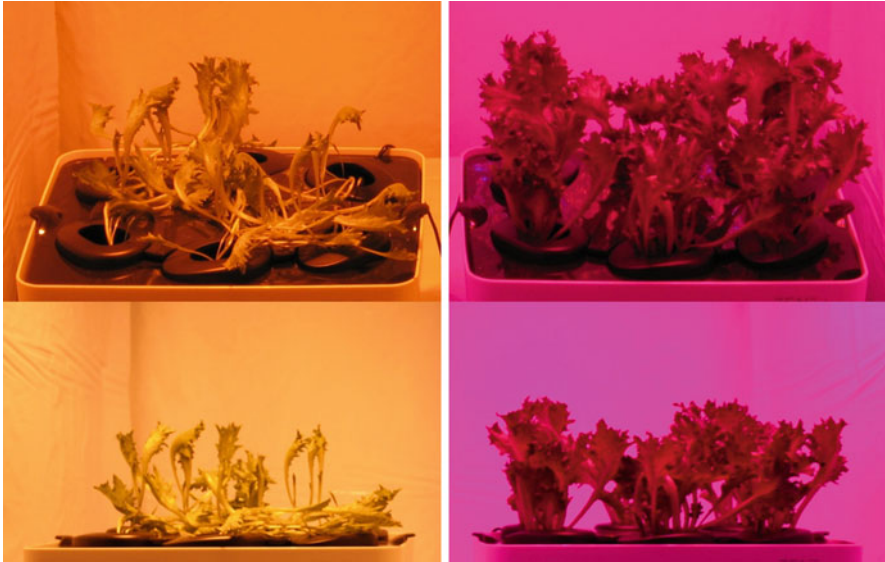


Fig. 1 Side-by-side comparison of light spectrum influence on morphology and fresh weight accumulation of lettuce (*Lactuca sativa*) plants grown under conventional high-pressure sodium lamps (*left*) and a combination of red and blue LEDs (*right*) at similar lighting intensity levels and environmental conditions

(Rizzini et al. 2011). UV-A radiation can influence the development of plants (e.g., stem elongation, dry weight, leaf area, photosynthetic activity, and flowering).

Light quality (i.e., light spectrum) can have a remarkable influence on growth, morphology, and phytochemical composition of plants (Fig. 1). Eventually, one of the first known scientific reports on spectral influence of light on growth and development of plants is dated from 1881 (Darwin and Darwin 1881). In his work, Charles Darwin presents perhaps the first proofs which demonstrate that the heliotropic (i.e., phototropic) movement of plants was influenced by radiation in the blue region of the electromagnetic spectrum comprised between 400 and 500 nm wavelength interval. In spite of this early discovery on the importance of light spectrum on plant growth, the type of interactions and the nature of the interdependence between known and not yet known groups of photoreceptors which mediate growth responses are still not yet well understood. Evidence of this can be found in the existing contradictory knowledge of the importance of green light (i.e., 500–600 nm) on plant development. Under white light illumination (e.g., daylight), the color of plant leaves is commonly perceived by the human eye as green. This perception results from the optical properties (i.e., spectral reflectance, absorbance, and transmittance) characteristic of green leaves. It is also known that in green leaves, the absorption of red and blue light is stronger than green light. However, this may not necessarily indicate that green light is less effective than red light in promoting photosynthesis and contributing to the healthy and balanced growth of plants. In fact, there are

indications of the existence of green light sensory systems which adjust the development and growth of plants in orchestration with red and blue photoreceptors (Folta and Maruhnich 2007; Wang et al. 2011).

Plant growth and development are not solely influenced by the qualitative aspect of light. Plant productivity is strongly dependent on light quantity through photosynthesis from which carbohydrates, such as sugar or glucose, and oxygen are synthesized from carbon dioxide and water using the captured energy of light (i.e., photon's energy). This radiant energy is harnessed by using specialized photoreceptors such as chlorophylls and converted into chemical energy to be used by plants in their metabolic process. Unfortunately, in spite of its 3.7 billion years of existence, the conversion efficiency of oxygenic photosynthesis is still surprisingly low. The maximum conversion efficiency of natural photosynthesis (i.e., solar energy to biomass) in green plants has been estimated to be 4.6–6 % (Zhu et al. 2008). Furthermore, if secondary processing such as growth is also considered, the efficiency may not exceed 1–2 %. In spite of these facts, photosynthesis still is the most important, widely available, and vital photochemical process on Earth. Therefore, the efforts to increase its conversion efficiency can be extremely important in order to respond to the continuous increasing demand for food and energy globally. It is estimated that the photosynthetic conversion efficiency has to be increased by 50 % in order to double agricultural crop production and respond to future global needs (Zhu et al. 2010). Naturally, the conversion efficiency of photosynthesis is also an important aspect to horticultural lighting. High photosynthetic conversion efficiencies will minimize production costs and environmental impacts related to the usage of electricity. Therefore, the main aim of horticultural lighting should be to maximize production efficiency by maximizing crop yields while minimizing the electrical energy consumption.

Quantification and Characterization of Horticultural Lighting

The existing metrics and methods for quantifying radiation used by plants in photosynthesis are very confusing. Radiometric, quantum, and photometric systems of units are frequently and indiscriminately used to quantify and report light for plants. The radiometric system, which is the basis of the photometric system, uses radiance power as the basic quantity and watt (W) as the basic unit. This quantity represents the flow rate of radiant energy in joule (J) per unit time or second (s). However, radiant energy should not be used to quantify light for plants because it does not properly correlate with photosynthetic rates. This is mainly due to the photochemical nature of the photosynthesis process. The radiometric quantities are mostly intended for the characterization and measurement of optical radiant energy within a wide band of wavelengths typically ranging between 10 nm and 1,000 μ m which include the ultraviolet, visible, and infrared regions of the electromagnetic spectrum.

The photometric system of units has been the only system formally defined by the Bureau International des Poids et Mesures (BIPM) for the measurement of

photobiological quantities in the International System of Units (SI). In spite of that, the use of photometric quantities to quantify radiation used by plants during photosynthesis should be in most of the cases avoided or at least carefully considered. This is because the photometric system of units is based on average spectral sensitivity response curves of the human eye. Therefore, the measurement of photometric quantities covers a narrow band of wavelengths defined by the average spectral sensitivity range of the human eye comprised between 380 and 760 nm. Photometry is conventionally used in lighting applications related to support human vision and activities. Therefore, the comparison of the photosynthetic performance of light sources with different spectral qualities using photometric units or quantities is neither recommendable nor accurate.

The Stark-Einstein law directly relates the amount of photosynthetic photons incident on a plant leaf with the amount of chemical change in molecules (Hart 1988). Due to the photochemical nature of photosynthesis, the photosynthetic rate, which represents the amount of oxygen evolution or the amount of carbon dioxide fixation per time unit, can be correlated with the number of photons falling per unit area per second on a leaf surface. Therefore, it is recommended that the measurement system of units intended to quantify light or radiation used by plants during photosynthesis is based on quantum system of units, having the mole (i.e., unit for amount of substance) as its base unit and photon particles (i.e., quanta of radiant energy) as its elementary entity.

The 14th Conférence des Poids et Mesures (CGPM) in 1971 included mole (mol) increasing to 7 the total number of base units composing the SI (Terrien 1972). The possibility and the necessity of revising the definitions of some of the base units of the present SI in order to improve it have been discussed over the past two decades (Waldemar 2010). In the future, apart from the candela, all the base units are intended to be defined in terms of universal physical constants. Mole is one of the seven SI base units which has been considered to be redefined by choosing exact numerical values for the Avogadro's constant. These new definitions are only intended at improving the SI and will not interfere with the continuity of the present measurements. It was also during the early 1970s that several measurements were performed and a comprehensive set of data on the action spectrum, spectral absorbance, and quantum yield of CO₂ uptake for 22 species of crop plants was gathered (McCree 1972a, b). Based on this data, in 1993, the Commission Internationale de L'Eclairage (CIE) provided guidelines stating that photosynthetically active radiation (PAR) for plants should be reported as the total photon exposure comprised between 400 and 700 nm wavelength region of the electromagnetic spectrum and that instantaneous measurements made with a flat or hemispherical sensor should be reported as photosynthetic photon flux density (PPFD) and mol m⁻² s⁻¹ as its unit (Tibbitts 1993). In practice, PPFD measures the number of moles of photons falling per unit area and per unit time falling on a surface.

Similarly to conventional light sources, horticultural light sources are commonly characterized by their optical (e.g., light), electric, and physical properties. However, depending on the lighting application, different systems of units can be utilized, for instance, to measure and characterize the optical performance of systems and sources. Table 1 shows the main equivalent quantities, symbols, and units of the

Table 1 Main radiometric, photometric, and quantic symbols, units, and quantities commonly used for measuring and reporting of light and radiation

Radiometry			Photometry			Photon quantum		
Quantity	Symbol	Unit	Quantity	Symbol	Unit	Quantity	Symbol	Unit
Radiant power	Φ_e	W	Luminous flux	Φ_v	lm	Photon flux	Φ_p	mol s ⁻¹
Radiant intensity	I_e	W sr ⁻¹	Luminous intensity	I_v	cd	Photon intensity	I_p	mol s ⁻¹ sr ⁻¹
Radiance	L_e	W m ⁻² sr ⁻¹	Luminance	L_v	cd m ⁻²	Photon luminance	L_p	mol m ⁻² s ⁻¹ sr ⁻¹
Irradiance	E_e	W m ⁻²	Illuminance	E_v	lx	Photon flux density	E_p	mol m ⁻² s ⁻¹

radiometric, photometric, and quantum systems. These quantities are commonly and in some cases arbitrarily used measurements for reporting of light in plant growth.

As mentioned earlier, energy efficiency plays an important role in many areas of applications. Horticultural lighting is no exception. Therefore, the characterization of horticultural light sources in terms of its energy efficiency performance is nowadays recommendable and desirable. However, in order to correctly assess or quantify the energy efficiency performance of horticultural light sources, a careful assessment should be carried out. Although the use of right units and quantities is of great importance to correctly quantify the energy performance of horticultural light sources, the crop specie and the growth response should be also considered. For instance, the light quality and quantity that might be suitable to increase the productivity of tomato plants in relation to a reference light source may not necessarily produce the same effects on cucumber plants. Eventually, the most reliable way to identify the energy performance of light sources is by carrying out a life cycle assessment (LCA) of the light source by selecting a relevant functional unit which directly relates to the intended crop yields.

It is known that light quantity strongly influences the photosynthetic rates and important growth parameters such as fresh or dry weight accumulation. On the other hand, its effects on other physiological responses, such as flowering, still are not totally understood (Thomas 2006). Although photon flux densities do closely correlate with photosynthetic rates due to the photochemical nature of photosynthesis, it does not take into account any existing photosynthetic spectral sensitivity curve of plants. Already in 1965, McCree was calling attention to the fact that there was not any evidence that plants have a linear spectral response to radiation (McCree 1965). Additionally, the quantification and characterization of light used by plants during photosynthesis is very complex due to several influencing factors involved (e.g., temperature, humidity, plant genomics, quantity, quality, and periodicity of light). This might be the main reason why there is not yet a universally accepted system of units which can fully characterize and quantify the influence of light on the growth and development of plants. The establishment of a method to coherently quantify and characterize light sources used in plant growth applications will allow a more appropriate design and optimization of future horticultural lighting installations. Also, in respect to the economics of this, it is expected that a coherent metrology will better forecast and correlate investments in lighting with the expected and desirable benefits. In this way, a universally accepted system of units can also be beneficial to horticultural industry, lighting industry, and associated activities and enterprises.

Horticultural Light Sources

The sun is the main source of visible and invisible electromagnetic radiation to Earth. This natural light source is also the main factor responsible for the existence of life on Earth. Although approximately one-third of the sun's radiant energy incident on Earth is reflected back to space, the net yearly solar energy reaching the Earth surface

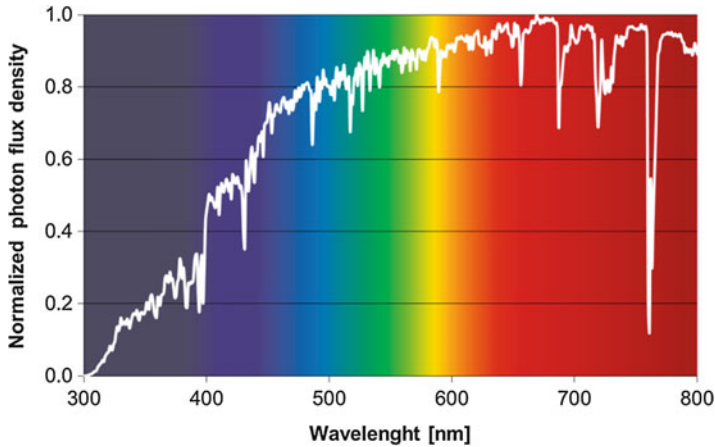


Fig. 2 Normalized spectral photon flux density distribution of sunlight on Earth surface derived from the ASTM G173 global solar spectral irradiance reference measured on a south-facing surface tilted 37° from horizontal and at an angle of incidence of approximately 48° and at air mass 1.5 (AM 1.5G) (ASTM); Gueymard et al. 2002)

is approximately 2,800,000 EJ (i.e., 265 EBtu) (Hart 1988). This value is more than 5,000 times higher than the world's total annual primary energy consumption in 2010, estimated in 523.9 PBtu (i.e., 552.7 EJ) (EIA 2013). The spectral bandwidth of the sun's radiation, as it can be measured at the Earth's surface, is comprised between approximately 250 and 2,500 nm. However, only 50 % of this radiation is comprised within the PAR region (Boyle 2004). The solar spectral photon flux density on the surface of the Earth at an angle of approximately 48° of incidence (AM 1.5G) within the PAR region is relatively constant as shown in Fig. 2.

The advent of electric lighting may have already started during ancient periods of human existence due to the almost magical attraction exerted by the sun. It is interesting to speculate that its imitation has led to the first incipient forms of artificial lighting based on flame (i.e., pyroluminescence) and later on to the first electric light sources such as the incandescent lamp. The incandescent lamp patented by Thomas Edison in 1879 is also the first known electric light source to be used in horticulture (Muijzenberg 1980).

The light emission of incandescent lamps is mainly based on the incandescence mechanism produced by an electric current flowing through the tungsten filament. The heating effect will produce a continuous spectrum similar to a blackbody radiator characterized by a strong emission on the far-red and infrared regions of the electromagnetic spectrum. The electrical conversion efficiency of incandescent lamps, as given by the ratio between radiant energy within the PAR spectral region and input electrical power, is very poor. Typically, just around 10 % of the input power is converted to PAR and the rest radiated as heat. Unbalanced spectrum and poor energy efficiency are the main reasons why incandescent lamps are rarely used anymore as a main artificial light source in horticulture. Due to the high amount of

far-red radiation emitted, incandescent lamps are mainly used in plant growth applications, to control photomorphogenetic responses, such as flowering, through the mediation of the phytochromes. For instance, floral responses can be initiated with long-day responsive plant species through overnight exposure to low photon flux densities using incandescent lamps (Yanagi et al. 2006). The incandescent lamps have the lowest relative lifetime expectancy (e.g., typically around 2,000 h) in comparison to any other electric light source. Lifetime is an important aspect in commercial greenhouse cultivation due to the large area to be irradiated and the high number of lamps typically involved. Longer lifetime expectancies can represent important reductions on the total maintenance and operational costs by decreasing the work labor costs related, for instance, to re-lamping. Since 2005, incandescent lamps have been gradually phased out.

Fluorescent lamps are more frequently used as a horticultural light source than incandescent lamps. The main reasons for this are related to the higher electrical efficiency, longer lifetime expectancy, and more balanced spectrum. Typically, tubular-type fluorescent lamps can achieve electrical efficiencies around 30 % with more than 90 % of the emitted photons within the PAR region. The lifetime expectancies of especially designed fluorescent lamps can reach very long life expectancies. Spectrum is another advantage of using fluorescent lamps in plant growth applications. Depending on the correlated color temperature (CCT) of the lamp, the amount of photon emitted between 400 and 500 nm (i.e., blue light) can reach more than 10 % of the total emission inside the PAR region. It is known that blue light is an important factor in achieving a balanced morphology of most crop plants through the mediation of the cryptochrome family of photoreceptors (Goins et al. 1997). Light is generated from the down conversion of the primary ultraviolet emission by layers of phosphor materials placed in the inner surface of the glass tube during manufacturing. Therefore, the spectrum of fluorescent lamps can be optimized for horticultural lighting usage by appropriate selection of phosphor materials and quantities. Eventually, in order to take advantage of this approach, both electrical efficiency of the lamp and photosynthetic performance of the plant have to be carefully considered. Fluorescent lamps are commonly used as a replacement of daylight in growth rooms and chambers and as replacements of incandescent lamps to promote flowering. Although halogen replacement lamps offer the same light quality as incandescent lamps, fluorescent lamps (e.g., compact fluorescent lamps) and LED lamps have been used due to their longer lifetime expectancies and higher energy efficiencies.

The emission of light by metal halide lamps is based on the luminescent effect. The optimization of light spectrum is achieved to a certain extent by an appropriate use of metal halides during manufacturing. Metal halide lamps are used as horticultural light source to totally replace daylight or for partially supplementing it during periods of lower availability. Typically, the spectrum of metal halide lamps are characterised by a balanced spectral power distribution across the PAR region with approximately 20 % of the total PAR emission in the blue region (Brown et al. 1995; Schuerger et al. 1997). This characteristic makes metal halide lamps a suitable horticultural light source for the sturdy growth of leafy vegetables with compact

morphological features. Additionally, the high electrical efficiency (i.e., approximately 25 %) and long lifetime expectancy (i.e., around 20,000 h) make metal halide lamps a viable light source for plant cultivation and an option for year-round commercial crop cultivation.

The high-pressure sodium (HPS) lamp has been the preferred light source to supplement daylight in greenhouses in northern latitudes. The main reasons are related to its low cost, high PAR emission, long lifetime expectancy, and high electrical efficiency. Approximately, 40 % of the input energy is converted into emission of photons inside the PAR region and almost 25–30 % into far-red and infrared. However, the main drawback of HPS lighting is related to the poor quality of its spectral emission which is predominantly in the yellow-green and infrared regions of the electromagnetic spectrum. Therefore, plants growing under HPS lighting alone may suffer from unbalanced morphology expressed by excessive leaf and stem elongation (Tibbitts et al. 1983; Wheeler et al. 1991). This is due to the low R/FR (i.e., red to far-red) ratio and low blue light emission. Another drawback of HPS lamps is the excessive heat emission which obstacles its use in close proximity with plants.

There are other lighting technologies such as induction and sulfur lamps which have been occasionally experimented as light sources for plant cultivation. In spite of the relatively high energy efficiency and long lifetime expectancies, the usage of these electrodeless lamps in horticulture has been hindered by the high initial costs of the installations. Sulfur lamps have been evaluated for plant growth applications and considered in the past as the prime candidate for the development of hybrid lighting systems for bioregenerative life support in space (Both et al. 1997; Cuello 2002). However, the high cost, noisy operation, and the lifetime of magnetrons, which poses reliability problems, were important obstacles for its use in many applications including horticultural lighting (Johnston 2003). Additionally, sulfur lamps have a high emission in the blue-green region, which can reach more than 60 % of the total photon PAR emission. This may not provide a balanced spectral emission suitable to be used as a main light source for plant cultivation.

In spite of the technological developments that have taken place over the past century, the electrical efficiency of most of artificial light sources did not improve significantly with the exception of inorganic light-emitting diode (LED) lighting technology. The potential of LEDs in lighting applications is very well known. In horticultural lighting, in addition to the high energy efficiency potential, LEDs do offer the possibility of optimization of light spectrum. The combination of different colored LED components in the same luminaire permits the optimization of the light spectrum in order to promote specific growth responses of plants. This is possible due to the narrow spectral bandwidth characteristic of colored LEDs which can closely match with the main absorption peaks of photopigments in plants known to play important roles in the mediation of specific growth responses.

It is expectable that LED luminaires will gradually become a common option in horticultural applications as the main lighting solution to support vegetative growth as long as the initial cost continues to drop and efficiency to rise. Also, with the phaseout of incandescent lamps from commercialization and production, LED lamps

are proved to be a viable alternative to control photomorphogenic responses such as flower inhibition. In some short-day plant species, LEDs with a moderate to high R/FR ratio are effective at preventing premature flowering (Craig and Runkle 2013). For all these reasons, the LED lighting technology has been seen with high interest and expectations by the horticultural industry. However, its wide adoption by the horticultural industry has been relatively slow due to several factors. Perhaps the most relevant factor still is the high initial costs of LED lighting in comparison to conventional lighting. Another important factor is related to the unconventional electrical, optical, and thermal characteristics of LEDs which requires the definition and standardization of several aspects such as lifetime and measurement procedures. Eventually, in horticultural lighting, the scenario might get slightly more challenging due to the lack of standardized measurement procedures and system of units for radiation used by plants in photosynthesis. The LED technology indeed offers new possibilities for plant cultivation in general and for horticulture in special. Lighting can now be tailored and optimized according to plant type, species, or stage of development. Therefore, it is likely that a large variety of LED luminaires with different spectral qualities and from different manufacturers will appear in the market claiming their unique benefits and excellent performances. In order to facilitate safer and more reliable decision making by the grower or consumer, standardized measurement procedures for the characterization of these luminaires and systems should be developed and implemented.

Less conventional light sources for plant cultivation are the organic light-emitting diode (OLED) and light amplification by stimulated emission of radiation (LASER). OLEDs do share most of the potentialities offered by its counterpart, the inorganic LED, to be used as a viable horticultural light source. The unique optical characteristics of OLEDs do offer interesting application possibilities as horticultural light source in the future. The application of LASER diodes as horticultural light sources has been limited to small-scale growth trials (Yamazaki et al. 2000, 2002). Due to the monochromatic emission characteristic of LASERs, luminaires can be spectrally optimized in order to match closely with the main absorption peaks of photosynthetic pigments in plants. Additionally, the high electrical efficiency potential and small size make LASER diodes an interesting option as viable horticultural light source in the future.

Energy, Environment, and Market Aspects

The energy efficiency, environmental impact, and market competitiveness are huge challenges to horticultural industry (Pinho et al. 2012) (Fig. 3). Greenhouses are among the most intensive energy consumption systems in agriculture. For instance, in the recent years, one of the most relevant technological developments in greenhouse cultivation has been the increasing utilization of artificial lighting. In northern latitudes where daylight availability during winter period is insufficient to sustain healthy plant growth and development, artificial lighting is commonly used as supplement. For instance, based on historical measurement data of the average

Fig. 3 Energy efficiency, environmental impact, and market competitiveness are the main challenges currently faced by the horticultural industry

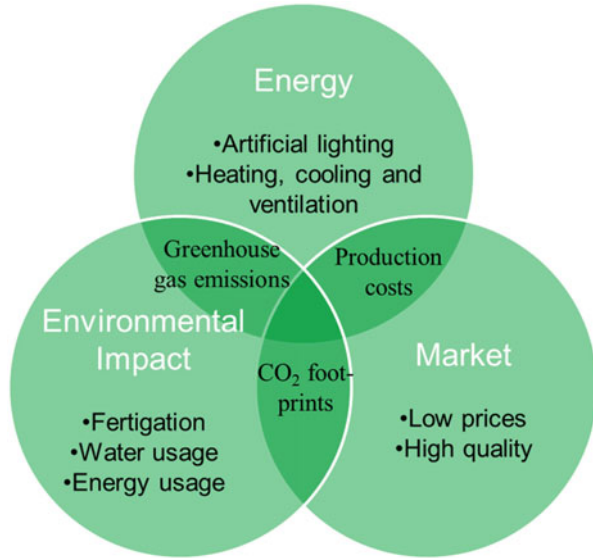
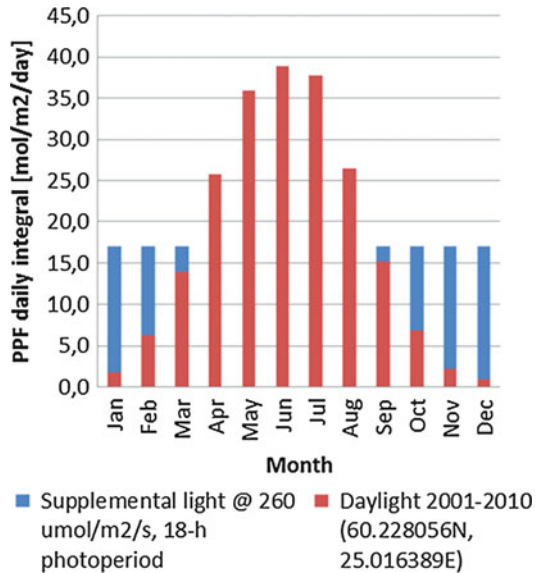


Fig. 4 Based on average daylight availability in Helsinki region between the years of 2001 and 2010 and a minimum daily PPFI requirement of 17 mol m², supplementary artificial lighting would be necessary during winter and autumn seasons from March to September



monthly solar irradiance at the Helsinki region between the years 2001 and 2010, it is possible to estimate that in order to guarantee a minimum of 17 mol/m² of daily photosynthetic photon flux integral (PPFI) in the growth area, supplementary lighting has to be used typically from September to March as shown in Fig. 4. The effort of using supplementary artificial lighting has decisively contributed to significantly increase crop productivity and allow for year-round production. For instance, in

Scandinavia, the annual productivity of cucumber in greenhouses increased from 50 to 250 kg/m² due to the utilization of modern cultivation methods and technologies such as artificial lighting. However, due to inexistence in the past of a less costly and more energy-efficient alternative to HPS lighting, the increasing utilization of artificial lighting was contributed to significant increases in electricity consumption and related costs. In some cases, the costs related to electricity usage can reach 30 % of the total production costs. Unless more energy-efficient horticultural light sources and lighting systems are developed, the costs related to electricity may continue to rise, following the global trend of energy prices. Hence, reductions in energy consumption without decreasing crop productivity and quality are desirable. These are also crucial factors to increasing production efficiency and guaranteeing the economic viability as well as environmental sustainability and market competitiveness of year-round horticultural production in closed and controlled environments.

Improvements on performance of energy usage can also contribute to reduce the negative environmental impacts of horticultural production in closed and controlled environments. Considering that electricity consumption due to lighting accounts for an important part of the total energy consumption in greenhouses, appropriated assessment of environmental impact of horticultural light sources and systems is indispensable. The life cycle assessment (LCA) provides the means to evaluate the environmental impact of products and systems. A pioneer academic study on the LCA of horticultural lighting indicates that the environmental burden of LED lighting can be significantly smaller than HPS lighting in greenhouse cultivation of cucumber seedlings (Grigoriadis 2013). This is mainly due to the higher energy efficiency and photosynthetic performance of LED luminaires used in this study in comparison to HPS luminaires. Even though LED luminaire raw material production and manufacturing processes are more energy intensive than equivalent HPS luminaire processes, the energy usage phase plays a decisive role in the obtained results. Naturally, environmental impacts of horticultural lighting can also be reduced significantly if renewable energy sources are utilized especially during the usage phase of the luminaires.

The rapid knowledge gathering on the effects of light on plant growth and development will allow for more energy-efficient production of plants in protected environments. To fully achieve this objective, appropriate control will be required. The integration of artificial light parameters in a holistic design approach of cultivation systems in controlled and closed environments is one of the possibilities which may allow for efficient use of energy. In greenhouses, the combination of daylight light sensors and smart control schemes may be used to provide more appropriate control of supplementary lighting (Chang et al. 2013). Control of the light intensity is more relevant when supplemental lighting is used. However, in these situations, the control strategies of the lighting intensity should not be only based on instantaneous measurements of the average PPFD at growth area but also take into account the daily photon flux integral (Pinho et al. 2013).

There is an increasing trend for consumers to become more aware of the environmental impacts of purchased products, namely, of food products. Therefore,

improvements on energy-efficient horticultural lighting can contribute to increasing the market demand for high-quality horticultural products by consumers. In order to address the energy, environment, and market challenges faced by the horticultural industry, innovative technology and intensive investment in research are needed. This should be aimed at the optimization of systems involved in all steps of the horticultural production chain in closed environments. The need for innovative and energy-efficient technologies can partially explain the wide interest in LEDs.

Final Remarks and Future Aspects

The rapidly growing of the world's population presents a remarkable challenge to the agriculture sector and global food security in the future. The agriculture sector is responsible for feeding this rapidly growing population. Unfortunately, this sector is also responsible for a significant part of the primary energy requirements in developed countries that are directly or indirectly related to important negative environmental impacts (e.g., greenhouse gas emissions). To feed the world's growing population in the future, which is forecasted to reach 9.1 billion people by 2050, presents a huge challenge for crop production due to the limited and unevenly distributed resources of available area of arable land and fresh water resources. Climate change and resultant increasingly frequent episodes of extreme weather conditions make crop production based on conventional agricultural methods more susceptible to failure.

In order to address the previously described challenges, it is necessary to develop new crop production methods and technologies which can make more efficient use of energy, arable land, and fresh water resources independently of weather conditions and geographic location. Horticultural crop production in controlled and closed environments can make more efficient use of land and water. However, these facilities do require considerable amounts of energy inputs to maintain and control the optimized artificial growth environment, making them extremely energy-intensive agricultural practice. The energy is used for several purposes such as heating, cooling, ventilation, fertigation, and lighting. However, the importance of artificial lighting is becoming more apparent especially in cultivation facilities where the possibilities of the utilization of daylight are reduced or nonexistent (e.g., vertical or underground cultivation). The use of artificial light allows the optimization of plant growth, development, and nutritional quality. This can be achieved through appropriate control of spectrum, intensity, and periodicity of the light. However, in order to achieve environmentally sustainable crop production, innovative and energy-efficient technologies are needed. LED lighting is among these technologies. However, the transition to these new technologies should take into consideration not only economic but also environmental aspects. In spite of the higher initial costs of horticultural LED lighting, there are indications that its environmental impact is significantly less than conventional lighting solutions. In the future, the performance of plant growth facilities can benefit from the full integration of lighting parameters in the overall automatic control management. The possibility of full control of light

intensity and spectrum with LEDs can be carried out simultaneously with control of other abiotic parameters such as CO₂ concentration, temperature, and humidity in order to minimize energy usage and maximize crop productivity.

Finally, the importance of horticultural lighting may not be just limited to industrial or commercial applications. It is known that horticulture, and in particular the proximity to plants, promotes the well-being of individuals by providing positive and physical environment and surroundings to work and live (Relf 1992). Intensive plant cultivation carried out in protected or indoor environments such as office buildings, house buildings, or dedicated installations (i.e., vertical farms, urban greenhouses, or plant factories) located in urban areas near to consumers is generally referred to as *urban horticulture*. Urban horticulture will facilitate the proximity between production and consumption and could contribute to reduce the CO₂ footprints of horticultural products by reducing transportation needs and the related environmental burden. Additionally, urban horticulture can contribute to the prevention of diseases of populations living in these areas by promoting healthier diets and food habits. It is known that plants are source of food, medicines, and oxygen which are essential for human survival. Urban horticulture offers new possibilities to improve the well-being of populations living in urban areas by promoting physical activity, healthier food habits, and reducing stress levels.

Light is crucial for the existence of life on Earth. Therefore, its vital importance is irrefutable. Artificial lighting and electrification were also important for the democratization and development of knowledge by allowing people to artificially extend the daytime which could be used, for instance, for learning activities. Similarly, artificial lighting may now provide an important contribution to address important global challenges which are deeply related to our existence. Access to food and the urgent need of promotion of healthy diets based on fruits and vegetables, grown sustainably and independently of weather conditions and geographic location, are among these challenges.

References

- ASTM G173-03 (2008) Standard Tables for Reference Solar Spectral Irradiances: Direct Normal and Hemispherical on 37° Tilted Surface, ASTM International, West Conshohocken, PA, www.astm.org
- Both AJ, Albright LD, Chou CA, Langhans RW (1997) A microwave powered light source for plant irradiation. *Acta Hort (ISHS)* 418:189–194
- Boyle G (2004) *Renewable energy: power for a sustainable future*. Oxford University Press, Oxford in Association with The Open University, Milton Keynes
- Brown CS, Schuerger AC, Sager JC (1995) Growth and photomorphogenesis of pepper plants under red light-emitting diodes with supplemental blue or far-red lighting. *J Am Soc Hortic Sci* 120:808–813
- Cashmore AR, Jarillo JA, Wu YJ, Liu D (1999) Cryptochromes: blue light receptors for plants and animals. *Science* 284:760–765
- Chang C, Hong G, Li Y (2013) A supplementary lighting and regulatory scheme using a multi-wavelength light emitting diode module for greenhouse application. *Light Res Technol* 46:548–566

- Chen M, Schliep M, Willows RD, Cai Z, Neilan BA, Scheer H (2010) A red-shifted chlorophyll. *Science* 329:1318–1319
- Christie JM, Briggs WR (2001) Blue light sensing in higher plants. *J Biol Chem* 276:11457–11460
- Christophe A, Moulia B, Varlet-Grancher C (2006) Quantitative contributions of blue light and PAR to the photocontrol of plant morphogenesis in *Trifolium repens* (L.). *J Exp Bot* 57:2379–2390
- Craig DS, Runkle ES (2013) A moderate to high red to far-red light ratio from light-emitting diodes controls flowering of short-day plants. *J Am Soc Hortic Sci* 138:167–172
- Cuello JL (2002) Latest developments in artificial lighting technologies for bioregenerative space life support. *Acta Hort (ISHS)* 580:49–56
- Darwin C, Darwin F (1881) *The power of movement in plants*. Appleton, New York
- EIA (2013) *International energy outlook 2013* DOE/EIA-0484
- Fankhauser C (2001) The phytochromes, a family of red/far-red absorbing photoreceptors. *J Biol Chem* 276:11453–11456
- Folta KM, Maruhnich J (2007) Green light: a signal to slow down or stop. *J Exp Bot* 58:3099–3111
- Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C (2010) Food security: the challenge of feeding 9 billion people. *Science* 327:812–818
- Goins GD, Yorio NC, Sanwo MM, Brown CS (1997) Photomorphogenesis, photosynthesis, and seed yield of wheat plants grown under red light-emitting diodes (LEDs) with and without supplemental blue lighting. *J Exp Bot* 48:1407–1413
- Grigoriadis M (2013) *Life cycle assessment in horticultural lighting*. Dissertation, Aalto University
- Gueymard CA, Myers D, Emery K (2002) Proposed reference irradiance spectra for solar energy systems testing. *Sol Energy* 73:443–467
- Hart JW (1988) *Light and plant growth, Topics in plant physiology: 1*. Unwin Hyman, London
- Johnston CW (2003) *Transport and equilibrium in molecular plasmas: the sulphur lamp*. Dissertation, Technische Universiteit Eindhoven
- Mangon H (1861) Production de la matière verte des feuilles sous l'influence de la lumière électrique. In: Mallet-Bachelier (ed) *Comptes rendus hebdomadaires des séances de l'Académie des sciences*. MM. les secrétaires perpétuels, Paris, pp 243–244
- McCree KJ (1965) Light measurements in plant growth investigations. *Nature* 206:527–528
- McCree KJ (1972a) The action spectrum, absorptance and quantum yield of photosynthesis in crop plants. *Agric Meteorol* 9:191–216
- McCree KJ (1972b) Test of current definitions of photosynthetically active radiation against leaf photosynthesis data. *Agric Meteorol* 10:443–453
- Muijzenberg EW (1980) *A history of greenhouses*. Institute of Agricultural Engineering, Wageningen, Netherlands
- Pinho P, Jokinen K, Halonen L (2012) Horticultural lighting – present and future challenges. *Light Res Technol* 44:427–437
- Pinho P, Hytönen T, Rantanen M, Elomaa P, Halonen L (2013) Dynamic control of supplemental lighting intensity in a greenhouse environment. *Light Res Technol* 45:295–304
- Relf D (1992) Human issues in horticulture. *Hort Technol* 2:159–171
- Rizzini L, Favory J, Cloix C, Faggionato D, O'Hara A, Kaiserli E, Baumeister R, Schäfer E, Nagy F, Jenkins GI, Ulm R (2011) Perception of UV-B by the *Arabidopsis* UVR8 protein. *Science* 332:103–106
- Schuenger AC, Brown CS, Stryjewski EC (1997) Anatomical features of pepper plants (*Capsicum annuum* L.) grown under red light-emitting diodes supplemented with blue or far-red light. *Ann Bot (Lond)* 79:273–282
- Spalding EP, Folta FM (2005) Illuminating topics in plant photobiology. *Plant Cell Environ* 28:39–53
- Terrien J (1972) News from the Bureau International des Poids et Mesures. *Metrologia* 8:32
- Thomas B (2006) Light signals and flowering. *J Exp Bot* 57:3387–3393
- Tibbitts TW (1993) *Terminology for photosynthetic active radiation for plants*. Publication of CIE. Vienna, Austria, 106, pp 42–46

- Tibbitts TW, Morgan DC, Warrington IJ (1983) Growth of lettuce, spinach, mustard, and wheat plants under four combinations of high-pressure sodium, metal halide, and tungsten halogen lamps at equal PPFD. *J Am Soc Hortic Sci* 108:622–630
- Waldemar N (2010) The quantum SI – towards the new systems of units. *Metrologia* 47:139
- Wang Y, Noguchi K, Terashima I (2011) Photosynthesis-dependent and -independent responses of stomata to blue, red and green monochromatic light: differences between the normally oriented and inverted leaves of sunflower. *Plant Cell Physiol* 52:479–489
- Wheeler RM, Mackowiak CL, Sager JC (1991) Soybean stem growth under high-pressure sodium with supplemental blue lighting. *Agron J* 83:903–906
- Yamazaki A, Tsuchiya H, Miyajima H, Honma T, Kan H (2000) Application of red laser diode as a light source for plant production. In: Kubota C, Chun C (eds) *Transplant production in the 21st century*. Kluwer Academic, Dordrecht, pp 119–124
- Yamazaki A, Tsuchiya H, Miyajima H, Honma T, Kan H (2002) Growth of rice plants under red laser-diode light supplemented with blue light. *Acta Hort (ISHS)* 580:177–181
- Yanagi T, Yachi T, Okuda N, Okamoto K (2006) Light quality of continuous illuminating at night to induce floral initiation of *Fragaria chiloensis* L. CHI-24-1. *Sci Hortic* 109:309–314
- Zhu X, Long SP, Ort DR (2008) What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Curr Opin Biotechnol* 19:153–159
- Zhu X, Long SP, Ort DR (2010) Improving photosynthetic efficiency for greater yield. *Annu Rev Plant Biol* 61:235–261

Museum and Exhibition Lighting

Jean-Jacques Ezrati

Contents

Introduction	722
What Is a Museum?	722
Museum Exhibitions	722
Light and Lighting	722
Light as a Degradation Factor	723
Sensitivity of Cultural Heritage Objects	723
Energy of Light	724
Characteristics of Light Sources	725
Reducing the Damage Caused by Light	729
Light as an Ergonomic Element	730
Illuminance Levels	730
Light Quality	731
Lighting as a Means of Expression	731
The Luminous Variables	732
Meaning Units	732
Lighting Recommendations for Other Parts of the Museum	734
Conservation and Restoration Workshops	734
Maintenance Workshop	734
Artworks Storage Area	734
Summary	735
References	735

Abstract

When considering indoor lighting, a museum has unique requirements in terms of the quality of illumination needed, taking into account that it may host a variety of different activities simultaneously. This article will consider these requirements,

J.-J. Ezrati (✉)

Centre de restauration et de recherche des musées de France - C2RME, Paris, France

e-mail: jean-jacques.ezrati@wanadoo.fr

developing a particular focus on exhibition lighting in museums. It will not be limited to conservation issues and problems of visual ergonomics, which are objective factors, but it will also emphasize the role played by lighting as an element influencing subjective factor, similar to other elements of scenography.

Introduction

What Is a Museum?

A museum is a nonprofit, permanent institution in the service of society and its development, open to the public, which acquires, conserves, researches, communicates, and exhibits the tangible and intangible heritage of humanity and its environment for the purposes of education, study, and enjoyment (ICOM 2007).

This definition reveals that the exhibition is only one activity among many others and that when we work on museum lighting, the requirements of these other activities should be integrated into the general lighting plan. In particular, consideration should be given to exhibition management including preparation workshops, storage space for materials, etc.; collection management including workshops for conservation; and public areas including auditoriums, libraries, and cafeterias.

Museum Exhibitions

We can differentiate between two types of exhibition: the permanent exhibition and the temporary exhibition. If the permanent exhibition is the heart of the museum (often around a collection that was purposely built for the museum), temporary exhibitions are the lungs which keep the museum refreshed. It is mainly through temporary exhibitions that the museum will shine, if the main collection is not of primary importance as in the great museums of fine arts such as the Louvre, the British Museum, or other national institutions.

The lighting requirements are slightly different for these two kinds of exhibition, especially with regard to ongoing technical maintenance.

Another important difference is the type of museum. To give the two extremes, in the case of a fine arts museum, the lighting will often be as discreet as possible, whereas society museum lighting may have an integral expressive value, as important as the other elements of the museum such as the color and the spatial arrangements.

Light and Lighting

We should also define two terms: light and lighting. In this context, light is considered as a material, even if in reality it is energy. This material can be in its natural state – as daylight – but it can also be produced by artificial means, as electric light. Any natural or artificial light can be described by its spectral composition and

its luminous flux (and other criteria that derive from them). These performances have developed over the years with relevant to technological advances. Lighting is the technique of light control in response to an identified need. It is based on the knowledge acquired, not only of light sources and control technology but also on more specific properties (historical, sociological, artistic, etc.) of the objects being illuminated.

Light for Museum

As we mentioned above, the activities of museums are many and varied, so the light requirements will differ. For example, for conservation workshops we will focus on color rendering, and for exhibitions we will also take into consideration the quality, lifetime, and light distribution. There is no “perfect” light source for any museum but only a light source which can be adapted to a specific task. We need to know the characteristics of each source and then make our choice accordingly. The main characteristics of a light source are (AFNOR 2000; IBN 1980; IESNA 1996; Michalski 1987; Shaw 2001; Thomson 1986):

- Spectral composition
- Luminous efficiency
- Color rendering
- Lifetime

Knowledge of these characteristics will ultimately determine our choice of light source in any given situation.

Exhibition Lighting

Exhibition lighting is designed to support other facets of scenography (such as sound, writing, color, and placement of objects) to enhance communication, learning, and enjoyment on the part of the guest, but simultaneously the exhibition lighting has to ensure the conservation of exhibits on display. In the context of the museum, the exhibition lighting is:

- An important factor of contributing to degradation
- An element of visual conditions
- A means of expression

Light as a Degradation Factor

Sensitivity of Cultural Heritage Objects

Many cultural heritage objects are sensitive to light, especially objects with organic origin. It is necessary, first, to know the physicochemical nature of the objects in order to deduce their sensitivity to light. It is customary to classify them into four categories: no sensitivity, low sensitivity, medium sensitivity, and high sensitivity (Table 1).

Table 1 Light sensitivity classification of cultural property from CIE 157:2004 (2004)

Category	Material description
1. No sensitivity	The exhibit is entirely composed of materials that are insensitive to light . Examples: most metals, stone, most glasses, ceramic, enamel, most minerals
2. Low sensitivity	The exhibit includes durable materials that are slightly light sensitive . Examples: most oil and tempera paintings, frescoes, undyed leather and wood, horn, bone, ivory, lacquer, some plastics
3. Medium sensitivity	The exhibit includes fugitive materials that are moderately light sensitive . Examples: most textiles, watercolors, pastels, prints and drawings, manuscripts, miniatures, paintings in distemper media, wallpaper, and most natural history exhibits, including botanical specimens, fur, and feathers
4. High sensitivity	The exhibit includes highly light-sensitive materials. Examples: silk, colorants known to be highly fugitive, most graphic art, and photographic documents

Energy of Light

Damage depends not only on the type of material but also on the spectral composition of light. Light is a form of energy. Ultraviolet (UV) and visible radiation contains enough energy to fuel the chemical reaction in organic materials which can lead to degradation of the materials in the forms of color fading, yellowing, weakening, and disintegration of the materials. It is important to note that although the energies of optical radiation can be compared to the bond energies found in organic compounds, many degradations are not result of the direct rupture of covalent bonds by light. Instead, the light provides energy to make the molecules into their excited states which are much more reactive than the molecules in the ground state (before absorbing the light). With the presence of oxygen and moisture, these electronically excited molecules will react to form new molecules, causing most of photodegradation (Brill Thomas 1980).

The photon energy (E) is inversely proportional to its wavelength.

$$E = h.c/\lambda$$

where

h is Planck's constant: $6,625 \times 10^{-34}$ j.s

c is the velocity of light: $2,998 \times 10^8$ m/s

E, the energy, can be expressed in kilocalories/moles (Kcal/m), Joules (J), or electron volts (eV). In the latter case $E_{(eV)} = 1,240/\lambda(\text{nm})$.

λ in nm		380	450	550	650	760	
	UV						IR
E in eV		3,26	2,75	2,25	1,90	1,63	

The essential characteristic of the energetic influence of optical radiation is that its action is cumulative, so we need to take into account illuminance but also the exposure time. That is to say that the effect of an illuminance (illumination level) of 100 lx for 10,000 h (3–4 years of exposure in a museum) – which builds to 1,000,000 lx – is the same as the effect of an illuminance of 500 lx for 2,000 h. This is called the reciprocity law. It is important to note that we use the concept of illuminance in lux and not the physical reality concept of irradiance, expressed in watts, for reasons of ease: it is far more usual to measure light (visible radiations) received by our human visual system with a light meter than the energy received by a material with a radiometer.

The heat generated by the light sources in exhibits and their environments (mainly from the sun and halogen lamps) is also a major cause of degradation. The most significant factor is the change in the relative humidity of the object and its environment, especially for objects containing organic materials, such as wood, ivory, paper, and textiles.

Characteristics of Light Sources

Daylight

Daylight comprises both direct light from the sun and sunlight scattered by particles in the atmosphere. Besides visible light, solar radiation contains, among others, ultraviolet and infrared radiation. Its spectral content, intensity, and spatial distribution depend on the weather conditions, the hour of the day, the season of the year, the latitude, and openings (windows). It can be modified by the use of different glasses, curtains, and so on. Accordingly, the color temperature can span a very wide range: it can be as low as 2,500 K (at sunset) and as high as 20,000 K (clear blue sky in a continental area) or even more under particular conditions and in certain sites. The illuminance can also vary to a large extent (several tens of thousands of lux for direct sunlight) (CIE (1970)).

Incandescent Lamps

Nowadays, the only incandescent sources in use are tungsten-based lamps. A tungsten filament is heated by the passage of an electric current, resulting in an emission spectrum that approximately follows that of a black body radiator. In tungsten-halogen lamps, the light bulb is filled with a halogen gas to increase the lifetime of the lamp. Most of the radiation emitted by tungsten lamps is located in the near-infrared region, and only a small fraction of UV radiation (about 1 % of the visible radiation) is emitted. Although the percentage of ultraviolet radiation emitted by tungsten-halogen lamps is greater than that from tungsten lamps, as mentioned above, most tungsten-halogen lamps sold in the EU and the USA are of the UV-STOP variety (or a similar brand name). They have an ultraviolet filter incorporated in the glass envelope to reduce ultraviolet emission below the level emitted by a standard tungsten lamp. The addition of the halogen gas shifts the emission peak toward the shorter wavelengths and, accordingly, the color temperature rises from

about 2,700 K (tungsten filament) to about 3,200 K. The general color rendering index CRI Ra (Norme XP CEN/TS 16163, AFNOR, Paris 2014) is excellent (about 100; the CRI Ra is calculated from the first eight colors of the CIE test color sample, of the Munsell atlas, are relatively low saturated colors. These eight colors selected by CIE, over 60 years ago (Nickerson 1960), were chosen, primarily, to classify fluorescent tubes. Waiting for a real quality index color rendering, to judge the quality of white LEDs, the lighting community requested adding a ninth color, saturated red, R9, to the CIE-Munsell color chart which made a total of 14 colors, R1 to R14.) For the purpose of energy saving, it is recommended that only ECO or IRC types of halogen lamps are used, as these use roughly 30 % less energy. Their luminous efficacy is about 25 lm/W and their life duration about 4,000 h. These lamps are easily dimmed but with a reduction of color temperature (CEN/TS 16163 2014).

Metal Halide Lamps

These are gas discharge lamps, like fluorescent lamps, but under a higher pressure. The discharge takes place in a small space: the burner, in which mercury is added to rare earths and other metal halides. For high-quality lamps, the emission spectrum consists of a multitude of lines equivalent to a continuous spectra of up to a CRI near 95. The latest ceramic burners offer greater stability of flux and color temperatures. For these lamps, the color temperature ranges between 3,000 and 6,000 K. The electric power required for exhibition areas ranges from 20 to 400 W, with a lifetime ranging from 10,000 to 15,000 h and a luminous efficiency greater than 90 lm/W. They emit a low amount of UV radiation. These lamps are **not** easily dimmed (CEN/TS 16163 2014).

Fluorescent Lamps

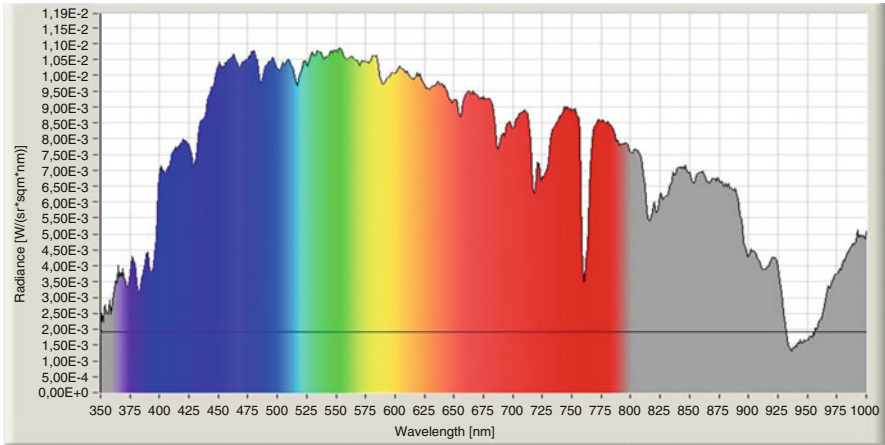
Fluorescent lamps are based on the discharge of electricity between two electrodes in a tube filled with low-pressure vapor or gas (typically mercury). The tube is coated with phosphors, which reemit radiation in the visible region when they absorb the UV radiation generated by the electric discharge. The emission spectrum of these lamps is a continuum over which the emission lines of the gas are superimposed. The color temperature and the color rendering index of fluorescent lamps depend on the blend of phosphors, and they span, as a rule, from 3,000 K up to about 6,500 K and more (7,500, 8,000 and 15,000 K) and CRI from 80 to 98, respectively. Luminous efficacy is about 100 lm/W and the life duration about 10,000 h. These lamps are easily dimmed when equipped with appropriate dimming ballasts (CEN/TS 16163 2014; Fig. 1).

Solid State Lighting

Solid-state lighting (SSL) is a type of lighting source which uses semiconducting materials to convert electricity into light. We can distinguish further between light-emitting diodes (LEDs), organic light-emitting diodes (OLEDs), and polymer light-emitting diodes (PLEDs).

At present, LEDs are more and more widely used for lighting in museums and galleries. They are based on semiconductors which emit light after the application of a suitable current. Individual LEDs emit a single color of light; however, it is possible to obtain white light if the surface of a blue LED is coated with fluorescent materials emitting in the green to red region. This converts the original emitted color and produces a clean white light. The color temperature of white LEDs can vary between 3,000 and 6,000 K. Luminous efficacy of the complete lighting system is about 100 lm/W at present, with a CRI R_a of about 95 and $R9$ above 90. The luminous efficacy and CRI increase with each new generation. The lifetime (for 70 % of luminous flux) is about 30,000 h and above but should improve in the future with further technical development. The lifetime is a function of the thermal

Daylight



Halogen

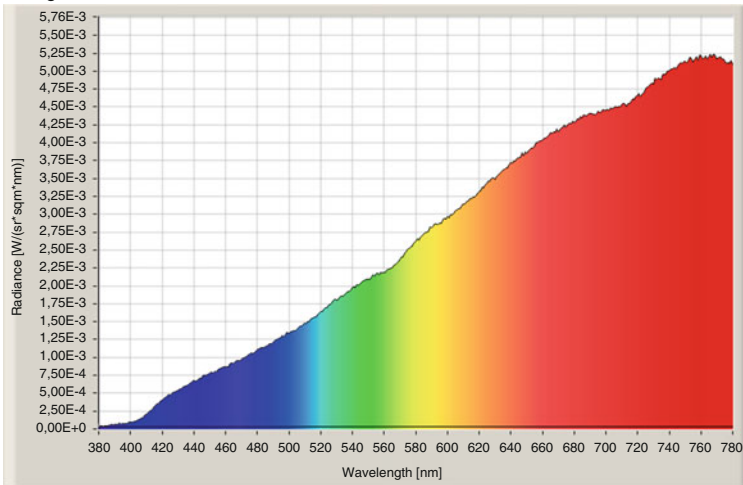


Fig. 1 (continued)

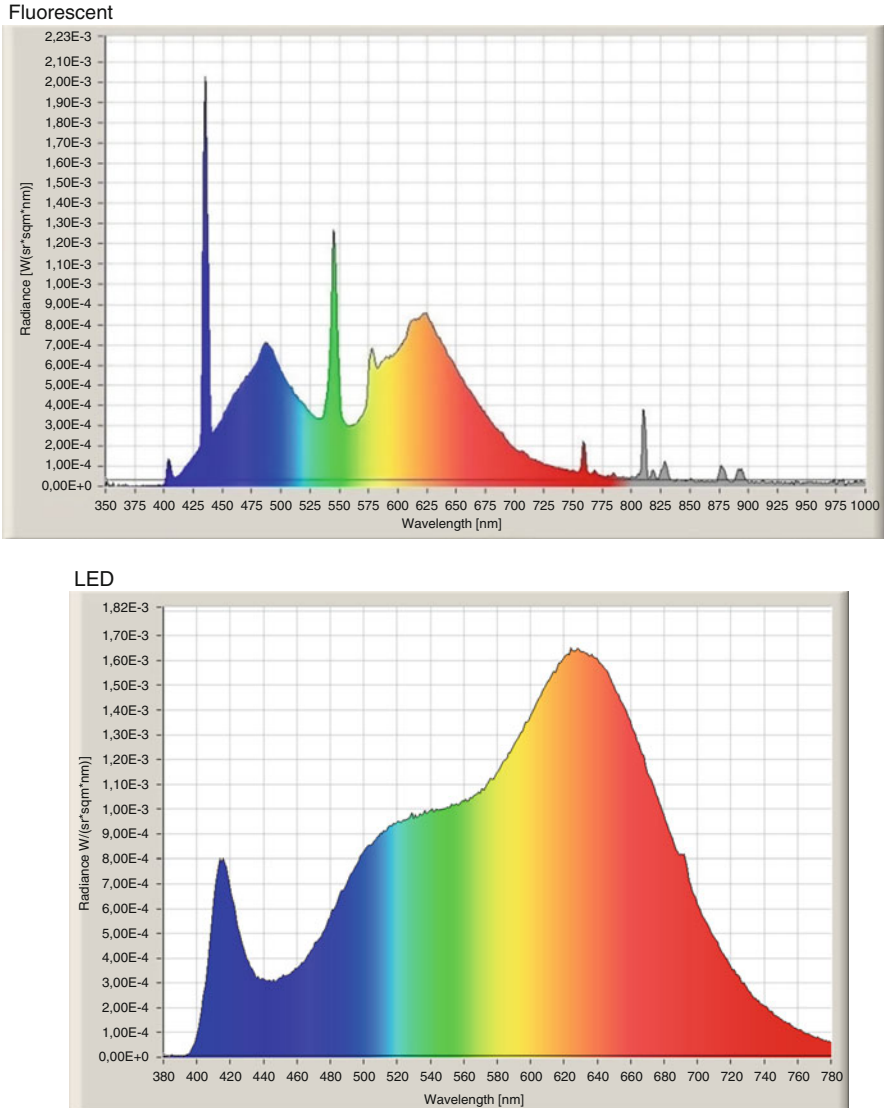


Fig. 1 Emission spectra of light sources with a CRI > 90

dissipation and any increase of the operating temperature, especially for high-power LEDs (with power greater than 1 W), will reduce their lifetime. It is possible to dim these sources when equipped with appropriate dimming drivers.

OLEDs and PLEDs are a rising technology and these lights are generally produced in the form of a flexible panel or screen. This potentially makes them useful for back lighting of transparencies. SSL sources are easily dimmed when equipped with appropriate dimming electronic power supply (CEN/TS 16163 2014).

Reducing the Damage Caused by Light

Strategies for reducing the degradation caused by light are based on very simple principles: eliminating the ultraviolet and infrared radiation and reducing the visible radiation. Knowing that the action of light is cumulative, we have to judge the total amount of light, that is to say, the light exposure (illumination level by the number of hours of exposure to light) that will be measured in lux hours.

Ultraviolet Elimination

Ultraviolet radiations are very energetic because of their short wavelength, as we have seen previously. These radiations thus represent a major risk to many organic objects if they exist in large proportions (about 6 % of optical radiations). One must try to eliminate them.

From Natural Light (Daylight)

Any new construction at a cultural heritage site must be fitted with laminated glazing. This glazing can be clear or tinted, with properties designed for other purposes in addition to UV elimination (security, thermal insulation, etc.). Glazing reduces entry of ultraviolet radiation by more than 95 %. In existing buildings, it is possible to stick on the windows some plastic films sold specifically for this purpose which are found in the stores of security film installers. These films can be clear or tinted.

From Artificial Lights (Electric Lights)

First, we choose conventional light sources (halogen and metal halide) with a glass envelope (bulb) designed to absorb UV. These are labeled with “no UV” or “UV stop.” This is especially important for a metal halide source emitting UV radiation in relatively large amounts. If not, it is necessary to equip the devices using these sources with relatively expensive glass filters. For fluorescent lamps, the temperature of the lamp is not very important so a simple anti-UV polyester filter (# 226 LEE Filters or # 1510 GAM) will be sufficient. The easiest way now is to use LEDs which are naturally devoid of UV emission.

Infrared Elimination or Reduction

Infrared radiation is also damaging and the resulting effects are typically irreversible. Most often, we try to minimize them.

For natural light, we should proceed by a reduction of sun openings side or the use of solar protection glazing, or films, and outdoor blinds.

If using artificial lighting, infrared filters or lamps with infrared reflectors (dichroic) or optical fiber systems should be employed. Again solid state lighting now offers the best solution for reduction of IR emission.

Reduction and Control of Visible Radiation

Visible radiation is the cause of photochemical degradation, so it must be controlled with great attention. Degradation due to light is cumulative, hence the importance

Table 2 Annual luminous exposure limits for different material sensitivities

Material classification	Annual luminous exposure
1. Insensitive	No limit (for conservation)
2. Low sensitivity	600,000 lx · h per year
3. Medium sensitivity	150,000 lx · h per year
4. High sensitivity	15,000 lx · h per year

given to the dose received per unit of time. For reasons of convenience, we work with visual units of illuminance and light exposure doses which are lux (Lx) and lux hours per year (Lx h/a). The measurement made with a light meter is an instantaneous measurement, which will be accurate if the lighting is stable. For dynamic measurements, integrating light meters exist, which will measure the brightness level at fixed intervals (e.g., every minute) to integrate in an hour resulting in a number of lux hours for the period studied. There are also chemical dosimeters such as the LightCheck[®] (Römich et al. 2004) ones, used to estimate the dose received by the exposed object during a period of time. They are less accurate than a light meter integrator but are very simple to use for the monitoring of light exposure. Table 2 shows the annual light exposure limits for different material classifications.

Reduction and control of visible radiation of daylight is possible by using solar control glazing or films, with shutters or roller blinds and limitation of the opening hours. It is easier with electric lighting to reduce light exposure using combinations of elements including detection systems, time lag switches, and dynamic lighting.

Light as an Ergonomic Element

Seeing and enjoying an artwork requires not only a certain level of illumination but also a certain quality of light. In addition, we should avoid all the visual noise – such as reflections and glare – in most situations to arrange a setting, you can make some choices that do not appeal to everyone, but in terms of ergonomics, there can be no question of semantic choices.

Illuminance Levels

For objects which have the most sensitivity to light, the recommended lighting levels are at the limit of color vision not sufficient for viewing purposes (50 lx compared with 1,500 lx recommended in the color industry or for the restoration of a work of art). Our natural aging increases the difficulty; we must, for the same reading performance, double the light quantity required for a 60-year-old compared to the light needed at the age of 20. This has to be taken into account to avoid depriving a large number of visitors of an interesting visit to the exhibition.

Light Quality

The human visual system has evolved and adapted to be stimulated by daylight, which is to say that it responds to a complete and balanced spectrum in the visible range. In contrast, luminous efficiency, which was the only criterion considered by the industrial lighting sector, can be produced by sources whose spectral composition consists of three individual spectra that, by additive mixing, restituted white light. Based on this principle, we see the development of fluorescent tubes (so-called high efficiency or compact) which are invading the lighting space. Complementing daylight, this trichromatic light can be acceptable, but as a light alternative (in the absence of natural lighting), it is not satisfactory. Our visual system is constantly asked to rebalance the spectrum, and increasingly we feel visual fatigue. When viewing colorful works with multiple hues, the failure to provide continuous spectral composition does not facilitate the fine discrimination of colors. In addition, some of these sources can give rise to problems of metamerism, that is to say two different colors of shade under a full spectrum light, such as sunlight, may seem similar under these sources. Out of halogen lamps, only full spectrum fluorescent tubes or metal-halide lamps or LEDs with a CRI >95 are recommended. (References to these tubes begin by any number 9 as 930, 950, and 965 depending on the selected color temperature. Their color rendering 95 will warm tint (3,000 K) to 98, the coldest shade (6,500 K).)

Lighting as a Means of Expression

If the lighting choices related to conservation and visual ergonomics were pragmatic and objective, those made of plastic criteria represent the subjective part of the exhibition lighting. Lighting design comprises a number of choices; as such, it is not meaningless and has all the characteristics of a semiotic system. Lighting is a language with its own syntax. An idea or a spatial scenario – which are part of the scenic language – can be enhanced through lighting: lighting is meaningful.

Choosing a specific type of lighting for a scene can induce a particular emotion or feeling in the viewer. One must follow rules to build and choose lighting conditions: *those that determine the constitution of units, those that preside over the combination of these units, and those which preside over pragmatic use of the latter.*

Units can be organized in three levels; the first one is the luminous variables (This term, luminous variables, is an extrapolation of that set by Fernand Saint-Martin (1987) in his book *Semiotics of visual language*, themselves derived retinal variables of Jacques Bertin in 1967 in *Graphic Semiology*.) which is defined by the ten basic variables including the intensity, the direction, and the color temperature of a light spot. The second level consists in the combination of at least two luminous variables to make sense units. And the final one is the assembling of the sense units to make up the light design.

The Luminous Variables

These variables can be divided into several groups, those grouped under the term plastic variables, geometric variables (that appeal to the laws of geometry such as reflection, diffraction, scattering, etc.), spatial variables, and time-related variables.

Plastic Variables

- Light intensity, or more precisely the illuminance, on a dazzling vs. darkness axis. We can measure this variable as a relative value – a percentage from 0 % for black to 100 % for full light – or as an absolute value in lux, the illumination unit.
- The chroma: this variable measures the hue (blue, green, red, etc.) as well as the saturation of the color, thanks to the measurement of the chromaticity coordinates (e.g., x and y in the case of the CIE system).

Geometric Variables

- The extent – or the aperture angle – of a part (or all) of the illuminated object (on an open vs. closed axis or flood vs. spot axis): this angle can be expressed in degrees.
- The texture, that is to say, the sharpness of the spot – from full transparency to full diffusion – can be measured by a diffusion factor ranging from 0 to 100.
- The shape of the light beam, developed around concepts such as fuzzy vs. net.

Spatial Variables

- The location (or the specific point) is the choice to illuminate a specific part of the scene – global versus local.
- The contrast – or brightness ratio – between the background and the subject will be held on a full-face versus backlight axis.
- Direction, i.e., the projected shadow: on a left versus right axis or front versus rear or East versus West and North versus South.

A Time-Related Variable

- The movement: on a chronological axis (before vs. after) or in a 3D space (a point or a topographic sequence in the (u, v, z) space). This variable can be related to the movement of light but also to the viewer's movement

Meaning Units

The previous variables can be combined together in order to carry a message.

Working within a particular semiotic, that of visual semiotics, moreover, exhibition lighting (i.e., an applied semiotic), rather than using the “Saussurean” schema, we will rely on a triadic model of Peirce inspiration (Klinkenberg 1986). We will illustrate these notions with two examples, the first one dealing with general lighting in a showroom and the other one dealing with lighting focused on an artwork.

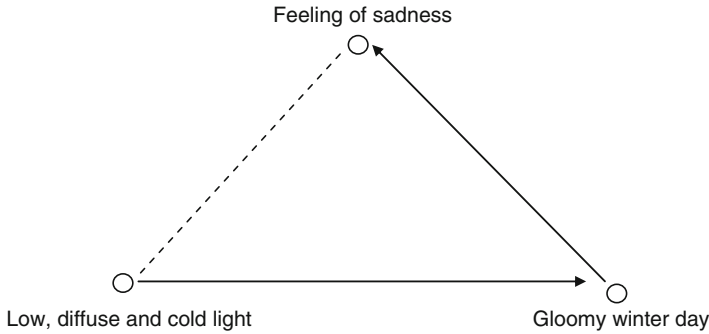


Fig. 2 Semiotic process of the building of a lighting atmosphere

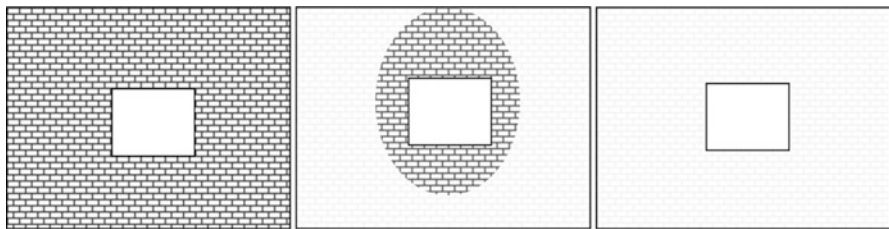


Fig. 3 Localized lighting typology: washed lighting, focused and framed lighting

As the first example, let us imagine a room uniformly illuminated by a low, diffuse, and cold light. In our personal experience of human being, is such a light not the characteristic of one gloomy winter day? If so, in which state of mind does this gloomy winter day bring us? It would not fill us with happiness; most commonly we would feel sad and depressed. As we are thinking with the “Peircean” perspective, we can illustrate this with the following diagram (Fig. 2):

This example underlines the use of a meaning unit, using three of the ten luminous variables described above, with a certain value for each of them. Giving different values to these variables would result in quite different scenes:

Representamen	Object	Interpretant
Low, warm, and direct light	A fireplace	Tranquility, peace
Strong, warm, and diffuse light	A fire	Anxiety
Strong, cold, and direct light	A beautiful sunny summer day	Well-being

So there is a connotation action between the sign and the object, which will denote a thought to our mind; that thought can give rise to a new image (object) in our mind, which will lead to a second thought and it may continue again and again.

Let us analyze the second example: a localized lighting on a two-dimensional object. We want to work on the contextualization/decontextualization axis by the ratio between the background and the object. The choice will be made depending on this ratio: a washed light, a focused light, or even framed light (Fig. 3).

In the first case, a washed light (wall washer) will link together in the same context the object with the background. The influence of the nature of the background view in conjunction with the object is important, and the act is significant. Equally significant, but carrying a different meaning, the second case is that of focused lighting. The emphasis is on the object, and the background loses its importance and becomes less influential. However, in the last case – framed lighting – the very negation of the environment gives a strong effect to this lack, to this maximum contrast, by the imposed decontextualization.

Also in the case of this localized where we have created special units of meaning by the choices made on the shape and dimension lighting, we can add a choice of lighting levels and color temperatures. These choices, like previous ones, are not neutral and will influence our perception.

Lighting Recommendations for Other Parts of the Museum

Conservation and Restoration Workshops

There is no single type of restoration workshop, but several, all linked to the types of objects to be processed. The conceptual approach is again the same. It begins like any ergonomic study by analyzing the task, which includes the subject and the object, or one that does the job, to pursue the knowledge of the means available to help in order to decide which solutions can be considered. Clearly it is important to be able to see and see very well, that is to say, not only see the differences in color (hue, saturation, and brightness) but also clearly distinguish cracks and thicknesses, to work out the details while restoring them in a set. It is also required that one can distinguish shapes as well as the quality of the shadows. The light can be diffuse, like a cloudy sky, or directional like the sun. Naturally balanced tonality should be sought in the visual field of the operator (avoiding glare and specular highlights) and a color balance or a color neutrality of the surrounding surfaces is required.

Maintenance Workshop

General lighting with natural light is a requirement of well-being. We will preferably choose northern lighting because of its stability throughout the day. Whatever the quality of this light is, it should be supplemented by artificial lighting. If the need for artificial lighting is constant (small aperture), its high quality is essential with sources with a CRI of over 95 and a color temperature of at least 5,000 K.

Artworks Storage Area

These are special areas in the museum. Works in storage should be kept in darkness, but a reservation is constantly visited to meet the needs of the museum (changing

exhibits, loans of works, inventory, maintenance, etc.). Hence there is a need to illuminate these spaces.

The area distribution of light switches and controls is of great importance. We clearly separate the circulations, the reserve areas, and spaces dedicated to specific work performed in the same premises as photography or scanning, which often cannot be done in another dedicated space.

Summary

The art exhibition and museum lighting is not limited to the conservation of heritage objects and takes into account the needs of visitors in terms of vision but contributes to the understanding as the other elements of the design. Technological developments such as programming and automation of lighting and the performance of new sources (LEDs, OLEDs) do not change the art of lighting but facilitate its implementation.

References

- AFNOR (2002) Prescriptions de conservation des documents graphiques et photographiques dans le cadre d'une exposition. Norme AFNOR Z 40 010, Paris
- Brill Thomas B (1980) Light. Its interaction with art and antiquities. Plenum Press, New-York
- CEN/TC346 Conservation of Cultural Heritage (2014) Guidelines and procedures for choosing appropriate lighting for indoor exhibitions
- CEN/TS 16163 (2014) Conservation du patrimoine culturel – Lignes directrices et procédures concernant le choix d'un éclairage adapté pour les expositions en intérieur. AFNOR, Paris
- CIBSE (1994) Lighting guide LG8: lighting for museums and art galleries. CIBSE, London
- CIE (1970) Daylight. Publication no 16
- CIE 089/3:1991 (1991) On the deterioration of exhibited museum objects by light. CIE Technical Collection
- CIE 157:2004 (2004) Control of damage to museum objects by optical radiation. CIE, Vienna
- Ezrati J-J (2000) Museum lighting. In: Professional lighting. Design no 16. VIA, Gütersloh
- IBN (1980) Code de bonne pratique de l'éclairage des œuvres d'art et objets de collection, NBN L13-003:1980. IBN, Bruxelles
- ICOM (2007) 21st general conference of the international council of museums, Vienna, 19–24 Aug 2007
- IESNA (1996) Museum and art gallery lighting: a recommended practice, RP-30-96. IESNA, New York
- Klinkenberg J-M (1986) Précis de sémiotique générale. De Boeck, Bruxelles
- Michalski S (1987) Damage to objects by visible and ultraviolet radiation. In: Lighting in museums, galleries and historic houses. Museums Association, London
- Nickerson D (1960) Light sources and color rendering, in JOSA 50, Washington
- Römich H, Graham M, Lavedrine B, Bacci M (2004) Lightcheck: a new tool in preventive conservation. Conservation Journal 47, Victor and Albert Museum
- Saint-Martin F (1987) Sémiologie du langage visual. PUQ, Québec
- Shaw K (2001) Lighting the show. In: Lord B, Lord GD (eds) Manual of museum exhibitions. Altamira Press, Walnut Creek
- Thomson G (1986) The museum environment, 2nd edn. Butterworths, London

Landscape Lighting

Janet Lennox Moyer

Contents

Introduction	738
Design Tools that Affect Lighting Scenes	738
Power Distribution in Landscape Lighting	739
LED is Simply a New Light Source	741
Selecting the Appropriate Type of LED Lamp	743
LED Replacement Lamps	746
Landscape Lighting Fixtures with Integral LED Modules	749
Conclusion	752
References	752

Abstract

The practice of landscape lighting varies dramatically from interior lighting. No walls, no ceilings, darkness, and continual change encompass most of the variables that separate how we approach landscape lighting from interior lighting. In addition, landscape spaces typically don't have "visual tasks." Landscape lighting connects humans to their landscapes that have otherwise been essentially erased by darkness. It has the ability to calm people and help them relax in the "after-work hours." Designing for our psychological and physiological responses allows us to create spaces that feel good to people. One example is that humans visually respond to vertical surfaces before horizontal, so when we light hedges, trees, or walls, we give visitors a clue as to the boundary of a space. This aids their sense of safety.

J.L. Moyer (✉)
Janet Lennox Moyer Design, Troy, NY, USA
e-mail: janmoyerdesign@mac.com

Introduction

Landscape lighting designers need light sources that produce as little light as possible. That sounds completely opposite to what most engineers work on and strive for, yet it isn't really. The light source can be efficient, producing as many lumens per watt as possible, just not much of it. In a typical landscape, the amount of wattage used in 2013 ranges from less than 1 W to typically 5–7 W and, when a very high light level is expected or a very large tree is lit, maybe 9.5–12 W (Lennox Moyer 2013a). With halogen that would have been 20–37 W, and in most landscape spaces, over 20 W of HID is way too much. Even in “city environments,” the amount of light that we need is less than most people expect. Outside a city, we need very little light to function once our eyes adapt from higher inside levels to lower outdoor levels. Our brain/eye function has switched from photopic to at least mesopic, if not scotopic vision, so our rods are the main functioning part, and it guides our sense of a space by comparing brightnesses, identifying movement, and giving us most information in our peripheral vision.

Design Tools that Affect Lighting Scenes

Providing balance in brightness relationships from one area to another is significantly more important than the light level. The human system needs little illumination to function in a dark environment. What is necessary for providing a sense of comfort is visual connection across a space. The goal is to balance the amount of light reflecting off objects within an acceptable contrast range across a scene (Lennox Moyer 2013b). See Fig. 1.

Controlling glare is a critical issue for landscape lighting. When a viewer can see the light source directly, including any *lensing*, or the *inside of a fixture housing*, that brightness will be the highest level in the overall scene. When that is the case, it is nearly impossible to balance the brightness in a space. That will always distract from the scene and can create confusion or add to a sense of discomfort in a space. Light sources must be shielded in landscape lighting in order to create a comfortable space (Lennox Moyer 2013c). See Fig. 2.

Without walls or ceilings, we don't have large reflecting surfaces that frame the context of a room. The way surfaces are lit creates the understanding of space, and so the decision-making process of “what to light” becomes the critical starting point for creating a lit night space. Deciding on a hierarchy of brightness frames how we respond to night spaces. Lighting can “show” people how to move through a space and direct them where to go. Or it can confuse people and detract from their ability to reach their destination. Using light on the boundaries of a space, lighting can create an outdoor room. See Figs. 3 and 4.

Landscapes include both “softscape” or a palette of plant materials and water features and “hardscape” – paths, stairs, sculptures, buildings. The plant materials continually change. Plants grow, they flower, and some die. They change appearance from one season to another, from one year to another as they mature, and then people

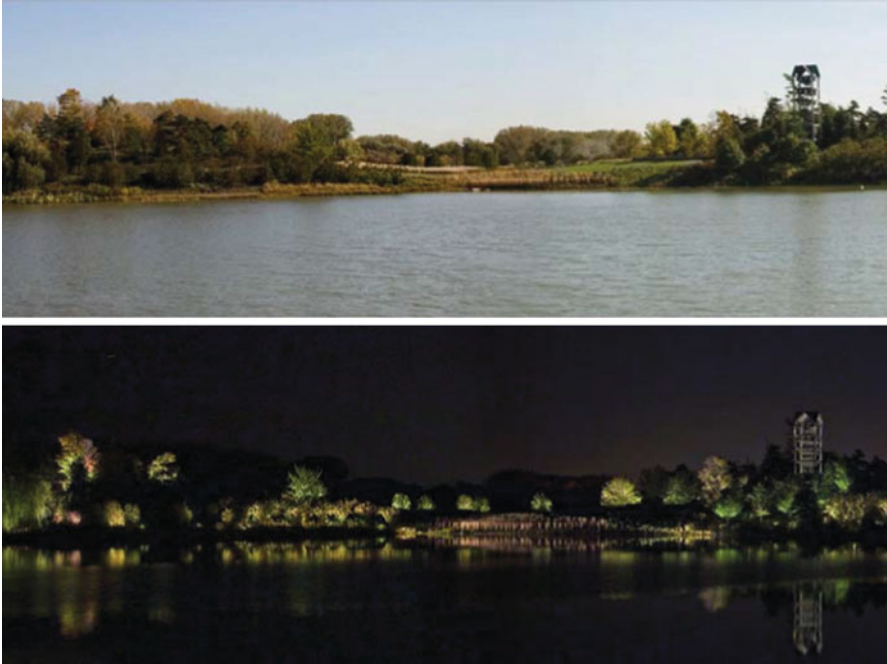


Fig. 1 Above the daytime photo of Evening Island at the Chicago Botanic Garden provides a starting point to understand the use of brightness provided across the scene. As viewed from across the lake, this scene creates the night experience, highlighting the carillon on the right – lit entirely with neon. Lighting Design, Jan Moyer Design (Photographs by George Gruel)

make changes. People add or move walkways, they introduce sculpture or new buildings, and they change the usage of space. Landscape Lighting needs to respond to all of that – from the beginning (Lennox Moyer 2013d).

Power Distribution in Landscape Lighting

Once a designer has a sense of the lighting composition for a space, the first document to be produced for construction is the power distribution plan. For the most part, today, as we retrofit existing projects to LED, the existing power distribution means that no power changes be done. One exception to that is when the load on a transformer is too small, requiring changing the transformer or transformer cassette to a lower wattage so that the LED lamps can activate.

Most landscape lighting utilizes low-voltage power (Lennox Moyer 2013e). The lighting consists of a system of fixtures, transformers to convert the voltage level, and some form of control system. It could be just on/off or it could be a computer-based control system either stand alone or integrated with the buildings' controls.



Fig. 2 On the *left* your eyes are drawn to the brightness of the inside of the fixture housings. No fixture is aimed more than 35° off nadir, so, no lamps are visible. On the *right*, an angled glare shield, made as a custom accessory in the 1980's, hides that glare and allows the viewer to see the scene. Lighting Design, Janet Lennox Moyer (Photographs by Ken Rice Photography – kenricephoto.com)



Fig. 3 Lighting, both up and down lighting creates the sense of walls, shows the ground plane, and using increased brightness adds two important clues. The first is the brightness at the arbor, which acts as a warning to be cautious as someone approaches the stairs. In the distance, the sculpture is lit more brightly than the surrounds creating a visual destination. Lighting Design, Jan Moyer Design (Photograph by George Gruel)



Fig. 4 In this photo of the same space shown lit by halogen lights in Fig. 3, all that has changed is the lighting. LED replacement lamps now create the “same” scene. Note how the color is quite different, both foliage and flowers appears more lush. The other change is that many of the shadows, important to the visual composition, have disappeared. This project is one of the learning projects teaching us that beam spread selection between halogen and LED must change. The LED beamspread needs to be narrower. Lighting Design, Jan Moyer Design (Photograph by George Gruel)

The concerns facing specifiers about LED include issues about the lamps themselves, the equipment required to now constitute the lighting fixture, and many complexities regarding how to control the LED.

LED is Simply a New Light Source

As the lighting industry evolves from our traditional sources of incandescent/halogen, fluorescent, and, HID to LED, nothing about lighting design has changed. LED, like all “new” light sources before it, is simply a new light source. See Fig. 5a, b. However, the players involved in the industry have changed. Now, a multitude of people familiar with electronics continue to design products with a short life span. Working in other, nonrelated fields before they entered lighting, they have programmed the public to expect to replace these devices regularly.

That model does not work in architectural or landscape lighting. People do not change the layout or planting of their gardens regularly. Some gardens still exist from hundreds of years ago. While those spaces have required maintenance, and the plant materials have grown substantially, the garden layout and its essence remain more or



Fig. 5 (a) On the *left, top* and *bottom*, is a typical halogen MR16 used in landscape lighting fixtures from the late 1980s until recently. The three other lamps are various LED replacement lamps. Note how the size and shape varies from the halogen lamp. Existing fixtures in gardens were designed for the halogen lamp. When considering an LED retrofit, designers need to be certain the lamp will physically fit and that the lamp is rated for outdoor use in an enclosed fixture (Photograph by George Gruel). (b) These images show four integral module LED sources. One issue designers need to consider about these LED sources is whether they are replaceable in the field. This will allow continuing to modify the existing landscape lighting over time as LED lamps evolve (Photograph by George Gruel)

less constant. Owners of gardens often balk at the maintenance cost of the garden, and then we add maintenance cost when we introduce lighting.

In 2013, when a landscape lighting specification gets to the installing contractor to produce a bid for a client, the listed products may no longer exist. If they do exist, and get purchased for the project, they may not exist 1 year later. Gardens continually expand, and the lighting needs to be consistent from one area to another. Designers and installers used to be able to get more of the same products as projects expand. That is what is needed in landscape lighting. Projects last for decades and

Fig. 6 (a, b) On the *left*, this California hillside was lit in the early 1990s. On the *right*, the owner called the lighting designer to make some changes as the garden had been renovated. First looking at these changes required some thinking time to comprehend how to have the lighting system respond. It required substantial updating (Photographs by Jan Moyer and George Gruel)



expand over that time. The lighting system needs to be flexible to respond to expansion, incorporating the existing products, or updates to the existing products. See Fig. 6a, b.

Selecting the Appropriate Type of LED Lamp

Designers have two main options today when choosing to utilize electronic light sources or light-emitting diodes. With either an existing or new system that uses standard/traditional fixtures, LED replacement lamps can replace/or be used instead of the existing traditional lamp. For new projects, designers typically specify fixtures with integral LED modules.

Several issues constitute the criteria and decision-making issues determining how LED gets incorporated into landscape lighting. The specifier needs to understand what is available and be aware of the issues that affect using this new light source. Some of those have been discussed above; specific issues will be addressed here

before we look at the two basic options available in 2013: *LED replacement lamps* and *fixtures incorporating integral LED modules*.

LED produces light differently from the traditional sources. The color production is different and does not match traditional sources. Starting with blue, the entire spectral composition shifts and will never have the same spectral makeup of the halogen or metal halide lamps typically used in landscape lighting. This brings up several issues. First, people are used to the color of the traditional sources and have expectations, especially when a halogen lamp is dimmed, shifting in color and becoming warmer. Second, in some situations both in existing or new projects, there usually is more than one light source utilized. When a project continues to include halogen or metal halide, the appearance will be different from either traditional source to the LED. Color variation can be a design tool, but it has to be controlled carefully when used as a visual tool in a composition. More typically, a color difference becomes an element in a composition that prevents visual balance in the scene – essentially causing a viewer’s eye to jump erratically from one element to another. Third, color shift due to too wide a binning range, to color shift change over time, or from one manufacturer’s LED to another causes the same problem. One may think this to be a small concern, but in reality it is a huge problem in visual composition.

While binning has improved dramatically over the last few years, variation in one set of LED lamps can still be a problem. LED lamps installed in landscape settings for over a year are experiencing color shift. While it happens gradually and isn’t dramatic, it changes the appearance of trees – both trunk and foliage impression. The biggest issue is that most landscape lighting projects include more than one manufacturer’s lighting equipment. Fixture manufacturers want their products to be different from others, and this means that the LED chip, the drivers, etc. vary from one manufacturer to another and can vary even within one manufacturer’s range of products. Designers used to have to worry about the color finish on the lighting fixture; now they have to consider the perceived color of the light produced and how that affects the objects to be lit.

One big difference in the LED light produced versus traditional light sources is a sense of vibrancy. The LED lighting seems to be more vivid, brighter, and more even across a scene. See Fig. 7a, b (Rodan front sculpture halogen and then LED). The Lambertian distribution of an LED seems to minimize or eliminate the shadows created for existing lighting scenes. This is a new phenomenon, and designers are just beginning to understand this and how they need to rethink the design as they change from a traditional source to an LED source. All LED lamps produce a main “beam” of light spread and then include more light in the “field” and beyond (often called “spill”) than traditional sources. This translates to a considerably wider distribution – no matter what beam spread is selected – and it causes lit scenes to be more even, erasing some of the shadows that were an essential part of a visual scene.

While a balance in brightness contrast is important for visual and physiological comfort, creating too even a wash over a scene makes it dull and uninteresting. This means that lighting designers have to learn a new sense of light distribution. When

Fig. 7 (a) This lighting scene uses Halogen MR16 lamps in uplights so that the four fruit trees frame the view to the sculpture at the end of the lawn. Then, uplighting behind the sculpture grazes the hedge as a backdrop. A total of four lamps, two upcoming from the left, and two downlights coming from the homes' eaves on the right highlight the sculpture. Lighting Design, Jan Moyer Design; Photograph by George Gruel.

(b) This shows the same garden as (a), but LED lamps replaced the halogen and viewed from a different angle. Note how much "brighter" the scene feels and how the areas specifically left dark are gone, as are most of the shadows. These changes to the appearance of the lighting scene may be considered acceptable or may require rethinking to more closely reproduce the original composition



retrofitting an existing project, the designer cannot simply look at wattage, beamspread, candlepower, and color temperature. While designers are starting to understand that an LED source will create more brightness than a halogen lamp having roughly twice the candlepower of the LED, designers need to rethink the coverage an LED will produce in order to try to recreate the original effects as closely as they can or to create essentially a completely new effect in a new "shade" of white.

Landscape lighting designers are familiar with the color of a halogen lamp. The color produced will vary from one manufacturer to another but in a relatively small range. LED sources start with a different white to begin with, and various color temperatures are available. Several years ago, the LED electronics specialists were trying to convince landscape lighting fixture manufacturers to use color temperatures way out of the range of normal color temperatures used in lighting – 7,000 K up to 15,000 K. The lighting industry had to push back hard and familiarize the electronics industry with the white color range that is normal and acceptable. Now, we can choose from 2,400 K, 2,700 K to 2,800 K, 3,000 K up to 6,000 K. Landscape lighting projects use the lower end of available white colors, typically between 2,700 K and 3,000 K. For a dimmed appearance- the warmer color of an

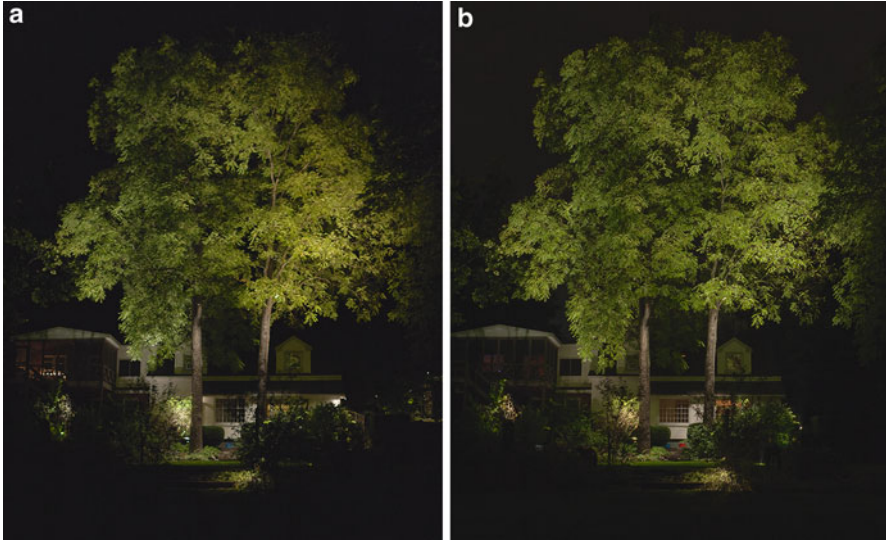


Fig. 8 (a, b) These two images compare two light sources in each photo to show how color affects a trees' appearance. The *left* photo (a) has halogen uplighting the left tree. On the *right* is 2,700 K LED shifting the perceived appearance towards yellow, the color these leaves turn in the fall as they die. The *right* photo (b) continues to have the same halogen uplighting the left tree and uses 3,000 K on the right tree. The LED lighting, in both cases, seems brighter, even though the lamps have less than half the candlepower at 0° in photometric reports

incandescent as it is dimmer – lower temperatures such as 2,400 K are becoming available, and some developers are working on providing a similar change of color, toward the warmer end of the spectrum, when dimming.

The color of white that will be preferred in landscape lighting will vary by the designers' preference, the owners' preference, and the colors in the scene. On the leaves of deciduous trees, 3,000 K looks better than a warmer temperature. See Fig. 8a, b to see a side-by-side up-lighting comparison on black walnut trees. Figure 8a shows halogen lighting on the left tree and 2,700 K Sora LED replacement lamps on the right tree. Figure 8b has halogen still on the left tree and 3,000 K Sora LED replacement lamps on the right tree.

The 2,700 K in Fig. 8a makes the leaves look as though they have started to turn color in the fall as they are dying. They do not look as good as they should. The 3,000 K in Fig. 8b makes the tree look better than either the 2,700 K LED or the halogen, and the 3 K LED lamp color emphasizes the leaves' natural appearance.

LED Replacement Lamps

For existing projects, LED replacement lamps are the easy option. Simply take out the existing lamp and put in an LED. There are many providers offering LED

replacement lamps. Designers need to be cautious in selecting an LED replacement lamp and consider the following factors:

1. **Outdoor rating** – either damp location, wet location, or an IP with the second number of 6, 7, or 8.
2. **Enclosed fixture rating** – the lamp and fixture require testing to ensure that the p-n junction is within normal working temperature range. The LED lamp specification sheet and the lamp box need to list that the lamp is rated for use in an enclosed fixture.
3. **Technical data** including voltage, wattage, life, candlepower, lumens, beam spread, color temperature, dimmability, and method required for dimming. All this data, except for lumens, is critical for a specifier to have to evaluate whether a lamp will work in each fixture that is specified for every project.
4. **Physical size**- does a specific lamp physically fit into the existing fixture. This may seem like a simplistic concern, but not all lamps fit in all fixtures. The only way to know for sure is to test the combination together.

Not all LED replacement lamps are rated for outdoor use. Typically, the moisture rating being given to lamps that can be used in an exterior setting is “Damp Location”. While the lamp itself is not usually exposed directly to rain, irrigation, etc., the fixtures, while completely enclosed, do have to be opened to remove the existing and install the new lamp. Moisture can get in at that point. If for any reason the fixture is opened up again, that introduces moisture, and many below grade have to be opened for aiming, while above grade, either stake mount or tree mount, will get opened to add or change accessories, such as louvers. A Damp Location rating is introducing a risk in landscape lighting. Wet Location or IP68 are the only fail-safe ratings to ensure that water won’t compromise the lamp.

LED replacement lamps must be rated for use in enclosed fixtures for landscape lighting. Yet, that does not ensure that a particular lamp will work with a specific fixture. Which party is responsible for making the determination that a specific pairing will allow an LED lamp to function properly? Some companies are cross testing to ensure reliable results. In the USA, a program called “Works with Soraa”(<http://www.soraa.com/wws/overview>) has the fixture manufacturer test the LED replacement lamp in their fixture(s), and the LED manufacturer tests the same combination(s). When the testing is completed, they compare the results. When satisfied that the mechanical, thermal, and electrical conditions will be satisfactory, the LED replacement lamp manufacturer allows the fixture manufacturer to market their fixture(s) with the “Works with Soraa” name.

Technical data about a specific lamp needs to be available. Many companies are listing light output in either lumens only or in foot-candles. Both of these are a problem. None of the incandescent/halogen lamps have ever had lumens listed in a lamp manufacturer’s catalog. What has traditionally been listed in the lamp manufacturer’s catalog is candlepower. So that is what landscape lighting designers know about the lamps they will be comparing the new lamps against. Foot-candles are a problem as humans can’t see them; they see reflected light. Specifiers need to know

the voltage range that an LED can work successfully within; what distance a remote low-voltage transformer or driver can be located; life to L70 or better; wattage; candlepower; beam spread; color temperature – where it lands on the black body would be very helpful in understanding color produced; dimmability; and what is required to have successful dimming. These factors are far more involved than previously required. This means that a designer has to be considerably more careful in specification and needs substantially more data immediately available to them without having to search at length. Putting all this on a specification sheet on the website would be very useful.

The availability of replacement lamps is a huge issue of concern for designers. Currently, the industry has lost at least one manufacturer of high-quality lamps in less than 1 year. There are dozens of companies offering LED replacement lamps today of varying quality. Of all those, only a handful print on their specification sheets or the lamp box that the lamps are rated for outdoor use in an enclosed fixture. That narrows the field very quickly. Then, designers need to consider the light output.

To replace the MR16 halogen lamps typically used on all projects before LED, specifiers need multiple beam spreads and light outputs. Landscape lighting typically uses a 20–60° beam spread. Sometimes, narrower beams are useful for very small objects or portions of a sculpture being lit with multiple sources. All lamps should be available in at least three beam spreads in all wattages available. It would be helpful to have standardization of sockets, beam spreads, etc.

Landscape lighting needs very low output, starting with 1-, 2-, 3-W lamps. This is useful for close fixture locations – as example on interesting rock walls, see Fig. 9.



Fig. 9 To show the undulating form, texture, and patterning of this stone wall, 1 and 2 W LED uplights are placed within inches of the wall. More wattage would be too bright. The fixture location has to do both with revealing the texture, using a grazing technique, and low enough wattage to not create too much light, often called a “hot spot.” Lighting Design, Jan Moyer Design (Photograph by George Gruel)



Fig. 10 In this lighting scene, a combination of wattages are needed to create the balance of light providing cohesion across the scene. For the hedge and the tree, at right, the wattages were in the range of 4–6 W. Lighting Design, Jan Moyer Design (Photograph by George Gruel)

The next level of wattage needed is 4-, 5-, 6-W. These are useful for small trees. A variation of wattages is needed as the trunk and branching structure of plant material varies dramatically. See Fig. 10. For larger objects including trees, 7-W up to 12-W are needed. See Fig. 11.

While the ability to change lamp in the same fixture is a benefit for LED replacement lamps, this works best with existing lighting systems, as these lamps don't offer the individual fixture dimming that integral module fixtures offer.

Landscape Lighting Fixtures with Integral LED Modules

Using LED lamps in landscape lighting started making a great deal of sense when fixtures could integrate LED source modules and the fixture featured dimming or “tuning” of its output separate from all other fixtures. These units need to have multiple, interchangeable optics to provide at least three beam spreads. These need to be provided with each fixture purchased, as it is not possible to determine what beam spread will be required and projects continually change. The beam spread used during the initial installation may need to be changed 1 year later or multiple years later. Getting an accessory for a product multiple years later is difficult.

It is also critically important that the LED module be replaceable. This allows for changing the output and for upgrading over time when either the module stops working or a new innovation has been released to the field.



Fig. 11 This large Beech Tree was planted nearly 200 years ago. It towers over the walkway between the house and the garage. For trees of this stature, the wattage of an LED replacement lamp needs to be in the 7–12.5 W range. Lighting Design, Jan Moyer Design (Photograph by George Gruel)

The individual fixture dimming is the feature that has captured the attention of designers for landscape lighting. For the first time, designers are able to set the right amount of light for each fixture. One fixture may be located near the trunk, requiring a very low amount of light to visually tie the canopy to the ground. Four or more fixtures may be located around the trees' canopy, and the location of branches and leaves requires varying amounts of light and differing beam spreads. See Fig. 12. This allows the designer to utilize one fixture and set the beam spread and output as needed in the field. Since landscapes continually change – trees grow, for example – a big element of landscape lighting is continually checking the lighting composition and making changes to the beam spread, light output, fixture aiming, glare shield setting, and even the fixture location.

Several landscape lighting fixture manufacturers offer integrated fixture dimming. How it is offered varies. One approach uses a tool inserted at a spot on the fixtures to dial the level or a magnet to dim the light level and/or set three predetermined levels. See Fig. 13.

Another manufacturer utilizes an in-line connection on the cable of each fixture fitted with the electronics to be dimmed via a temporarily wired control unit. This allows easy dimmability without need of being at the fixture – especially helpful for fixtures mounted in trees which can be 30 f. above the ground or for fixtures in difficult-to-reach locations. See Fig. 14.



Fig. 12 The two larger trees in the mid-ground of this photo are magnolia trees approximately 100 years old and the smaller tree in the foreground was planted within the last year. This variation of size and form is typical in landscape lighting. As trees grow or change in other ways, the lighting is adjusted. Fixture location or aiming may be changed. Sometimes the lamp output or beam spread needs to be changed. Previous to fixtures using LED sources with tuning capability this meant either the owner, contractor, or designer needed to stock multiple lamps which could be thousands of dollars in inventory. Tunable fixtures allow the designer to easily change the output in the field whenever necessary. Lighting Design by Design Teams B, H, and i of the International Landscape Lighting Institutes' (illi) first Landscape Lighting Exhibition. See the illi photos as <http://illionline.org/llc> or <http://illionline.org/llc/leipad>

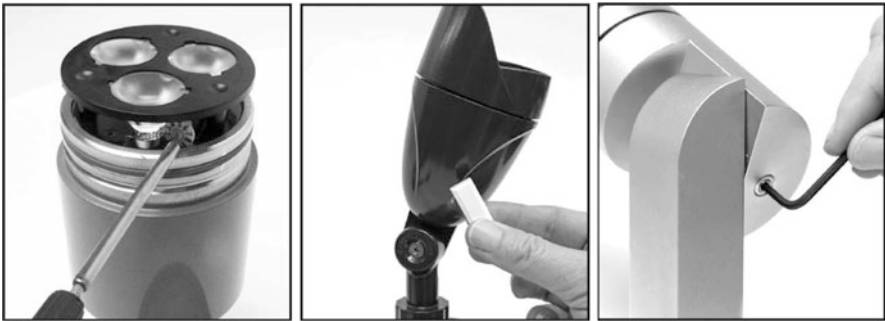


Fig. 13 Shows two fixtures using a tool to dial the dimmed level and the middle photo shows using a magnet. Manufacturers are BK Lighting, *left*, EcoLink, *middle*, and, HK Lighting Group *right* (Photos by George Gruel)



Fig. 14 LED “tunable” fixtures are located behind this patio furniture and under the hedge. The lighting accentuates the branching structure of this 100 year old hedge and provides some general light with no perceived glare for the sitting area. They are not easily accessible and they don’t need frequent adjustment. This means the Cast Lighting tuning system provides a good control option. Lighting Design by Design Team i of the International Landscape Lighting Institutes’ (illi) first Landscape Lighting Exhibition. See the illi photos as <http://illionline.org/ille> or <http://illionline.org/illeipad>

Conclusion

Landscape lighting provides safety and security in all instances, but it offers the possibility of revealing the beauty of our land, our architecture, and, our art. It makes most sense in the winter months or in areas that experience short days throughout the year. It removes the dark mask of night and the fears that go with it. Lighting designers need all the players in the electronic lighting community to communicate and work together to invent and develop, then market products that will be sensitive to our environment, not create unnecessary glare, have a realistic long life, continue to be available, and be modular so that products can be updated as technology evolves. Landscape lighting equipment needs to last the lifetime of a project and be responsive to landscape and technological changes.

References

- Lennox Moyer J (2013a) Light sources, Chap 6. In: *The landscape lighting book*, 3rd edn. Wiley, Hoboken, pp 87–104. ISBN 978-1-118-07382-7 (ebook is available)
- Lennox Moyer J (2013b) Luminous composition, Chap 3. In: *The landscape lighting book*, 3rd edn. Wiley, Hoboken, pp 19–32. ISBN 978-1-118-07382-7 (ebook is available)

-
- Lennox Moyer J (2013c) Light fixtures, Chap 7. In: The landscape lighting book, 3rd edn. Wiley, Hoboken, pp 105–136. ISBN 978-1-118-07382-7 (ebook is available)
- Lennox Moyer J (2013d) Garden evolution – changes that affect lighting, Chap 15. In: The landscape lighting book, 3rd edn. Wiley, Hoboken, pp 251–263. ISBN #978-1-118-07382-7 (ebook is available)
- Lennox Moyer J (2013e) Wiring, Chap 10. In: The landscape lighting book, 3rd edn, Wiley, Hoboken, pp 169–191. ISBN #978-1-118-07382-7 (ebook is available)

Part VI

Human Factors and Performance

Human Vision and Perception

Mahalakshmi Ramamurthy and Vasudevan Lakshminarayanan

Contents

Anatomy of the Visual System	758
Physiology of Information Processing	764
Photoreceptor Density and Spectral Sensitivity	764
Receptive Field Organization	766
Some Aspects of Visual Perception	768
Detection and Discrimination Thresholds	768
Information Integration	772
Color Psychophysics	776
Summary	780
References	780

Abstract

This chapter covers some basic aspects of the fundamentals of the human visual perception. Given the enormity of the subject, we will only give a very limited account of this huge area of intellectual activity. We will first discuss the structure of the visual system that is necessary to understand the functional processing of visual information and its limitations. Next, we will elaborate on some classic visual psychophysical results that delineate the limits of detection and

M. Ramamurthy (✉)

Department of Psychology, Developmental and Brain Sciences, University of Massachusetts,
Boston, MA, USA

e-mail: zz.maha@gmail.com

V. Lakshminarayanan

School of Optometry and Vision Science and Departments of Physics and Electrical and Computer
Engineering, University of Waterloo, Waterloo, ON, Canada

Department of Physics and Department of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, MI, USA

e-mail: vengu@uwaterloo.ca

discrimination. With these basics, we will introduce the reader to three important streams of information processing pertinent to lighting and display technologies, namely, spatial vision, flicker fusion, and color vision. We do not discuss many aspects of visual perception such as binocular visual perception, shape and form recognition, face recognition, etc. due to space constraints. Good overviews and detailed discussions of various aspects of visual perception can be found in a number of books, e.g., Cornsweet (1970), Schwartz (2010), Palmer (1999), Norton et al. (2002), and Werner and Chalupa (2013), and Chalupa and Werner (2003) or in the chapter by Lakshminarayanan and Raghuram (2003).

Anatomy of the Visual System

The enormous amount of visual information in the world is contained in the light reaching the eye. What we perceive and how we perform depend on the information that is processed by each successive structure of the visual system. In this section, we will traverse the anatomy of the visual system as the light stimuli would and will elaborate its structural and functional aspects. For more details, please see Oyster (1999).

The eye is embedded in a protective framework of bones and connective tissues called the orbit. At birth, the average axial length (distance between cornea and retina) of the eye is about 16.8 mm and grows to an average length of 24.5–25 mm by adulthood and remains fairly constant thereafter (Gordon and Donzis 1985).

The first part of the visual system consists of structures that remain transparent to enable maximum, undisturbed transmission to light. These structures, namely, the cornea, pupil, lens, and the humors (aqueous and vitreous) of the eye, collectively determine the optics of the eye. The quality of the image formed on the retina (which is the neuronal layer) greatly depends on the optical properties of the eye. A good book on the optics of the eye is by Atchison and Smith (2000). The optics of the eye is analogous to that of a camera in terms of its image-forming properties. This analogy holds good until we reach the neuronal layer, the retina, where information encoding gets complex and interesting. A schematic cross section of the eye is given in Fig. 1.

The Cornea: This is the first refractive surface that the light encounters. Anatomically there are five corneal layers of which the thickest is the stromal layer. The stromal layer is made up of collagen fibrils neatly arranged in the form of a lattice, which facilitates the transparency of the layer. The refractive index of cornea is 1.376. Cornea is aspherical and has variable thickness across center and surround, center being the thinnest (~0.56 mm) among other regions (Liu et al. 1999). The anterior surface has a radius of curvature of approximately 7.8 mm, while the posterior surface is about 6.53 mm (Edmund 1994). The asphericity of the anterior and posterior corneal surfaces is independent of the radius of curvature at the vertex (Muller et al. 2001). The cornea contributes to two thirds of the dioptric power of the eye (roughly about 40 diopters of power). The cornea is nearly transparent to visible spectrum (Fig. 2), absorbing less than 10 % of the incident light at 800 nm and less

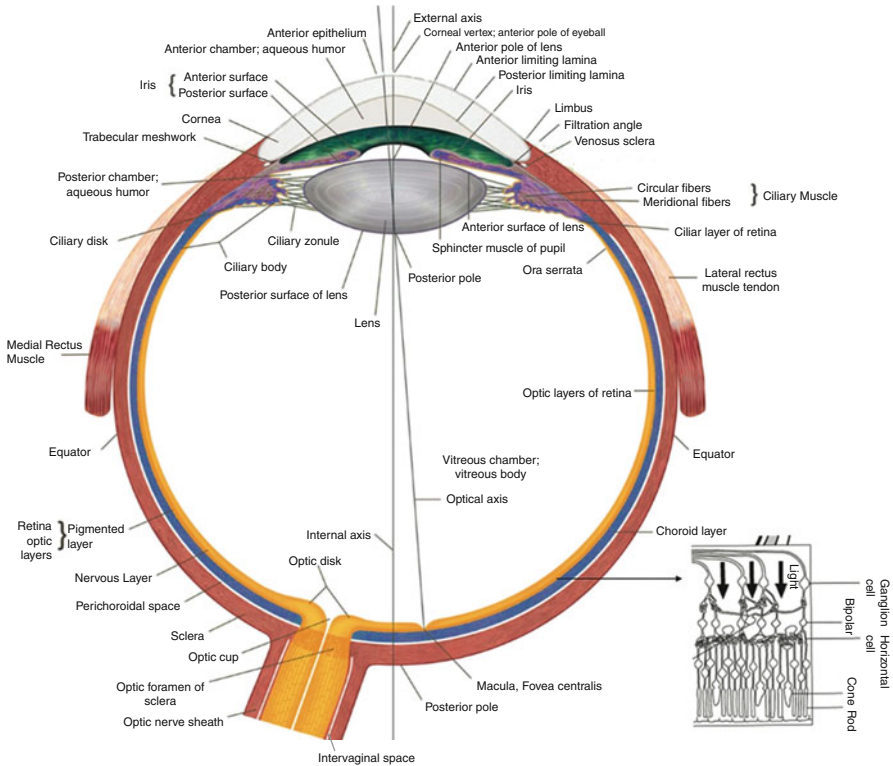


Fig. 1 Shows the cross section of the human eye. The figure on the *right corner* shows the direction of light and the arrangement of the retinal layers that the light first meets

than 20 % of the incident light at 400 nm. For wavelengths less than 300 nm, the corneal absorption increases to more than 99 %, protecting the lens from the short wavelength exposure (Boettner and Wolter 1962; Van Best et al. 1988).

Aqueous and Vitreous Humors: The distance between the posterior corneal surface and the iris is called the anterior chamber of the eye and is filled with an optically clear medium called the aqueous humor. The pressure exerted by the aqueous humor helps in maintaining the structural integrity of the eye and is called the intraocular pressure (IOP). The normal IOP ranges between 12 and 16 mm of Hg. The space between the lens and retina is the posterior chamber and is filled with a colorless gelatinous medium called the vitreous humor. Both aqueous and the vitreous humors absorb less than 10 % of the incident light at all wavelengths between 400 and 800 nm (Fig. 2).

Lens: The crystalline lens is an avascular, transparent elliptic structure, suspended from the surrounding ciliary body by zonular fibers. The contraction of the ciliary muscles can thus cause a change in the lens curvature. This process by which the lens changes curvature to focus objects at different distances on to the retina is called *accommodation*. It should be noted that the ability to focus at near distances as one

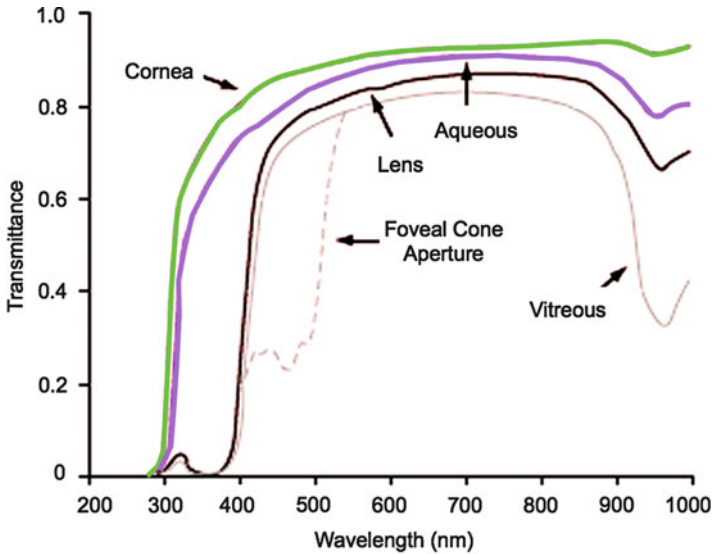


Fig. 2 The transmittance of the ocular media based on measurements on freshly enucleated eyes. The data are replotted from Boettner and Wolter (1962). Each curve is the transmittance at the rear surface of the labeled structure

ages decreases leading to the condition called presbyopia. The lens contributes to about one third (about 20 D) of the total dioptric power of the eye. The lens has a gradient refractive index (GRIN) distribution, the refractive index being a factor of protein concentration within the lens fibers, and it increases from anterior to center and decreases from the center to posterior. The pigments of the lens absorb short wavelengths so strongly that often lens absorption is substituted for the total absorption of the ocular media at visible ranges. In young adults, lens absorption is less than 10 % between 450 and 900 nm. Absorption has been measured in excised lenses, by comparing the spectral sensitivities of the eye with and without lenses and by taking the ratio of the intensities of the Purkinje images originating from the front and back surfaces of the lens (Van Norren and Vos 1974). Unlike other ocular filters, lens transmission changes with age. The human lens contains short wavelength absorbing pigment that decreases in concentration over the first 5 years of development (Cooper and Robson 1969; Werner 1982; Werner et al. 1990). After 30 years of age, there is an increase in scattering within the lens that reduces the amount of transmission at all wavelengths. Lens thickness and pigment density increase with age, making absorption at short wavelengths stronger. Between the second and sixth decade of life, the lens density for 400 nm increases, on an average by 0.12 log units per decade. After the sixth decade, the density increase accelerates to 0.4 log units per decade on an average (Pokorny et al. 1987). Please see the chapter on “► [Lighting and the Elderly.](#)”

Refractive anomalies:

The furthest and the closest points that we can see clearly are the near and far points of the eye. The difference between the near and far point forms the accommodation range of the eye. An eye with its far point at infinity is said to be normal and is termed an emmetropic eye. When the far point is not at infinity, the eye is said to be anomalous and is termed ametropic. Hyperopia (far sight) and myopia (near sight) typically show shifts in the far point; in the former, the far point shifts behind the eye, and in the latter, the far point is at some finite distance in front of the eye. The origin of refractive anomalies can either be the cornea or the lens or in the axial length (Borish 1954).

Pupil: Light passes through an aperture in the pigmented iris, called the pupil, to reach the crystalline lens. The smooth muscles of the iris can alter the size of the pupil, thereby altering the amount of light entering the eye. The size of the pupil is adjusted based on the pupillary light reflex. The reflex is controlled by retinal illumination that is signaled by retinal ganglion cells that project to the pretectum (Edinger-Westphal nuclei) in the midbrain (Rodieck et al. 1993, 1998). The midbrain, in turn, sends signal to the ciliary ganglion that innervates the iris sphincter muscles that causes the pupil to constrict when illumination is high and to dilate when illumination is low.

The pupil size affects image quality. The optics of the eye gives rise to aberrations; therefore, some of the incident (parallel) rays fail to converge perfectly resulting in a blurred image. The retinal image of a single point of light on the retina is called the point spread function (PSF) of the eye's optics. The point spread function provides a complete description of image quality at the retinal location for a given wavelength of light. A constricted pupil reduces aberration of the incoming light and improves the quality of image on the retina, but at the cost of increasing diffraction, that distorts the image (see Fig. 3). Considering all factors, a size of 2–3 mm produces optimum image quality (see Atchison and Smith 2000 or Wandell 1995).

The second part of the visual system consists of the neural components that make up the visual pathway. The light that reaches the retina is subjected to the limitations posed by the optics of the human eye and is now ready to traverse the neuronal layers.

Retina: The innermost layer of the eye is the neuronal layer, retina. The central retina is called *macula* and encompasses a high-resolution cone-rich zone called the foveola. Anatomically the retina has ten layers. We will briefly examine the ten layers of the retina and their functions. Broadly, the retina is divided into outer retina that is proximal to choroid (the second layer of the eye) and derives nutrients from the vascular-rich choroid, and the inner retina (that is fuelled by the central retinal vessels) is composed of layers proximal to vitreous.

Retinal Pigment Epithelium (RPE) is the outermost layer of the retina and is not responsive to radiant energy and therefore plays no role in encoding visual information. The darkly pigmented RPE absorbs photons that escape absorption by the photoreceptive layer and thus helps in reducing light scatter within the eye.

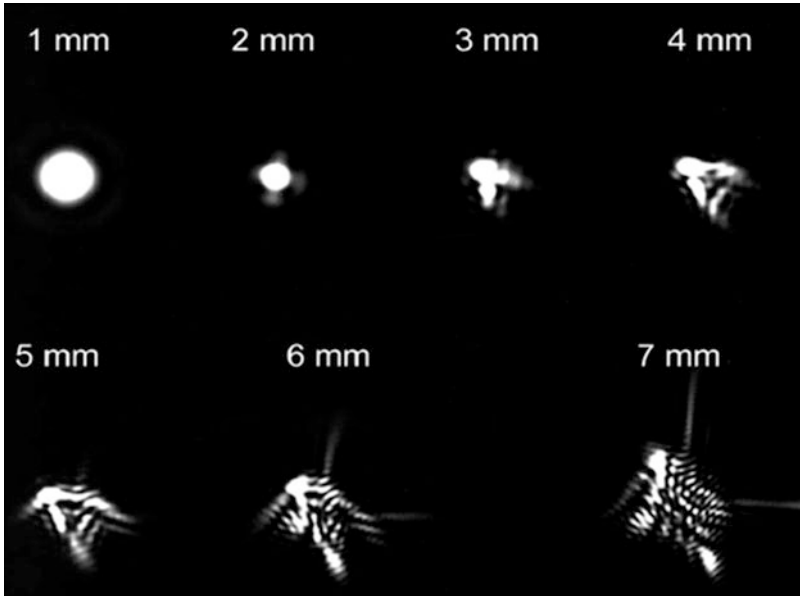


Fig. 3 The effect of pupil size on PSF for a typical human eye. At small pupil diameters, diffraction dominates, while at large sizes, aberrations contribute more to retinal blur (Courtesy Dr. Austin Roorda, Roorda and Williams (1999))

RPE plays a crucial role in providing metabolic support and in phagocytizing the continuously shed photoreceptor outer segment (Young 1971).

Photoreceptors are cells that manifest very high metabolic rates and therefore are located close to choroidal blood supply. There are two kinds of photoreceptors: **rods**, which are the foundation for night vision, and **cones**, which are the basis for daytime vision. The outer segments of the photoreceptors contain photo pigments that absorb photons that, in turn, initiate series of photochemical reactions (see Lakshminarayanan 2005 for discussions of photon absorption and transduction; see also Rodieck 1998). The outcome of such a reaction is phototransduction. Based on the absorption spectra of photopigments, the cones are categorized into short wavelength absorbing cones (S), long wavelength absorbing cones (L), and middle wavelength absorbing cones (M).

Outer limiting membrane is formed by interconnecting process of Müller cells (glial cells) that separate the photoreceptor outer segment from the inner segment. The inner segment consists of the photoreceptor nuclei that form the *outer nuclear layer*.

The photoreceptor cells synapse with the bipolar cell layer and horizontal cells to form the *outer plexiform layer*. The cell bodies of the bipolar, horizontal, and amacrine cells form the *inner nuclear layer*.

Ganglion cell layer consists of more than 20 different types of cells of which 80 % are accounted by three major classes of neurons: the smaller, more populated

(that make up ~80 % of the population of ganglion cells) midget cells; the larger, less populated (that make up ~20 % of the population of ganglion cells) parasol cells; and the less studied small bistratified ganglion cells. The axons of the ganglionic cells form the *nerve fiber layer*, exiting the eye through the optic disk to form the optic nerve (2nd of the 12 cranial nerves) that synapses at the midbrain. The optic disk has no photoreceptors and therefore forms a physiological blind spot in each eye (also called an absolute scotoma). This is why driving instructors advise drivers to turn and look for vehicles before changing lanes to avoid missing vehicles that might fall on the blind spot. This is differentiated from what is called relative scotomas, which are regions of the retina where visual sensitivity is lowered.

The *internal limiting membrane* forms an interface between the nerve fiber layer and the vitreous with strong adhesion at the vitreous base and at the optic nerve head.

Optically, what is of concern is the retinal vasculature that lies between the cornea and the photoreceptors. The human retina has a very small avascular zone, ranging from 120 to 600 μm diameter around the fovea. With increasing eccentricity from the edge of the avascular zone, the capillary coverage increases to about 40 %. The light incident on the photoreceptor is filtered through a uniform layer of blood 2 μm in thickness. Given the spectral properties of blood that absorbs light between 400 and 450 nm, the vasculature absorbs short wavelength more strongly. The shape of the absorption band depends on the degree of oxygenated hemoglobin in the blood. The invisibility of retinal vasculature in normal viewing is presumably due to retinal gain changes, cortical adaptation, and failure to represent retinal regions beneath denser vessels (Adams and Horton 2002).

With regard to information processing, these ten layers can be compressed into three main neuronal layers: the photoreceptor layer \rightarrow bipolar cell layer \rightarrow ganglion cell layer. This arrangement reflects a feed forward or centripetal organization (as shown in Fig. 1). The axons that leave the retina decussate at the optic chiasm, so the information from each nasal hemiretina is sent to the contralateral hemisphere. The retinal ganglion cells project to three main subcortical regions: the pretectum, the superior colliculus, and the lateral geniculate nucleus (LGN). The LGN is the principle relay station that sends 90 % of the information to the visual cortex. The parasol and midget ganglion cells project, respectively, to the magnocellular and parvocellular divisions of the LGN (Perry et al. 1984, 1985; Rodieck and Stone 1965, 1979); the small bistratified cells project to the parvocellular layer (Rodieck 1991). The LGN is organized such that adjacent regions in visual space stimulate neighboring neurons. This property is called retinotopic organization. Information from the LGN is carried through optical radiations to the primary visual cortex (V1). Information from V1 transcends to deeper layers (higher visual cortical areas V2, V3, V3a, V4, V5, V6). Information processing up to the primary visual cortex is conventionally termed as lower level processing, whereas beyond V1, processing is more complex and in general is termed higher-level processing (mainly because it involves higher-order cognitive mechanisms). There are several computational approaches to model the visual cortex, and the goal of which is to capture mechanisms that explain how visual information is represented at every stage of the visual pathway and ultimately the visual cortex. To understand how a physical stimulus is

processed by the visual system, it is critical to understand the physiological capacities and constraints of information processing at each stage, starting from photoreceptor density, spectral sensitivity, receptive field organization, and pathways that carry information.

Physiology of Information Processing

Photoreceptor Density and Spectral Sensitivity

After traversing the layers of the retina, the light reaches the photoreceptors. The spatial distribution of rods and cones is highly nonuniform (Curcio et al. 1990). Figure 4a shows the density of photoreceptors as a function of retinal eccentricity. A rather striking feature of photoreceptor topography is the radial symmetry with

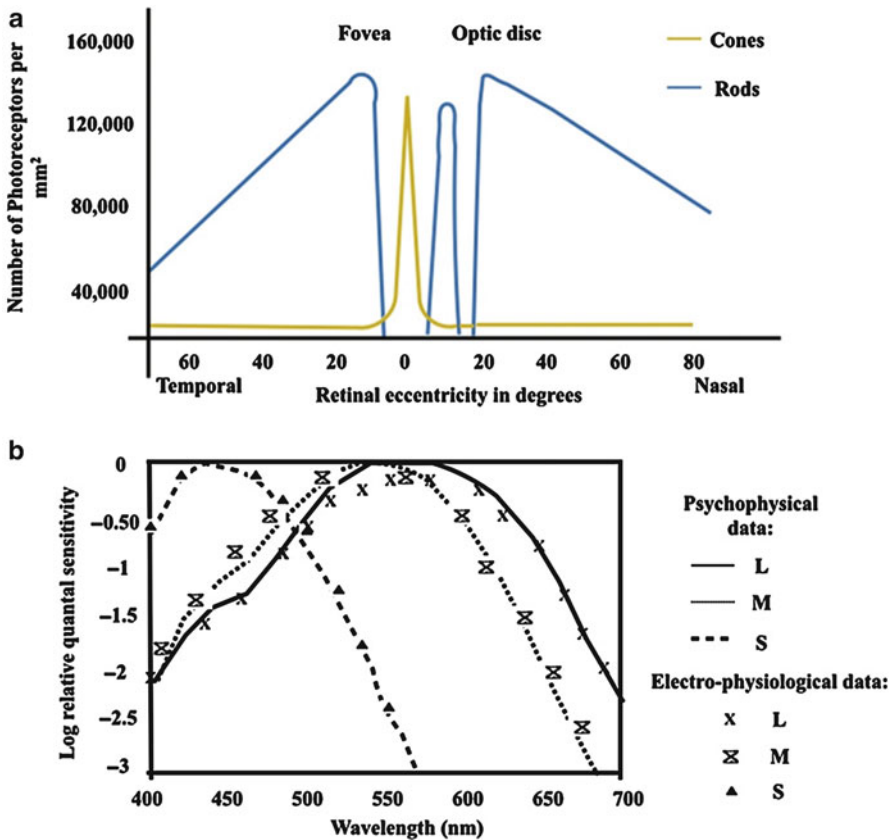


Fig. 4 (a) Density of photoreceptors as a function of retinal eccentricity (Adapted from Osterberg 1935); (b) spectral properties of the eye (Psychophysical data is from Smith and Pokorny (1975); electro-physiological data is from Schnapf et al. (1987))

which rods and cones are distributed. Notice that there are far more rods than cones. This might seem counterintuitive as cones function at high luminance levels and provide high resolution compared to rods that operate at low luminance levels. It should be recalled that cones are concentrated at the center and exhibit one-to-one connection, while rods at the periphery tend to pool information by showing many-to-one connection with the preceding layer (which is the bipolar cell layer). The quality of spatial, temporal, and color vision at high light levels is subserved by 4–5 million cones that are distributed unevenly across the retina. The densities and retinal distributions are different for different cone types. The S cone submosaic forms a rather regular lattice comprising less than 2 % of cones in the central fovea and somewhat less than 10 % of cones elsewhere (Marc and Sperling 1977; Monasterio et al. 1985; Ahnelt et al. 1990). S cones show different retinal distribution than other cones (Calkins 2001). Not only are they considerably less numerous than L or M cones (constituting only less than 10 % of the cone population), they are also not found in the central 0.3–0.4° of the human fovea (Williams et al. 1981; Curcio et al. 1990; Roorda et al. 2001; Carroll et al. 2004). It is difficult to discriminate the L and M cone submosaics, as there is no known morphological difference between the two. Psychophysical evidence suggests that the ratio of L to M cone is 2:1 (Nerger and Cicerone 1992; Cicerone 1990), but the available physiological data in monkeys suggest a ratio close to 1:1 (Bowmaker 1991; Schnapf et al. 1990). In humans, psychophysical tasks involving high temporal frequencies (that specifically target luminance channels) provide a ratio that is significantly different from unity with large interindividual variability. In contrast, tasks involving low temporal frequencies (targets red-green chromatic channels) provide L/M cone ratios that are close to unity. Measurements from retinal densitometry (uses second generation adaptive optics) correlate more with psychophysical data obtained using high temporal frequencies (Kremers et al. 2000). First retinal images of the three cone types from living human retina were captured by adaptive optics combined retinal densitometry. These images showed distinct patches with interleaved cone classes in a single mosaic, such that at each point on the retina, a single cone class samples the retinal image (Roorda and Williams 1998). Developments of retinal imaging systems with adaptive optics (AO) has made it possible to image the living human retina at a microscopic scale, something that was possible only with excised retinas. Adaptive optics provides supernormal acuity to normal vision, by correcting for eye's aberrations. Contrast sensitivity of fine spatial patterns was increased when observers viewed the patterns through adaptive optics (Liang et al. 1997). With supernormal resolution, AO has helped understand the sampling performance of the cones, by imaging the packing densities and patterns across the retina. In a study by Li and Roorda (2007), they found that the cones located at greater eccentricities are more randomly arranged and protect our visual system from perceiving aliased signals (cone density also decreases with higher eccentricity replacing aliasing with noise). Aliasing does not affect central retina, due to higher cone densities and hexagonal close packing, that extends beyond frequencies that can even pass through the optics of the eye (Lakshminarayanan and Nygaard 1992).

Spectral sensitivities: The spectral sensitivity of a photoreceptor is determined by the ability of its photopigments to absorb photons of different wavelengths and have been isolated using electrophysiological, microspectrophotometry, and psychophysical techniques. Electrophysiological data from single primate cones directly measure the stimulus intensity (for different wavelengths) required to elicit a criterion response. Thus the cone itself is being used as a univariant photon counter. The clearest data show that each of the cones contain one of three photopigments with peak sensitivity at 430, 530, 560 nm, respectively (Baylor et al. 1987). Spectrophotometric studies measure the spectral sensitivity of the three cone groups in situ, using microspectrophotometry (MacNichol 1986) or by measuring the light reflected back out of the eye (Rushton 1965).

Usually stimulus intensity is measured at the cornea so the spectral sensitivities of the receptors are confounded by the absorption properties of the optics of the eye. Despite the difficulties, spectrophotometric methods also showed the existence of three distinct populations of cones. Psychophysical measures of spectral sensitivity depend either on the fact that the visual system adapts (sensitivity reduces as ambient light intensity increases), or on individuals who genetically lack a cone pigment type. Adaptation studies typically use background light of wavelengths that affect one spectral channel more than others; that channel would be differentially desensitized, thus facilitating measurement of the spectrum of the remaining channel(s) (there are many variants to this approach; see, Stiles 1978; Pugh and Kirk 1986; Stockman et al. 1993; Hood 1998). The alternate approach is to use individuals, who genetically lack one or more cone types. Since the central retina of normal individuals are populated with L and M cones, individuals who lack either are assumed to have only a single cone type in the central retina. Figure 4b shows data from psychophysical and electrophysiological studies. The human psychophysical spectral data are taken from Smith and Pokorny (1975), for which stimulus intensity was measured at the cornea and therefore they include the effects of preretinal absorption. To compare these curves with the superimposed points from isolated cones (Schnapf et al. 1987), the cone spectra were corrected for preretinal absorption. The nice fit between the two sets of data is at least partly due to this correction for preretinal factors.

Receptive Field Organization

The output of retinal ganglion cells consists of a series of discrete electrical impulses called action potentials or spikes. Consider that we have microelectrodes placed in close proximity to the retinal ganglion cells such that it records electrical activity from one cell (extracellular single-unit recording). A light stimulus is directed onto various areas of the screen until an area is found which maximally changes the activity of the neuron we are recording. This area is called the receptive field of a neuron. Although this is referred to as receptive field of a neuron, it should be remembered that it depends on the properties of the entire visual pathway, starting from the optics of the eye. The receptive fields of the ganglion cells are

Cell type	Photoreceptors	Horizontal cells	Bipolar	Amacrine cells	Ganglion cells
Receptive field					
Response to center 					
Response to Surround 					

Fig. 5 Shows a schematic of center-surround response in ganglionic cells

approximately circular with a center region and an annular antagonistic surround. In on-center ganglion cells, light to the center increases response, whereas light on the surround reduces response; the opposite occurs in off-center ganglion cells (Fig. 5). This is often referred to as lateral inhibition (Kuffler 1953; Barlow et al. 1957). Lateral inhibition plays a major role in our perception. On increasing the area of a spot of light, the activity changes, reaches a maximum, and then drops. On further increasing the area, it no longer modulates activity of the neuron and is now beyond its receptive field. Figure 5a shows the receptive field shape at every stage of the retina (intracellular recordings from the retina of mudpuppy) and the corresponding spike response to a spot of light at center versus surround (adapted from Werblin and Dowling 1969). The receptive field configuration of ganglion cells reflects the collective properties of the neurons that precede them. The on-center midget bipolar cells synapse with the on-center midget ganglion cells, and off-center midget bipolar cells synapse with the off-center midget bipolar cells. Midget ganglion cells form the parvocellular pathway. Similarly, the on- and off-centers of the diffuse bipolar cells synapse with that of the diffuse parasol cells that make up the magnocellular pathway.

The receptive field properties of the LGN neurons (both parvocellular P and magnocellular M cells) are similar to their ganglion cells inputs and display similar center/surround configuration, spatial, temporal, and chromatic response properties. Parasol and midget ganglion cells project respectively to the magnocellular and parvocellular divisions of the LGN (Perry et al. 1984); small bistratified cells project to the parvocellular layers (Rodieck 1991). The internal organization of the LGN is locally complex with afferent fibers from ganglion cells making contact with relay

neurons that convey signals to cortex and with interneurons. The receptive field properties of the neurons in the primary visual cortex are substantially different from those in the retina or LGN; V1 neurons have elongated receptive fields that display considerable selectivity to size, orientation, direction of motion, and binocular disparity. Thus, the probability that a cortical neuron will be activated by an arbitrary retinal image is much lower than for retinal and LGN neurons. Essentially all V1 cells display nonlinear response characteristics but can be divided into two classes based on their nonlinear behavior. Simple cells are quite sensitive to spatial phase or position within the receptive field; complex cells are relatively insensitive to position within the receptive field (Hubel and Wiesel 1968).

Some Aspects of Visual Perception

The previous section outlined the anatomy/physiology of the visual system. Much of visual perception is predicated on the components of the visual system, its architecture and physiology that limit or govern the performance on different tasks. In some tasks, the outcome is limited by the optics, while in others performance is limited by computations at the level of visual cortex (see Marr 2010 or Frisby and Stone 2010). A major goal in the study of human vision is to relate performance to the underlying anatomical and physiological constraints. It is very interesting as to how the architecture of the visual system and its constraints are optimized to function in a certain way that best results in what we see and how well we see what we see. For example, what are the implications of having a center-surround type of receptive field organization? Or how does the three cone system (with their peak sensitivities) aid in the perception of colors that span the visible spectrum? By the time light reaches the retina, there is loss of information due to scattering and absorption at each structure that forms a part of the optics of the eye. Once the light reaches retina, it is further hampered due to aliasing/sampling effects and background noise (internal or sensory noise) throughout the retino-cortical pathway. Given all the constraints, how does the visual system compute information that ultimately results in recognition (De Valois et al. 1990) or decision-making? In this section, we will discuss some fundamentals aspects of visual perception and psychophysical experiments that help us frame the perceptual limits of the visual system. Detailed aspects of psychophysics methods can be found in the books by Geisheider (2013) or Lu and Doshier (2013).

Detection and Discrimination Thresholds

Absolute Sensitivity in Vision

The eye is extremely sensitive to light. But, how sensitive is the eye to light? The absolute sensitivity of the eye cannot be gauged by a single threshold value, since the minimum amount of light necessary for vision has been found to be dependent on the conditions of stimulation and on the quantum nature of the light source. Therefore, the absolute sensitivity of the visual system is best understood by looking into the

relationship between the absolute threshold and the conditions that determine its value. Exposing the eye to intense light desensitizes the eye. Sensitivity is recovered gradually by staying in dark for at least 1 h. The dark adaptation curve is traced by periodically measuring the observer’s absolute threshold during recovery period, as a function of time in the dark. In the classic experiments by Hecht et al. (1937), a test stimulus was presented to a region of the retina that has both rods and cones, and a biphasic dark adaptation curve was obtained (Fig. 6a). The first phase shows a rapid reduction in threshold with time and saturates after 5–8 min of dark exposure. The second phase that starts after 10 min is relatively gradual and takes a long time to saturate (about 40 min). The biphasic curve is caused by the intersection of these two curves that start at different intensities, have different slopes, and approach different asymptotes. The rod cone break is the point when the rod sensitivity actually exceeds cone sensitivity. Although the visible spectrum spans between 400 and 700 nm, the human eye is not sensitive to all wavelengths equally. The spectral sensitivity curves showing the absolute threshold as a function of stimulus wavelength have been

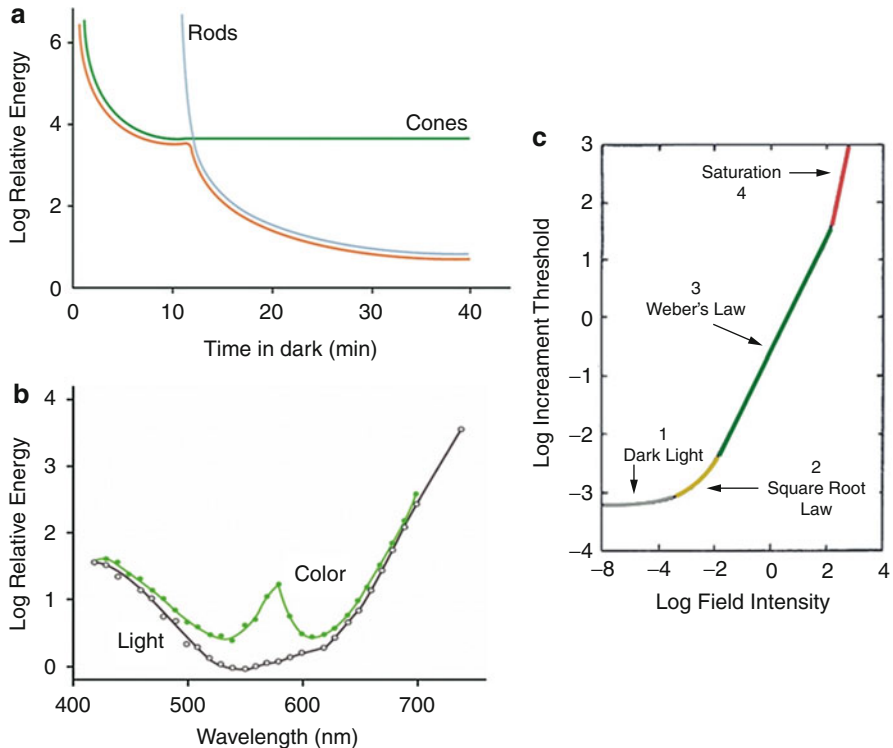


Fig. 6 (a) Shows the dark adaptation thresholds for both rods and cones for a 420 nm stimulus as function of time to dark adapt (Adapted from Schwartz (2010)); (b) shows the photochromatic interval as a function of wavelength (Adapted from Graham and Hsia (1969)); (c) threshold-versus-intensity function (TVI) showing regions of noise, square root law, Weber’s law, and saturation (Data adapted from Aguilar and Stiles; from Lakshminarayanan (2012a))

obtained for cones (and are called *photopic spectral sensitivity*) and rods (*scotopic spectral sensitivity*). Wald (1945) measured the absolute thresholds for 22 observers for detecting a 1° 40-msec test stimulus presented at different wavelengths either within the fovea or 8° above the fovea. Figure 6b illustrates that light at extreme wavelengths, red and blue, are relatively ineffective in producing a response. The peripheral retina is sensitive to 500 nm and the central retina is sensitive to 560 nm. The difference between rod and cone threshold is obtained by gradually increasing the intensity of light in an extrafoveal region, where both rods and cones are present. When the rod threshold is reached, the light appears but is colorless; on further increasing the intensity, a point is reached when the color appears which marks the cone threshold. This interval between the rod and cone intensities is called the *photochromatic interval*. The size of the photochromatic interval varies with wavelength and is smallest for short wavelengths and largest for long wavelengths (see Fig. 6b).

Lorentz in 1901 hypothesized that a just noticeable flash of light delivered approximately 100 photons to the cornea. It is difficult to predict the number of photons required for the perception of light from this number, due to uncertainties in the photons reaching the photoreceptors. In what became a classic experiment in vision science by Hecht et al. (1942) determined the amount of light at the retina necessary for vision under conditions yielding optimal sensitivity. To specify the number of quanta absorbed at threshold by the photochemical pigment (rhodopsin), the threshold values measured at the cornea were corrected for losses of light within the eye. After accounting for these factors, only 5–14 quanta were absorbed by rhodopsin. In the 10-min retinal area stimulated, there are about 500 rods, thus making it highly unlikely that more than one quantum will be absorbed at threshold. Thus, Hecht et al. 1942 concluded that in order to see, it is necessary that only one quantum of light be absorbed by a single photopigment molecule in each of 5–14 rods. Thus, the maximum sensitivity of the eye approaches a limit imposed by the nature of light. Given the exquisite sensitivity of the eye, you might wonder why a single photon is not sufficient for seeing. The answer to this is due to differential sensitivity of the eye. Sensory thresholds are greatly limited by the internal noise in the system. In this sense, absolute and differential thresholds are similar. In the measurement of differential sensitivity, it is the difference between two stimuli that needs to be distinguished, whereas in absolute thresholds, it is the difference between a stimulus and its internal noise that must be distinguished (Gescheider 2013).

Discrimination Thresholds

We have been discussing the detection of light at (or near) absolute threshold values. Here, we discuss the important issue of intensity discrimination – that is, how does the visual system determine the difference in intensity between two luminous stimuli? The quantum fluctuations provide a theoretical lower limit for intensity discrimination by an ideal observer. In an increment threshold measurement, a test stimulus of luminance L_1 is compared to an adjacent stimulus of luminance L_2 ,

which is a reference stimulus. The psychophysical task is to determine how different L_1 must be from L_2 for it to be seen as different (at some preassigned probability value, say 50 % of the time). Let this difference be given by ΔL . If we determine ΔL for a number of different values of the reference L , we get a curve called a threshold-versus-intensity function (TVI function). A typical TVI curve is shown in Fig. 6c. If we start from near absolute threshold, the threshold in the flat horizontal portion of the curve is determined by the internal noise (1 – dark light); increases in luminance does not change ΔL very much; this implies that observers would not be able to always detect an intensity difference between two flashed stimuli that on average differ in intensity by only a few quanta. As the background luminance is further increased, we move on to region two of the curve, the “square root” law region. Quantum fluctuations increase with number of quanta in the stimulus, and as stimulus luminance increases, the minimum discriminable threshold increases in proportion to the square root of the intensity level. This is known as the deVries-Rose law or the square root law and is expressed as

$$\frac{\Delta L}{\text{SQRT } L} = K$$

In this portion, the slope of the curve is $\frac{1}{2}$. For the rod pathway, a slope of 0.6 is often found. At low reference luminance levels, humans behave as ideal detectors and follow the deVries-Rose law.

As we further increase background levels, Weber’s law holds and the intensity discrimination threshold is higher than expected from an ideal detector. In this region, also called Weber’s law region, the slope is a constant. The constant proportional relationship between increment threshold (ΔL) and reference luminance (L) is called Weber’s law:

$$\frac{\Delta L}{L} = \text{constant}$$

This proportional change in threshold ΔL with L implies that the visual system is not detecting luminance differences at the theoretical limit. It should be noted that the Weber constant is affected by stimulus size, duration, wavelength, and retinal location (Blackwell 1946; Lynn et al. 1996; Harwarth et al. 1993; Gescheider 2013). Weber’s law implies that there is a limitation on intensity discrimination due to loss of information. At higher luminance levels, the Weber fraction becomes large – that is ΔL increases faster than L and the visual system saturates. To account for this trend, Gustav Fechner, in 1860, proposed that the relationship between the physical intensity and the perceived magnitude of a stimulus would follow a logarithmic relationship and is called Fechner’s law (and is the basis, e.g., the decibel scale used in acoustics). While Fechner’s law holds for a wide range of stimuli, it does not cover all. A more comprehensive relationship to entail all possible subjective response was given by Stevens.

Steven's power law:

$$\text{Perceived magnitude} = \text{constant} (\text{physical stimulus})^{\text{power}}$$

According to Steven's power law, the perceived magnitude is directly proportional to the physical intensity raised to some power. This power defines the relationship; if the relationship between perceived and physical stimulus is linear, then the power is unity (Stevens 1957; Gescheider 2013).

It is also important to point out that the visual system is directionally sensitive – that is, it is more sensitive to light coming in through the center of the eye pupil than to light entering from the periphery of the pupil. This is called the Stiles-Crawford Effect (Stiles and Crawford 1933) of the first kind. There is also a color effect, in that the perceived hue of a light changes depending upon its pupil entry position (the Stiles-Crawford effect of the second kind) (Stiles and Crawford 1937). These effects are due to the waveguide properties of retinal photoreceptors (for reviews, see: Lakshminarayanan and Enoch 2000/2010; Enoch and Lakshminarayanan 1991). The Stiles-Crawford effect helps in integrating and giving a summated response to light incident on the retina (Enoch and Lakshminarayanan 2009).

Information Integration

Consider a visual input displayed on a monitor; two important integration steps would be spatial and temporal. Spatial integration of information helps in pattern vision, and temporal integration of information helps in the perception of motion. In the previous section, we discussed the light sensitivity of the eye, which helps in understanding fundamental limits of perception. In this section, we will discuss contrast sensitivity function (basis of spatial vision) and critical flicker fusion frequency (critical concept in designing video displays).

Spatial Vision

Almost all we see in space should be accounted for by spatial vision. Conventionally the term is used to study how the visual system processes static two dimensional luminance patterns. The receptors are able to summate energy over space, as evident from the previous section, and the total number of quanta is constant at threshold, whether they are distributed sparsely over a large area or concentrated in a small area. This process is called spatial summation and is governed by Ricco's law ($L \times A = \text{Constant}$). Similarly, the receptors are also able to summate over time up to 0.1 s, since it has been shown that the quanta at threshold are the same when exposing the eye to a weak stimulus for a long time or a strong stimulus for a short time. This is called temporal summation and is described by Bloch's law ($L \times T = \text{Constant}$). Studies on human pattern vision can be divided into two types: studies that focus on the threshold assessment or low-level integration, where the value of given dimension of interest (say, e.g., contrast) at which the stimulus is detected or discriminated with a given probability is measured and studies where the value mentally assigned

to the dimension is assessed. Most of our understanding of the visual system is from threshold studies. Suppose we want to assess the most detectable spatial pattern, one simple approach would be to measure the sensory threshold for individual pattern, the only constraint being, infinite possibilities of pattern. To overcome this problem, all patterns can be decomposed to their fundamental components, and this is nothing but a Fourier decomposition of spatial patterns (Fourier 1822). Therefore, any pattern can be represented as sum of sine wave patterns with differential spatial frequency, amplitude, phase, and orientation. With this the visual system is conceived as a spatial filter (Campbell and Robson 1968; Baldock and Graham 2000). Our ability to perceive luminance variation across space is depicted by contrast sensitivity function (CSF). To obtain the CSF, sine wave gratings of different spatial frequency are presented on a background that has the same luminance as the average luminance of the test grating. The threshold contrast at which a given spatial frequency is detected (here it is more like discriminated from its background) is plotted as a function. The plot of this sensitivity as a function of spatial frequency is called the contrast sensitivity function.

The CSF of the optical neural system is analogical to the modulation transfer function (MTF) of an optical system. The quality of the image formed on the retina is solely obtained by the MTF (or the optical transfer function) and is delimited by the optics of the eye that precedes the retina. The neural system then operates on this retinal image to form further computations of the image to be sent to higher-order recognition systems. Unlike MTF, CSF is a band-pass filter. Typically, normal adults have maximum sensitivity between 2 and 6 cycles per degree of visual angle (Fig. 7a). There is a steep reduction in sensitivity at lower spatial frequencies and a gradual reduction at higher spatial frequency approaching unity (since the maximum contrast possible is 1, and since sensitivity is the inverse of the threshold contrast, the minimum possible sensitivity is 1) at about 60 cycles per degree. This high spatial frequency cutoff determines the resolution limit of the visual system. This is called the grating acuity. The high frequency cutoff is attributed to the optical limitations (aberrations of the eye) and the packing of cones (DeValois and DeValois 1988; Lakshminarayanan 2005). The lower frequency drop-off is attributed to lower spatial frequency stimuli covering both the center and surround of the receptive field with a uniform stimulation that results in lateral inhibition. In one view, the visual information processing system is considered as a Fourier analyzer (which deconstructs the retinal image into its spatial frequency components. This view assumes independent spatial frequency channels and not a single channel that is maximally sensitive to a single spatial frequency. On adapting the system to a given spatial frequency and on retracing the CSF after adaptation, the sensitivity to the adapted spatial frequency is reduced (Blackmore and Campbell 1969). This effect is called the notch effect and favors the notion of CSF being an envelope of multichannel responses (Fig. 7b). The contrast sensitivity is affected by various factors ranging from optical, neural, retinal, and adaptation state of the individual. The detection of these simple patterns is described invoking a set of linear filters that have various tuning properties (Wilson and Bergen 1979; Barten 1999). For detailed discussions, see Graham (2001) or DeValois and DeValois (1988).

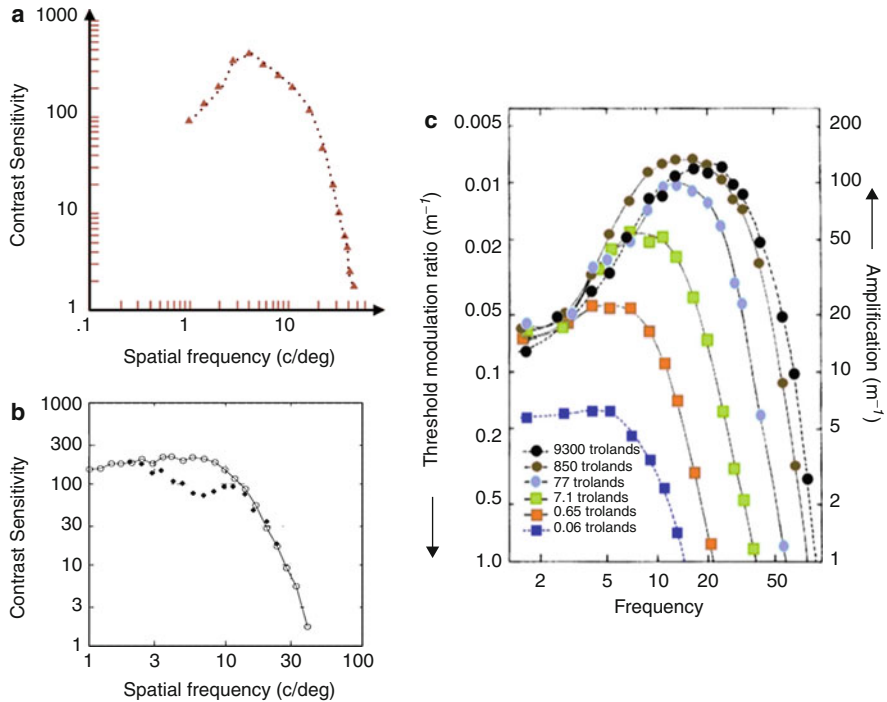


Fig. 7 (a) Shows normal human contrast sensitivity function obtained psychophysically; (b) shows the effect of pattern adaptation to a sinusoidal grating of 7.1 cpd, what is known as a notch effect (Adapted from Wandell (1996)); (c) shows the human contrast sensitivity function for several mean luminance levels (The data is adapted from Kelly (1961) and Lakshminarayanan (2012b)). The amplification scale on the right is the contrast sensitivity and the threshold modulation on the left is the threshold contrast

Temporal Vision/Critical Flicker Frequency

As is well known, images on a monitor are not continuous and are continuously refreshed. The rate at which the image is refreshed plays a crucial role in making the image appear continuous, even though the actual luminance of a point on the screen is intermittent. Because the visual system is sensitive to temporal changes, it integrates the responses with respect to time. Flicker arises when the display images are not repeated quickly enough. Flicker perception can be studied using grating stimuli whose luminance varies sinusoidally with time. Flicker perception depends upon stimulus size, luminance, retinal location, and temporal modulation among other factors. It has been found that chromaticity has little or no effect on CFF if the luminance is held constant (DeLange 1958).

If we present to an observer alternating repetitive cycles of low and high luminance of a temporal square (or sine) wave stimulus, the light will appear to flicker when the temporal frequency (in Hertz) is low. If we increase the temporal frequency, it will appear to be steady beyond a certain frequency. Psychophysically, we

define the CFF (critical flicker fusion frequency) as the frequency at which the stimulus is seen flickering 50 % of the time and as steady or fused 50 % of the time. The CFF is a measure of the temporal resolving power of the visual system. This minimum interval of resolution is analogous to the minimum angle of resolution in spatial vision (Lakshminarayanan 2012b). In this sense, temporal acuity is analogous to grating acuity in spatial vision. The neural basis of CFF is the modulation of firing rates of retinal neurons (Tyler and Hamer 1990; Lee et al. 1989).

When a light is flickering above the CFF, it will appear steady, and the time averaged luminance of a flickering light determines its brightness above the CFF. This time averaged luminance is called the Talbot brightness. The Talbot brightness can be easily calculated using

$$\text{Talbot Brightness} = L_{\min} + ([L_{\max} - L_{\min}] * f)$$

Here, L_{\max} and L_{\min} refer to the maximum and minimum luminance of the grating, and f is the fraction of time L_{\max} is present during the total period, i.e., the duty cycle.

There are also two brightness enhancement effects associated with temporal variation of light. The first one, called the Broca-Sulzer effect, occurs for suprathreshold flashes of light. It is found that if the flash duration is of the order of 50–100 ms, then they would appear brighter than stimuli of either shorter or longer durations. The second effect is called the Brucke-Bartley effect. This is another manifestation of the Broca-Sulzer effect. Here, a flickering light of about 10Hz will appear brighter than a steady light of the same average luminance (Bartley 1938).

Temporal contrast sensitivity function: A complete description of the temporal responsiveness of the human visual system is given by the temporal contrast sensitivity function (temporal CSF). Like its counterpart, the spatial CSF, the temporal CSF has a band-pass shape, with a peak, a high temporal frequency cutoff (the CFF) and a low temporal frequency roll-off (Fig. 7b). The CFF is the temporal analog of the minimum angle of resolution.

Figure 7c shows the human contrast sensitivity for different luminance levels. The peak contrast occurs at an intermediate flicker frequency. The temporal peak frequency shifts from approximately 20–5 Hz as the mean luminance decreases. The cutoff at high temporal frequency goes from 60 Hz to about 15 Hz as the luminance decreases. Because the visual system has a low CSF at low luminance, we can only see low temporal frequencies shift from medium to high contrast. Kelly (1961, 1964, 1972, 1979, 1969, 1974, 1971, 1966) in his classic experiments on flicker used a large flickering field with blurred edges to measure the temporal CSF functions (see also DeLange 1958). If a sharp-edged field is used, the visibility of low frequency flicker is enhanced. The presence of spatial detail improves the visibility of flicker at low temporal frequencies.

Methods to predict flicker in monitors:

Various empirical methods have been proposed to predict whether a particular VDT will appear to flicker in a given environment (see, e.g., Rogowitz 1988).

Many of these methods are cumbersome and time consuming. Farrell et al. (1987, 1988) has developed analytic methods for predicting whether a given VDT will flicker given screen phosphor persistence, refresh frequency, distance to VDT from the observer, etc.

Color Psychophysics

Color is defined as that characteristic of visible light by which an observer may distinguish differences between two fields of identical contours by just the differences in the spectral composition of the radiant power concerned in the observation (Wyszecki and Stiles 1982). Understanding human color perception has had vast implications on various color industries ranging from printing industry to video display industry. Detailed description of various aspects of color science can be found, for example, in the books by Malacara (2011), Kaiser and Boynton (1996), Gegenfurtner et al. (2000), and Shevell (2003). The human visual system is sensitive to electromagnetic radiation ranging from 400 to 700 nm – visible spectrum. The three types of cones clearly serve color vision. Since there is only one type of rod, the rod system is incapable of perceiving color. The sensitivity of the two systems with respect to the perceived brightness of various wavelengths is given by the spectral luminous efficiency function. For the rod system, the scotopic luminous efficiency function ($V'(\lambda)$) is identical to the spectral absorption of rhodopsin. In contrast, for the cone system, the photopic luminous efficiency function $V(\lambda)$ represents a combination of all three types of cones and is derived from psychophysical measurements of heterochromatic photometry (Wyszecki and Stiles 1982). These functions signify that under balanced intensities (as in the heterochromatic bipartite fields), the information regarding the wavelength of light is lost (*principle of univariance*) and therefore an observer with just this function will only be able to perceive relative intensity differences (as in the case of monochromats – a color deficiency defined by the presence of just one photopigment). Figure 8a shows the standard observer's photopic and scotopic luminous efficiency functions. The peak of the photopic luminous efficiency function is about 555 nm. The scotopic luminous efficiency function has a peak of 507 nm and is derived from brightness matches of stimuli viewed in low light levels, where threshold is measured under dark adaptation conditions as a function of wavelength. The V - λ function gives the transformation from radiometric to photometric units and vice versa. 683 lumens (photometric units) at 555 nm is equivalent to 1 W of radiant power at that wavelength (Palmer and Grant 2010). The cones generate sets of responses of identical shape regardless of the wavelength that excite them (Fig. 8b), meaning that wavelengths of light are discriminated by relative stimulation of the three cones. Thus, by additively mixing three wavelengths to the bipartite test field, the number of photons on either side can be adjusted and made indistinguishable to the matching field. These three wavelengths are chosen such that if the absorption of photons from the two sides of the bipartite fields is equal for one wavelength, they will not be the same for the

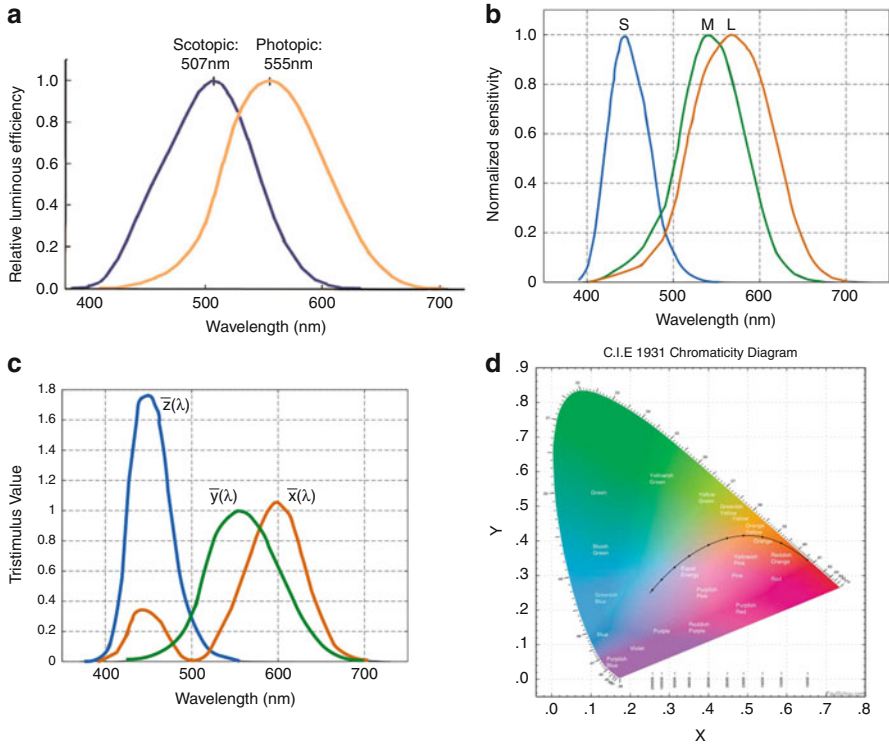


Fig. 8 (a) Photopic and scotopic luminous efficiency functions; (b) shows the three cone normalized spectral sensitivities; (c) CIE color matching functions; (d) CIE 1931 chromaticity diagram (Credits: Paulschou at en.wikipedia – <http://userpages.umbc.edu/~schou/>)

other wavelengths; such wavelengths are called *primaries* (Wyszecki and Stiles 1982; Billmeyer and Saltzman 1981)

Historically, color difference research has been based on two broad purposes. The first was to analyze color difference data in order to gain information about color discrimination mechanisms/pathways. The second purpose was to use the data as a metric to quantify perceptible color differences for industrial quality control purposes. The branch of science concerned with numerically specifying the color of physically defined stimuli such that the two stimuli that look the same (under certain conditions) have identical specifications is called colorimetry. The Commission Internationale de l’Eclairage has been developing an optimal system to best quantify color for the past 80 years. The key to understanding the CIE systems is the understanding of basic color mixing principles. Experimental laws that were derived and clarified around the beginning of the twentieth century are referred to as Grassman’s laws. Metamers are lights of dissimilar spectral radiation but are perceived as same by the observer. Metamers have three important properties that allow treatment of color mixture as a linear system (Grassmann 1853), they are:

1. Additive property: when two stimuli are metamers of each other, and when a radiation is added to each side of a color mixture field, the metamerism is unchanged.
2. Scalar property: When both sides of the color mixture field are changed in radiance by the same proportion, the metamerism is unchanged.
3. Associative property: A metameric mixture may be substituted for a light without changing the metameric property of the color fields.

According to Grassman's laws, a color match is invariant under a variety of experimental conditions that might alter the appearance of the match. Metameric matches will hold with the addition of a chromatic surround or adaptation to moderately bright chromatic field. These form the basis of color representation spaces. A fundamental property of human vision is the existence of metamers. It is possible to find a metamer for any light, by varying the energies of three fixed lights called the primaries, a formal requirement being that one primary color cannot be metameric to a mixture of other two primaries. The term *trichromat* and *trichromacy* refer to this property of human color vision (for brief review on color vision anomalies, Lakshminarayanan 2012a). Despite a century's collection of precise color matching data, there were difficulties in deducing the visual systems fundamental filtering properties. In response to this, in 1931 the Commission Internationale de l'Eclairage (CIE) settled with a specific set of spectral weighting coefficients given by $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ as shown in Fig. 8c. The use of these weighting coefficients solves the ambiguity with the selection of primaries, as the color matching functions are independent of the primaries we choose. What then defines a visual match is the tristimulus value of the stimuli, which can be derived from

$$X = k \sum_{360}^{830} E(\lambda) \bar{x}(\lambda) P(\lambda)$$

$$Y = k \sum_{360}^{830} E(\lambda) \bar{y}(\lambda) P(\lambda)$$

$$Z = k \sum_{360}^{830} E(\lambda) \bar{z}(\lambda) P(\lambda)$$

Here, $P(\lambda)$ is the spectral reflectance under an illuminant (standards defined by CIE) of relative spectral power $E(\lambda)$, and k is a normalizing factor. Tristimulus values are thus absolute values. The 1931 CIE system was derived using a stimulus field size of 2° of visual angle. In 1964, an additional set of color matching functions were derived based on 10° of visual angle and are preferred for many practical applications. This additional set of color matching functions is known as the 1964 or 10° standard observer. The 1931 CIE was particularly designed so that Y tristimulus

represents luminance. The chromaticity of the color can be calculated from the chromaticity coordinates (which are relative measures), given by

$$x = X/X + Y + Z$$

$$y = Y/X + Y + Z$$

The chromaticity diagram (Fig. 8d) reveals the characteristic horseshoe shape of the spectral locus. It is sometimes convenient to refer to the dominant wavelength (obtained by extending a line from the white point through the color stimulus to the spectral locus) and purity of color (ratio of the distance between the color stimulus from the white point to the distance between color stimuli and intersection point on the spectral locus). After the development of the color specification by the CIE, colors have been represented in terms of chromaticity coordinates and luminance in a three dimensional color space. The spaces initially represented the amount of the three primaries required to match a test light, but now the spaces are more appearance based with two orthogonal hue axes usually specifying a red-green dimension and a blue-yellow dimension and the third axis which is perpendicular to the hue axes usually specifies the luminance or black-white dimension. These newer spaces are the ones used most frequently in representing color standards in various industries (for a comprehensive list of color difference equations and its application, see Ramamurthy 2011). The reader is also referred to the chapter on “► [History of Color Metrics](#)” in this volume.

Studies on chromaticity differences can be further classified based on the experimental design. One design measures the perceptibility thresholds directly using psychophysical methods. In these experiments, the task is most often based on asking subjects to compare two stimuli of which one may differ in hue, chroma, luminance, or any combination of the three. The second design measures color discrimination as intraobserver precision in color matching (sample covariance) experiments. In such designs, colors that fall within 1–2 standard deviations of the mean match are considered to be identical to the reference color, whereas colors outside these boundaries are considered to appear different from the reference color. Wright in 1941 proposed one of the first systematic studies on color discrimination. His results showed just noticeable differences (JND) in color, mapped as unequal line segments in the CIE 1931 x, y, L space (Wright 1941). The unequal line segments indicated that the perceptible color difference thresholds were not equal distances within a color space derived from the color matching experiment and demonstrated that the CIE x, y, L was not a uniform space in terms of color appearance and discrimination. Wright’s results were overshadowed by MacAdam’s extensive data (MacAdam 1943; Wright 1941; Boynton 1979) on the precision of the color matching for a single observer. The target used was a bipartite 2° field in a dark surround. The task was to adjust the test field’s chromaticity to match that of the comparison field. Twenty-five standard colors were sampled across the CIE 1931 x, y space. Thresholds were equal to two standard deviations away from the mean

match point. On a two dimensional plane, MacAdam's data were represented as ellipses of various sizes and orientation. This finding has led to the never-ending quest of achieving a uniform color space and deriving a color difference equation such that an equal distance anywhere within the space represents an equal perceptual change in the stimulus appearance. (Further readings: Billmeyer and Saltzman 1981.)

Summary

This chapter is an overview of the human visual system and some aspects of perception. We started of the structural and functional description of every stage of the visual pathway starting from the optics of the eye to the visual cortex. Understanding the structure and functional limitations of the visual stream helps understand the physiological constraints it imposes on information processing at every stage. Psychophysical studies bridge our understanding of how these physiological constraints manifests as perceptual constraints in the context of detection and discrimination thresholds. The spine of visual perception lies in comprehending how information is integrated spatiotemporally and how color is processed. The way the visual system works helps our understanding of how images are processed and optimized for higher-order decision-making stages to quickly recognize objects and optimize performance accordingly. This finds special application in video display industry, since the core motive for research in video display technology (and also in lighting) is to develop visually appealing environments.

References

- Adams DL, Horton JC (2002) Shadows cast by retinal blood vessels mapped in primary visual cortex. *Science* 298(5593):572–576
- Ahnelt P, Keri C, Kolb H (1990) Identification of pedicles of putative blue-sensitive cones in the human retina. *J Comp Neurol* 293(1):39–53
- Atchison DA, Smith G (2000) *Optics of the human eye*. Butterworths, Boston
- Baldock R, Graham J (2000) *Image processing and analysis*. Oxford University Press, Oxford
- Barlow HB, Fitzhugh R, Kuffler SW (1957) Change of organization in the receptive fields of the cat's retina during dark adaptation. *J Physiol* 137(3):338–354
- Barten PG (1999) *Contrast sensitivity of the human eye and its effects on image quality*. SPIE Press, Bellingham
- Bartley SH (1938) Subjective brightness in relation to flash rate and the light/dark ratio. *J Exp Psychol* 23:313–319
- Baylor DA, Nunn BJ, Schnapf JL (1987) Spectral sensitivity of cones of the monkey macaca fascicularis. *J Physiol* 390:145–160
- Billmeyer FW, Saltzman M (1981) *Principles of color technology*. Wiley, New York
- Blakemore C, Campbell FW (1969) On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *J Physiol* 203(1):237–260
- Blackwell H (1946) Contrast thresholds of the human eye, *J Opt Soc Am* 36:624–632
- Boettner EA, Wolter JR (1962) Transmission of the ocular media. *Invest Ophthalmol Vis Sci* 1(6):776–783
- Borish IM (1954) *Clinical refraction*. Professional Press, Chicago

- Bowmaker J (1991) Visual pigments and colour vision in primates. In: From pigments to perception. Springer, Boston, pp 1–9
- Boynton RM (1979) Human color vision. Holt Rinehart and Winston, New York
- Calkins DJ (2001) Seeing with S cones. *Prog Retin Eye Res* 20(3):255–287
- Campbell FW, Robson JG (1968) Application of fourier analysis to the visibility of gratings. *J Physiol* 197(3):551–566
- Carroll J, Neitz M, Hofer H, Neitz J, Williams DR (2004) Functional photoreceptor loss revealed with adaptive optics: an alternate cause of color blindness. *Proc Natl Acad Sci U S A* 101(22):8461–8466. doi:10.1073/pnas.0401440101
- Chalupa LM, Werner JS (2003) The visual neurosciences, vol 2. MIT Press, Cambridge, MA
- Cicerone C (1990) Color appearance and the cone mosaic in trichromacy and dichromacy. In: Color vision deficiencies, McGraw Hill Professional, pp 1–12
- Cooper GF, Robson JG (1969) The yellow colour of the lens of man and other primates. *J Physiol* 203(2):411–417
- Cornsweet T (1970) Visual perception. Academic, New York
- Curcio CA, Sloan KR, Kalina RE, Hendrickson AE (1990) Human photoreceptor topography. *J Comp Neurol* 292(4):497–523
- De Lange Dzn H (1958) Research into the dynamic nature of the human fovea→ cortex systems with intermittent and modulated light. I. Attenuation characteristics with white and colored light. *Josa* 48(11):777–783
- de Monasterio FM, McCrane EP, Newlander JK, Schein SJ (1985) Density profile of blue-sensitive cones along the horizontal meridian of macaque retina. *Invest Ophthalmol Vis Sci* 26(3):289–302
- De Valois KK, Lakshminarayanan V, Nygaard R, Schluskel S, Sladky J (1990) Discrimination of relative spatial position. *Vision Res* 30(11):1649–1660
- DeValois RL, DeValois KK (1988) Spatial vision. Oxford University Press, New York
- Edmund C (1994) Posterior corneal curvature and its influence on corneal dioptric power. *Acta Ophthalmol* 72(6):715–720
- Enoch JM, Lakshminarayanan V (1991) Retinal fiber optics. In: Charman WN (ed) Vision and visual dysfunction: visual optics and instrumentation, vol 1. MacMillan Press, London, pp 280–309
- Enoch JM, Lakshminarayanan V (2009) Integration of the Stiles Crawford effect of the first kind. *J Mod Optics* 56:2240–2250
- Farrell J, Benson BL, Haynie CR (1987) Predicting flicker thresholds for video display terminals. *Proc SID* 28(4):449–453
- Farrell JE, Casson EJ, Haynie CR, Benson BL (1988) Designing flicker-free video display terminals. *Displays* 9(3):115–122
- Fourier J (1822) *Theorie analytique de la chaleur*, par M. Fourier. Chez Firmin Didot, père et fils
- Frisby JP, Stone JV (2010) Seeing: the computational approach to biological vision. MIT Press, Cambridge, MA
- Gegenfurtner KR, Sharpe LT, Boycott BB (2000) Color vision: from genes to perception. Cambridge University Press, Cambridge, UK
- Gescheider GA (2013) Psychophysics: the fundamentals. Psychology Press
- Gordon RA, Donzis PB (1985) Refractive development of the human eye. *Arch Ophthalmol* 103(6):785–789
- Graf V, Norren DV (1974) A blue sensitive mechanism in the pigeon retina: λ max 400 nm. *Vis Res* 14(11):1203–1209
- Grassmann H (1853) Zur theorie der farbenmischung. *Annalen der Physik* 165(5):69–84
- Graham N (2001) Visual pattern analyzers. Oxford University Press, New York
- Graham C, Hsia Y (1969) Saturation and the foveal achromatic interval. *J Opt Soc Am* 59(8):993–997
- Harwarth RS, Smith EL 3rd, DeSantis L (1993) Mechanisms mediating visual detection in static perimetry. *Invest Ophthalmol Vis Sci* 34(10):3011–3023

- Hecht S, Haig C, Chase AM (1937) The influence of light adaptation on subsequent dark adaptation of the eye. *J Gen Physiol* 20(6):831–850
- Hecht S, Shlaer S, Pirenne MH (1942) Energy, quanta, and vision. *J Gen Physiol* 25(6):819–840
- Hood D (1998) Lower-level visual processing and models of light adaptation. *Annu Rev Psychol* 49(1):503–535
- Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195(1):215–243
- Kaiser PK, Boynton RM (1996) *Human color vision*, 2nd edn. Optical Society of America, Washington, DC
- Kelly D (1961) Visual responses to time-dependent stimuli. I. amplitude sensitivity measurements. *J Opt Soc Am* 51(4):422–429
- Kelly D (1964) Sine waves and flicker fusion. *Doc Ophthalmol* 18(1):16–35
- Kelly D (1966) Frequency doubling in visual responses. *J Opt Soc Am* 56(11):1628–1632
- Kelly D (1969) Flickering patterns and lateral inhibition. *J Opt Soc Am* 59(10):1361–1368
- Kelly D (1971) Theory of flicker and transient responses, I. uniform fields. *J Opt Soc Am* 61(4):537–546
- Kelly D (1972) Flicker. In: *Visual psychophysics*. Springer, pp 273–302
- Kelly D (1974) Spatio-temporal frequency characteristics of color-vision mechanisms. *J Opt Soc Am* 64(7):983–990
- Kelly DH (1979) Motion and vision. II. Stabilized spatio-temporal threshold surface. *J Opt Soc Am* 69(10):1340–1349
- Kremers J, Scholl HP, Knau H, Berendschot TT, Usui T, Sharpe LT (2000) L/M cone ratios in human trichromats assessed by psychophysics, electroretinography, and retinal densitometry. *J Opt Soc Am A* 17(3):517–526
- Kuffler SW (1953) Discharge patterns and functional organization of mammalian retina. *J Neurophysiol* 16(1):37–68
- Lakshminarayanan V (2005) Vision and the single photon. In: “What is a photon?”, SPIE proceedings, vol 5866. pp 332–337
- Lakshminarayanan V (2012a) Light detection and sensitivity. In: *Handbook of visual display technology*. Springer, Berlin, pp 85–91
- Lakshminarayanan V (2012b) Flicker sensitivity. In: *Handbook of visual display technology*. Springer, Berlin, pp 101–108
- Lakshminarayanan V, Enoch JM (2000/2010) Biological waveguides. In: *Handbook of optics*, vol 3. Optical Society of America, Washington DC. Revised and updated chapter in *Handbook of optics*, vol 3, 3rd edn. McGraw Hill, New York, 2010
- Lakshminarayanan V, Nygaard RW (1992) Aliasing in the human visual system. *Concepts Neurosci* 3:201–212
- Lakshminarayanan V, Raghuram A (2003) Chapter 30, Visual considerations for the optical engineer. In: Wolfe W (ed) *Handbook of optical engineering*. SPIE Press, Bellingham, pp 593–659
- Lee BB, Martin PR, Valberg A (1989) Nonlinear summation of M-and L-cone inputs to phasic retinal ganglion cells of the macaque. *J Neurosci* 9(4):1433–1442
- Li KY, Roorda A (2007) Automated identification of cone photoreceptors in adaptive optics retinal images. *J Opt Soc Am A* 24(5):1358–1363
- Liang J, Williams DR, Miller DT (1997) Supernormal vision and high-resolution retinal imaging through adaptive optics. *JOSA A* 14(11):2884–2892
- Liu Z, Huang AJ, Pflugfelder SC (1999) Evaluation of corneal thickness and topography in normal eyes using the orbscan corneal topography system. *Br J Ophthalmol* 83(7):774–778
- Lu Z, Doshier B (2013) *Visual psychophysics: from laboratory to theory*. MIT Press, Cambridge, MA
- Lynn J, Feltman R, Starita R (1996) Principles of perimetry. In: Rich R, Shields MB, Krupin T (eds) *The glaucomas*. Mosby, St. Louis
- MacAdam DL (1943) Specification of small chromaticity differences. *J Opt Soc Am* 33(1):18–26

- MacNichol E Jr (1986) A unifying presentation of photopigment spectra. *Vision Res* 26 (9):1543–1556
- Malacara D (2011) *Color vision and colorimetry*. SPIE Press, Bellingham
- Marc RE, Sperling HG (1977) Chromatic organization of primate cones. *Science* 196 (4288):454–456
- Marr D (2010) *Vision*. MIT Press, Cambridge, MA
- Muller LJ, Pels E, Vrensen GF (2001) The specific architecture of the anterior stroma accounts for maintenance of corneal curvature. *Br J Ophthalmol* 85(4):437–443
- Nerger JL, Cicerone CM (1992) The ratio of L cones to M cones in the human parafoveal retina. *Vision Res* 32(5):879–888
- Norton T, Corliss D, Bailey JE (2002) *Psychophysical foundations of visual perception*. Butterworth-Heinemann, Boston
- Osterberg G (1935) Topography of the layer of rods and cones in the human retina. *Acta Ophthalmol Suppl* 6:1–103
- Oyster CS (1999) *The human eye*. Sinauer, Sunderland
- Palmer S (1999) *Vision science: photons to perception*. MIT Press, Cambridge, MA
- Palmer JM, Grant BG (2010) *The art of radiometry*. SPIE Press, Bellingham
- Perry VH, Cowey A (1985) The ganglion cell and cone distributions in the monkey's retina: implications for central magnification factors. *Vision Res* 25(12):1795–1810
- Perry V, Oehler R, Cowey A (1984) Retinal ganglion cells that project to the dorsal lateral geniculate nucleus in the macaque monkey. *Neuroscience* 12(4):1101–1123
- Pokorny J, Smith VC, Lutze M (1987) Aging of the human lens. *Appl Optics* 26(8):1437–1440
- Pugh EN Jr, Kirk DB (1986) The mechanisms of WS stiles: an historical review. *Perception* 15:705–728
- Ramamurthy M (2011) *Colour discrimination thresholds and acceptability ratings using simulated microtile displays*. MSc thesis, University of Waterloo, Waterloo
- Rodieck RW, Stone J (1965) Analysis of receptive fields of cat retinal ganglion cells. *J Neurophysiol* 28(5):833–849
- Rodieck R (1979) Visual pathways. *Annu Rev Neurosci* 2(1):193–225
- Rodieck R (1991) The density recovery profile: a method for the analysis of points in the plane applicable to retinal studies. *Vis Neurosci* 6(02):95–111
- Rodieck RW, Brening RK, Watanabe M (1993) The origin of parallel visual pathways. In: Shapley R, Lam Dm-K (eds) *Proceedings of the retinal research foundation symposium*, pp 117–144
- Rodieck R (1998) *The first steps in seeing*. Sinauer, Sunderland
- Rodieck RW, Rodieck RW (1998) *The first steps in seeing*, vol 15. Sinauer Associates, Sunderland
- Rogowitz BE (1988) The psychophysics of spatial sampling. In: 1988 Los Angeles symposium-OE/LASE'88, pp 130–138
- Roorda A, Williams DR (1999) The arrangement of the three cone classes in the living human eye. *Nature* 397(6719):520–522
- Roorda A, Williams DR (1998) Objective identification of M and L cones in the living human eye. *Investig Ophthalmol Vis Sci* 39:204
- Roorda A, Metha AB, Lennie P, Williams DR (2001) Packing arrangement of the three cone classes in primate retina. *Vis Res* 41(10):1291–1306
- Rushton WA (1965) The ferrier lecture, 1962: visual adaptation. *Proc R Soc Lond B* 162:20–46
- Schnapf JL, Kraft TW, Baylor DA (1987) Spectral sensitivity of human cone photoreceptors. *Nature* 325(6103):439–441
- Schnapf JL, Nunn BJ, Meister M, Baylor DA (1990) Visual transduction in cones of the monkey macaca fascicularis. *J Physiol* 427:681–713
- Schwartz SH (2010) *Visual perception: a clinical orientation*, 4th edn. McGraw Hill, New York
- Shevell S (2003) *The science of color*. Elsevier, New York
- Smith VC, Pokorny J (1975) Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vision Res* 15(2):161–171

- Stevens SS (1957) On the psychophysical law. *Psychol Rev* 64(3):153
- Stiles WS (1978) Mechanisms of colour vision: selected papers of WS stiles; with a new introductory essay. Academic, New York
- Stiles WS, Crawford BH (1933) The luminous efficiency of rays entering the eye pupil at different points. *Proc R Soc Lond B* 112:428–450
- Stiles WS, Crawford BH (1937) The luminous efficiency of rays entering the eye pupil at different points and a new colour effect. *Proc R Soc Lond B* 122:255–288
- Stockman A, MacLeod DI, Johnson NE (1993) Spectral sensitivities of the human cones. *J Opt Soc Am A* 10(12):2491–2521
- Tyler CW, Hamer RD (1990) Analysis of visual modulation sensitivity. IV. Validity of the ferry-porter law. *J Opt Soc Am A* 7(4):743–758
- Tyson R, Lakshminarayanan V (2012) Adaptive optics. *J Modern Opt* 59(12):1032–1033
- Van Best J, Bollemeijer J, Sterk C (1988) Corneal transmission in whole human eyes. *Exp Eye Res* 46(5):765–768
- Wald G (1945) Human vision and the spectrum. *Science* 101(2635):653–658
- Wandell BA (1995) Foundations of vision. Sinauer, Sunderland
- Wandell BA (1996) Book Rvw: foundations of vision. *J Electron Imaging* 5(1):107–107
- Werblin FS, Dowling JE (1969) Organization of the retina of the mudpuppy, *necturus macubsus*. II. intracellular recording. *J Neurophysiol* 32:339
- Werner JS (1982) Development of scotopic sensitivity and the absorption spectrum of the human ocular media. *J Opt Soc Am* 72(2):247–258
- Werner JS, Chalupa LM (2013) The new visual neurosciences. MIT Press, Cambridge, MA
- Werner JS, Peterzell DH, Scheetz A (1990) Light, vision, and aging. *Optom Vis Sci* 67(3):214–229
- Williams DR, MacLeod DI, Hayhoe MM (1981) Foveal tritanopia. *Vision Res* 21(9):1341–1356
- Wilson HR, Bergen JR (1979) A four mechanism model for threshold spatial vision. *Vision Res* 19(1):19–32
- Wright W (1941) The sensitivity of the eye to small colour differences. *Proc Phys Soc* 53(2):93
- Wyszecki G, Stiles W (1982) Color science: concepts and methods, quantitative data and formulae. Wiley, New York
- Young RW (1971) The renewal of rod and cone outer segments in the rhesus monkey. *J Cell Biol* 49(2):303–318

History of Color Metrics

Wendy Davis

Contents

Introduction	786
CIE 1931 Standard Observer	787
CIE 1964 Supplementary Standard Observer	789
Uniform Chromaticity Diagrams	789
Uniform Object Color Spaces	791
Physiologically Based Color Matching Functions	792
Color Rendering	792
Development of Color Metrics	794
Future Directions	795
References	796

Abstract

Though the physical principles and visual mechanisms underlying color have been studied for centuries, the modern colorimetric system used to quantify and specify the color properties of light sources and illumination had not even begun to be developed 100 years ago. Because color only exists when the physical properties of light and materials interact with the human visual system, any such system must essentially model the link between physics and perception. This is an enormously difficult task and the colorimetric system hasn't always been successful. Nonetheless, since the origin of colorimetry in the early twentieth century, remarkable progress has been made.

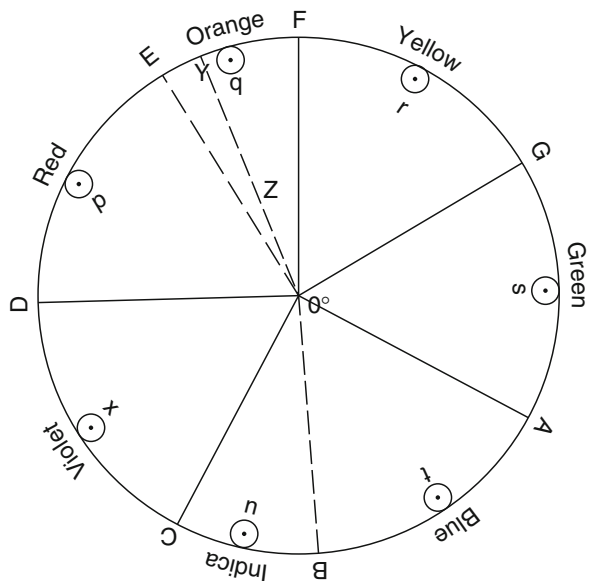
W. Davis (✉)
University of Sydney, Sydney, Australia
e-mail: wendy.davis@sydney.edu.au

Introduction

Attempts to quantify, specify, and predict color appearance arose from some of the earliest research on the physics of light and color. Isaac Newton's landmark research on the color mixing of light yielded a color circle figure composed of seven hues: red, orange, yellow, green, indigo, violet, and blue (Newton 1704). This circle, as shown in Fig. 1, illustrates the way that light of these different hues mixes to create other colors. For instance, a mixture of red and yellow light appears orange. Countless circular models of color have been proposed since – some illustrate color mixing of light sources, and others show color mixing of pigments (objects). Color models of many other shapes, including pyramids, spheres, cones, and cubes, have been developed over the past 250 years.

One of the more influential color specification systems for object colors was developed by artist Albert Munsell (1905) and is still used today. The Munsell Color Space is composed of colored samples arranged into an approximately spherical shape. The Munsell space was devised such that samples differed in perceived color by approximately equal steps. Lightness (called *value* in the Munsell system) changes along the polar (vertical) axis of the sphere, equivalent to latitude on Earth. The top of the sphere is white and the bottom is black. Hue varies between samples around the polar axis, equivalent to changes in longitude on Earth. The saturation or colorfulness of the samples (called *chroma* in the Munsell system) is maximal at the surface of the sphere and lowest at the polar lightness axis, which would correspond to the center of the Earth. Each sample is described by its hue, value, and chroma. Later research on the perceptual spacing of samples resulted in

Fig. 1 Newton's color circle
(Reproduced from Newton
(1704))



modifications to the system (Newhall et al. 1943). As this is a commercial color order system, physical samples can be purchased.

While history is rich with a fascinating variety of color specification systems (Kuehni 2003), they are of limited usefulness in lighting. Their primary limitation is that they do not define procedures for determining numerical specifications of any novel light source color.

CIE 1931 Standard Observer

Soon after the development of the spectral luminous efficiency function, $V(\lambda)$, by the International Commission on Illumination (CIE), the organization set out to build on work done by the Optical Society of America (Troland 1922) and define the functions that would underlie the modern system of colorimetry. Of primary concern was the quantification and specification of chromaticity, indicative of the color of light independent of brightness. Like perception of color, chromaticity is both a function of the physical properties of the light and the human visual system. Over 50 years prior, color matching experiments reported by Helmholtz (1866) would eventually establish the trichromatic nature of color vision – that is, that the perception of color arises from three channels in the visual system. An important consequence of trichromacy is that any light source color can be replicated by combining any three independent colors, often called *primaries*. In this context, “independent” means that none of the colors can be matched by a combination of the other two. While trichromacy ultimately limits the number of colors that humans can discriminate between, it enables the specification of light source (chromaticity) with only three numbers.

In the late 1920s, Wright (1928) and Guild (1931) independently conducted color matching experiments that would serve as the foundation of modern colorimetry. The details of their experiments varied; for instance, Wright used monochromatic primaries of 435.8, 546.1, and 700 nm, whereas Guild used red, green, and blue filtered lights. However, the general concept for both experiments can be illustrated in Fig. 2. The observers viewed a square stimulus that subtended the central 2° of their visual field. The figure was vertically bisected and an experimenter-set reference color was presented on one-half. The observer adjusted the relative intensities of the three primaries that lit the test side of the figure until it matched the reference color, as shown in the top panel of Fig. 2. Using this method, any reference color can be matched, with one caveat, which is shown in the bottom panel of Fig. 2. For some reference colors, the observers must add a small amount of one primary to the reference side of the figure to create a color match. This is referred to as *negative color mixing*.

Independently, both Wright and Guild mathematically transformed their data to a set of monochromatic primaries (700.0, 546.1, and 435.8 nm) that matched the National Physical Laboratory’s (NPL) standard white when of equal intensity (Broadbent 2004). The results were averaged and altered slightly to correct an error and smooth out irregularities. The resultant color matching functions (CMFs), \bar{r} , \bar{g} , and \bar{b} ,

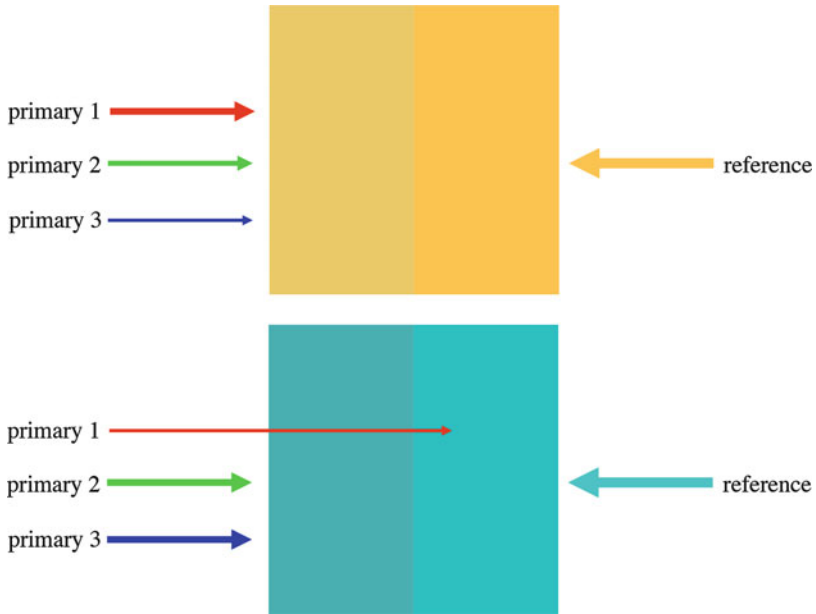
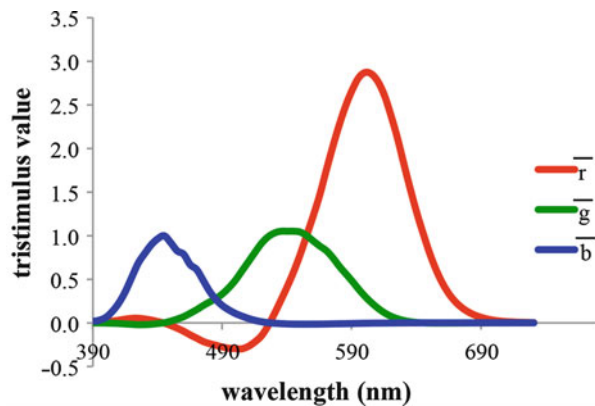


Fig. 2 General procedure for color matching experiments. *Top*: for most reference colors (*right*), observers create a matching test color (*left*) by adjusting the intensity of the three primaries. *Bottom*: for some reference colors (*right*), a small amount of one of the primaries must be added to the reference light to create a matching test color (*left*)

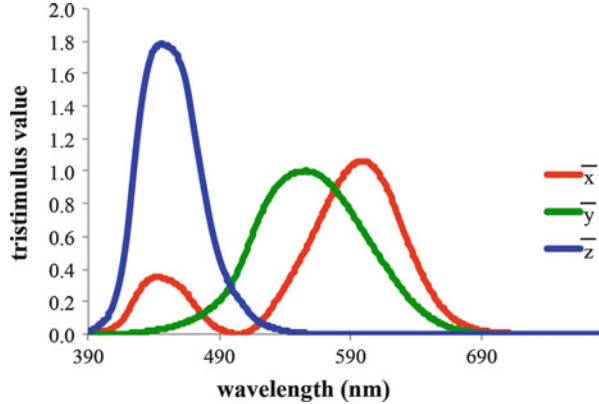
Fig. 3 CIE \bar{r} , \bar{g} , and \bar{b} color matching functions. Each curve shows the relative intensity of each primary (700.0, 546.1, and 435.8 nm) required to match the color of each wavelength of the visible spectrum



are shown in Fig. 3. The small negative peak of \bar{r} centered near 490 nm results from the negative color mixing illustrated on the bottom panel of Fig. 2.

The negative values in these CMFs were mathematically inconvenient, so they were transformed into a new set of functions, sometimes referred to as *imaginary color matching functions*. In addition to eliminating negative values, this transformation accomplished two other goals: one of the functions would be identical to

Fig. 4 CIE 1931 \bar{x} , \bar{y} , and \bar{z} color matching functions



$V(\lambda)$, and chromaticity coordinates for an equal-energy white-light source would be $\frac{1}{3}$. These CMFs, \bar{x} , \bar{y} , and \bar{z} , are shown in Fig. 4. These functions, also known as the *Standard Observer*, were adopted by the CIE in 1931 at its meeting in Cambridge, England (Fairman et al. 1997).

Subsequent research by Judd (1951) and Vos (1978) revealed that $V(\lambda)$ underestimates sensitivity for the short wavelengths of the visible spectrum. As \bar{y} was set to match $V(\lambda)$, this error also affects the CMFs. Though the CIE has published a modified spectral luminous efficiency function, $V_M(\lambda)$ (International Commission on Illumination 1988), it is not widely used. The original 1931 Standard Observer is still primarily used to calculate tristimulus values.

CIE 1964 Supplementary Standard Observer

The CIE Standard Observer is only applicable to centrally fixated 2° visual stimuli. Perception of chromaticity varies for larger and/or peripherally viewed stimuli due to the distribution of photoreceptors and macular pigment across the retina. Stiles and Burch (1959) conducted new color matching experiments with stimuli that subtended 10° of visual angle, instructing their observers to ignore the center 2° region of the stimuli. As was done in 1931, the experimentally derived CMFs were transformed into \bar{x}_{10} , \bar{y}_{10} , and \bar{z}_{10} , shown in Fig. 5. These CMFs are also known as the *Supplementary Standard Observer*. They are used for certain industrial color applications and are recommended for stimuli larger than 4° of visual angle, but are not nearly as widely used at the Standard Observer.

Uniform Chromaticity Diagrams

The Standard Observer and Supplementary Standard Observer were never expected to yield chromaticity diagrams that are *perceptually uniform*, which means that distances across a color space correspond to magnitudes of perceptual color

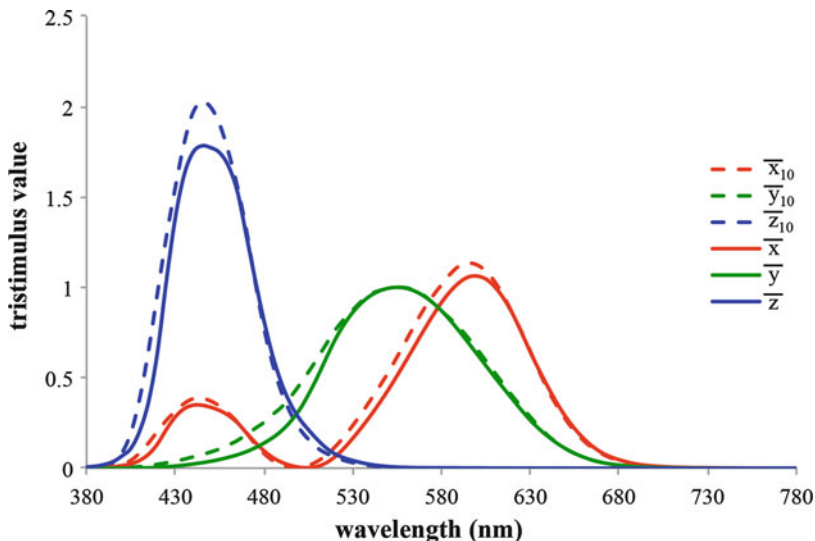


Fig. 5 CIE 1964 \bar{x}_{10} , \bar{y}_{10} , and \bar{z}_{10} color matching functions compared to CIE 1931 \bar{x} , \bar{y} , and \bar{z} color matching functions

differences. MacAdam (1942) systematically measured color discriminability across the CIE 1931 (x, y) chromaticity diagram. An observer adjusted a test light to match 25 separate reference points on the chromaticity diagram. The observer made adjustments to the test light from multiple directions in the chromaticity diagram relative to the reference color (e.g., above, below, left, right, etc.). In total, 25,000 color matches were made. This allowed MacAdam to draw a region around the chromaticity of each reference point that was indiscriminable in color (for isoluminant stimuli). If the CIE 1931 (x, y) chromaticity diagram were perceptually uniform, all of these regions would be circular and identical in size. This was not the case – the regions were elliptical and greatly varied in size. The ellipses were smallest in the violet portion of the diagram and largest in the green portion. This means that humans are sensitive to relatively small changes in CIE 1931 (x, y) chromaticity for the former colors and relatively insensitive to chromaticity differences for the latter colors. This was indicative of the nature of the chromaticity diagram, not the visual system’s ability to discriminate different colors.

The CIE has made a few attempts to develop perceptually uniform chromaticity diagrams, with coordinates that can be simply calculated from CIE XYZ tristimulus values or CIE 1931 (x, y) chromaticity coordinates. The first was the CIE 1960 (u, v) chromaticity diagram (International Commission on Illumination 1960). This diagram was indeed more uniform than its predecessor, but was not perfect. Another attempt at improvement was made with the development of the CIE 1976 (u', v') chromaticity diagram (International Commission on Illumination 1975). Again, it was more uniform than the CIE 1931 (x, y) chromaticity diagram, but still had some nonuniformities. In fact, for the nominally white chromaticities, the CIE 1960 (u, v)

system is considered to be more uniform, and it is still used for the calculation of correlated color temperature (CCT) (International Commission on Illumination 2004b). In both CIE 1960 (u, v) and CIE 1976 (u', v'), the Euclidian distance between two points is calculated to quantify color difference.

Uniform Object Color Spaces

As the CIE 1931 (x, y), CIE 1960 (u, v), and CIE 1976 (u', v') chromaticity diagrams do not consider lightness (i.e., do not include colors such as brown or gray), they are not appropriate for the specification of object color. Object colorimetry must consider both the object's illuminant and reflectance factor, as well as include three dimensions. Adams (1942) and Nickerson and Stultz (1944) developed color difference formulae that calculated color differences relative to the Munsell system, which has perceptually equal spacing between samples. The value (lightness) scale of the Munsell system was applied to the CIE XYZ tristimulus values.

The first uniform object color space recommended by the CIE was CIE 1964 (U^*, V^*, W^*). This color space was proposed by Wyszecki (1963) and was built off of the CIE 1960 (u, v) system. Lightness was indicated by W^* , while the combination of U^* and V^* specified hue and chroma. The Euclidian distance between two points represented color difference. This color space was not widely used for color difference measurement, and research revealed significant perceptual nonuniformities across the space (Robertson 1990).

Just over a decade later, the CIE (1975) recommended two uniform object color spaces: CIE 1976 (L^*, a^*, b^*) and CIE 1976 (L^*, u^*, v^*). In both of these systems, L^* indicates object color lightness, and the combination of the other two coordinates specifies hue and chroma. Formulae exist to calculate hue angle, chroma, color difference (Euclidian distance), etc. for both spaces. CIE 1976 (L^*, a^*, b^*) was based on a simplification of the Adams-Nickerson color difference formula, while CIE 1976 (L^*, u^*, v^*) was based on an improvement of the CIE 1964 (U^*, V^*, W^*) color space. CIE 1976 (L^*, a^*, b^*) has emerged as more popular over the years, though nonuniformities have been noted (Kuehni 1976; McLaren 1980).

Attempts have been made to improve the color difference measurements of CIE 1976 (L^*, a^*, b^*). Guided by experimental data of color difference detection, ΔE_{94}^* applied weightings to the lightness, chroma, and hue components (International Commission on Illumination 1995). Since that did not fully fix the nonuniformities, ΔE_{00}^* applied five additional corrections (Luo et al. 2001). CIE ΔE_{00}^* is significantly more mathematically complex than other color difference formulae and is not widely used for lighting. Some fields, such as color imaging and industrial color measurement, use this method, however.

The development of simple, universally perceptually uniform color spaces is not possible, because color perception is influenced by a number of factors that vary over time and across situations, such as the adaptation state of the observer, color and luminance of the stimulus' surroundings, etc. Therefore, color appearance models (CAMs) have become a popular area of research and development in recent years.

The latest CAM from the CIE, CIECAM02 (International Commission on Illumination 2004a), specifies mathematical procedures for making detailed predictions of the color appearance of reflective samples under well-described viewing conditions. The inputs to CIECAM02 include the luminance of the adapting field, adopted white point, luminance of the background, surround type, and degree of adaptation. Outputs of this model include lightness, brightness, redness-greenness, yellowness-blueness, colorfulness, chroma, saturation, hue composition, and hue angle. Uniform object color spaces have been developed based on CIECAM02 (Luo et al. 2006). Because these models and their associated color spaces are quite specific with regard to situations and viewing conditions, they are of limited usefulness for lighting applications.

Physiologically Based Color Matching Functions

In 1979, Boynton (1996) suggested that the CIE develop a set of CMFs that are based on the spectral sensitivity functions, often called *cone fundamentals*, of the three types of cone photoreceptors in the human retina. Over 25 years later, the CIE published 10° cone fundamentals that were mathematically derived from the CIE 1964 Supplementary Standard Observer (International Commission on Illumination 2006). From these data, 2° cone fundamentals were also reconstructed. Mathematical procedures were described for correcting for macular pigment transmission optical density, photopigment effective transmission optical density, ocular media optical density, and stimulus field size. These physiologically based CMFs can be customized to the age of the observer, as well as any specific stimulus size in the 2–10° range. The resulting CMFs can also be used to calculate chromaticity coordinates (l, s) in terms of relative cone responses.

There is quite a lot of enthusiasm for the relatively new physiologically based CMFs. However, history has shown strong resistance to changes of the fundamental functions that underlie photometry and colorimetry, even when new versions are clearly better. It remains to be seen whether these new CMFs will gain widespread use beyond research circles.

Color Rendering

Whereas chromaticity describes the color of a light source when directly viewed, color rendering describes the impact of the spectral composition of light source on the color of illuminated objects. Measurement of this property was not tackled immediately by the CIE because when modern colorimetry was being initially developed, incandescent lighting technology was dominant. The spectral power distribution (SPD) of these lamps is consistent and continuous and changes smoothly as a function of wavelength. Problems with the appearance of illuminated object colors were not noted. In the middle of the twentieth century, however, fluorescent lamps were becoming increasingly used. The spiky SPDs of these lamps impacted

object colors, usually in undesirable ways. Initial thoughts on the issue supposed that the relative smoothness or spikiness of the SPD determined color rendering performance (Bouma 1937). Therefore, the first color rendering metric recommended by the CIE employed a spectral-band method of assessment. The SPD of the light source was divided into eight sections, and the relative power in each band was compared to the relative power of a full radiator reference lamp in the same band (International Commission on Illumination 1948). This method penalized light sources with SPDs that exhibited abrupt changes in spectral power across the visible spectrum.

Subsequent research revealed that a smooth, full SPD was not required for good rendering of object colors and that color rendering could be evaluated by mathematically simulating the appearance of the reflective samples when illuminated by a test lamp and when illuminated by a known good color rendering illuminant (Nickerson 1958; Ouweltjes 1960). The CIE (1965) recommended a test-sample method for color rendering evaluation. In this method, the appearance of eight Munsell samples when illuminated by the lamp of interest is compared to the appearance of the same samples when illuminated by a reference illuminant, using the CIE 1964 (U^* , V^* , W^*) color difference calculation. The reference illuminant was matched to the CCT of the test lamp and was either a Planckian radiator (if $<5,000$ K) or a daylight simulator (if $\geq 5,000$ K). Since the SPDs of incandescent lamps approximate those of Planckian radiators, this technology was basically set as the “gold standard” for color rendering for chromaticities commonly used to electrically illuminate indoor environments. The color differences of the eight samples were averaged and scaled so that a warm-white fluorescent lamp (of the time) had a General Color Rendering Index (CRI) of 50.

Nine years later, this method was revised slightly (International Commission on Illumination 1974). Calculations were added to account for chromatic adaptation, which lessened the impact of chromaticity differences between the test lamp and reference illuminant. Six additional samples were added; they are not used in the calculation of the General CRI, but Special Color Rendering Indices can be considered. Another version of the recommendation was later been published (International Commission on Illumination 1994), but it did not include any changes to the calculation procedure.

Colorimetry has advanced markedly since 1974 (e.g., the replacement of CIE 1964 (U^* , V^* , W^*) with CIE 1976 (L^* , a^* , b^*) and CIE 1976 (L^* , u^* , v^*)), and numerous inadequacies with the CRI have been identified (e.g., Seim 1985), so multiple attempts have been made to revise the CRI or develop an entirely new metric for color rendering evaluation. None of them have been successful, though. One relatively recent attempt involved a technical committee, CIE TC 1–33 *Color Rendering*, which was established in 1991 and closed without issuing a recommendation in 1999. Updates proposed within this committee included the use of CIE 1976 (L^* , a^* , b^*) to measure differences in the color appearance of the reflective samples and the replacement of the eight Munsell samples with eight samples from the Macbeth Color Checker Chart. Ultimately, the committee was unable to reach an agreement and issue a recommendation (International Commission on Illumination 1999).

The CIE has acknowledged the problems with the CRI, particularly when it is used to assess light-emitting diodes (LEDs) (International Commission on Illumination 2007), and another technical committee, CIE TC 1–69 *Color Rendition by White Light Sources*, was created in late 2006. The members of the committee conducted research and proposed numerous new metrics (e.g., Davis and Ohno 2010; Smet et al. 2010; Rea and Freyssinier 2010), but were also unable to reach an agreement and issue a recommendation. One of the points of contention surrounded the very definition of color rendering and whether it encompasses multiple separate, meaningful, and useful concepts, such as color fidelity and color preference, that might be measured separately. When it became clear that this committee would not be successful, the CIE opened two more technical committees in 2013, TC 1–90 *Color Fidelity Index* and TC 1–91 *New Methods for Evaluating the Color Quality of White-Light Sources*.

Development of Color Metrics

All of the color metrics described in this chapter required at least some element of research during their development. Oftentimes, the desire to develop a new measurement method motivates the research. Other times, interesting experimental data serve as inspiration for a metric that is developed later. Users of color metrics should be mindful that some popular metrics arose from quite little research. For instance, MacAdam ellipses have been widely used in recent years to quantify color variability between LED products. Though MacAdam (1942) collected a lot of data, he did so with only one observer. While fundamental visual processes, such as color discrimination, are not highly variable between young observers with normal color vision, the sample size of one was very small by any standard.

As the current state of color rendering measurement illustrates, research is necessary but not sufficient for the development of a new color metric. Most successful color metrics are ultimately recommended or standardized by a standard development organization, technical body, or professional society through a consensus decision-making process. Typically, all expert stakeholders are welcome to participate in the process. For new test and measurement standards in photometry and colorimetry, committees typically consist of a combination of academics, employees of various levels of government, employees of manufacturers of lighting products, lighting designers, and/or members of relevant activist organizations (e.g., representatives from the International Dark-Sky Association are usually involved in the development of test and measurement standards for outdoor lighting products). Rules for membership on committees vary between organizations. Full members of the Illuminating Engineering Society (IES) participate on committees either by invitation or by volunteering. The CIE is composed of National Committees, rather than members of individual people. Participants in CIE activities become members of the National Committee where they live.

The definitions of consensus also vary between organizations. Until 2014, the CIE required unanimous agreement among members of a technical committee to

publish a technical report. This undoubtedly accounts for at least some of the CIE's inability to update or replace its color rendering metric over the past 20 years. Due to a change in procedures in 2014, technical reports may now be published with the agreement of two-thirds of the committee members. Most organizations require majority or (more commonly) supermajority agreement to issue recommendations or create standards. Most organizations also require the majority to formally respond to all dissenting opinions and make a strong technical case for disregarding the position of the dissenter. Even after the committee directly involved in creating a standard or recommendation reach an agreement, many organizations require voting at a higher level (e.g., member countries) before a report is published.

Though organizations like the IES and CIE often have paid staff performing administrative functions, volunteer experts perform nearly all of the work involved in developing color metrics. This also slows progress – by the time that most people rise to the point in their career when they are considered an expert in technical matters, they often have demanding jobs and struggle to find the time to commit to committee work.

Future Directions

Changes in lighting technologies often necessitate new or revised color metrics. The advent of fluorescent lamps motivated the development of the original Color Rendering Index (CRI). Now, new lighting technologies, particularly LEDs, are requiring changes to it. While color rendering is one of the more fraught colorimetric quantities, it is also critically important. Evidence suggests that poor color rendering was at least partially responsible for the slow adoption of compact fluorescent lamps (CFLs) (Sandahl et al. 2006). Manufacturers must be able to accurately evaluate and predict the color quality of new lighting technologies if widespread commercialization is to be successful.

Amidst the frustration surrounding the issue of color rendering measurement, the need for changes presents a tremendous opportunity – to rethink what is even meant by “color rendering.” Most researchers in the field recognize the issues surrounding the selection of reflective samples, color difference calculations, etc., even though there is disagreement about the best remedy for these problems. However, comparison of object color appearance to appearance under a reference illuminant is potentially just as limiting, if not more so, as the other problems of the CRI. For most light sources, the incandescent lamp is considered to have the best possible SPD for color rendering, even though studies have demonstrated that objects illuminated by incandescent lamps aren't necessarily judged by people to appear most natural or preferred (Jost-Boissard et al. 2009; Narendran and Deng 2002). Technology is able to produce SPDs that are unlike anything the lighting industry has used before. Though care must be taken, this can be leveraged to improve light quality as well as increase luminous efficacy. However, metrics that encourage or even pressure manufacturers to develop newer technologies that

behave just like the older technologies will surely stifle innovation. Lighting exists for people – not for the sake of colorimetry. The lighting industry is in the unique position to step back and reconsider lighting quality.

References

- Adams EQ (1942) X-Z-planes in the 1931 I.C.I. system of colorimetry. *J Opt Soc Am* 32:168–173
- Bouma PJ (1937) Colour reproduction in the use of different sources of ‘white’ light. *Philips Tech Rev* 2:1–7
- Boynton RM (1996) History and current status of a physiologically based system of photometry and colorimetry. *J Opt Soc Am A* 13(8):1609–1621
- Broadbent AD (2004) A critical review of the development of the CIE1931 RGB color-matching functions. *Color Res Appl* 29(4):267–272
- Davis W, Ohno Y (2010) Color quality scale. *Opt Eng* 49(3):033602
- Fairman HS, Brill MH, Hemmendinger H (1997) How the CIE 1931 color-matching functions were derived from Wright-Guild data. *Color Res Appl* 22(1):11–23
- Guild J (1931) The colorimetric properties of the spectrum. *Phil Trans R Soc A* 230:149–187
- von Helmholtz H (1866) Concerning the perceptions in general. In: *Treatise on physiological optics*, 3rd edn, vol III (trans: Southall JPC (1925) *Opt Soc Am*. Section 26, reprinted Dover, New York, 1962)
- International Commission on Illumination (1948) *Compte rendu 11th session*. CIE Central Bureau, Paris
- International Commission on Illumination (1960) In: *Proceedings of the CIE session 1959 in Brussels*. Publication 004, CIE Central Bureau, Paris
- International Commission on Illumination (1965) *Method of measuring and specifying colour rendering properties of light sources*. Publication 13, CIE Central Bureau, Paris
- International Commission on Illumination (1974) *Method of measuring and specifying colour rendering properties of light sources*. Publication 13.2, CIE Central Bureau, Paris
- International Commission on Illumination (1975) *Progress report of CIE TC – 1.3 colorimetry*. Publication 36, CIE Central Bureau, Paris
- International Commission on Illumination (1988) *Spectral luminous efficiency functions based upon brightness matching for monochromatic point sources, 2° and 10° fields*. Publication 75, CIE Central Bureau, Vienna
- International Commission on Illumination (1994) *Method of measuring and specifying colour rendering properties of light sources*. Publication 13.3, CIE Central Bureau, Vienna
- International Commission on Illumination (1995) *Industrial color difference evaluation*. Publication 116:1995, CIE Central Bureau, Vienna
- International Commission on Illumination (1999) *Colour rendering (TC 1–33 closing remarks)*. Publication 135/2, CIE Central Bureau, Vienna
- International Commission on Illumination (2004a) *A colour appearance model for colour management systems: CIECAM02*. Publication 159:2004, CIE Central Bureau, Vienna
- International Commission on Illumination (2004b) *Colorimetry, 3rd edn*. Publication 15:2004, CIE Central Bureau, Vienna
- International Commission on Illumination (2006) *Fundamental chromaticity diagram with physiological axes – part 1*. Publication 170–1:2006, CIE Central Bureau, Vienna
- International Commission on Illumination (2007) *Colour rendering of white LED light sources*. Publication 177:2007, CIE Central Bureau, Vienna
- Jost-Boissard S, Fontoynt M, Blanc-Gonnet J (2009) Perceived lighting quality of LED sources for the presentation of fruit and vegetables. *J Mod Optics* 56(13):1420–1432
- Judd DB (1951) Report of US secretariat committee on colorimetry and artificial daylight. In: *Proceedings of the twelfth session of the CIE, Stockholm, vol 1, p 11*

- Kuehni RG (1976) Color-tolerance data and the tentative CIE 1976 $L^*a^*b^*$ formula. *J Opt Soc Am* 66:497–500
- Kuehni RG (2003) *Color space and its divisions: color order from antiquity to the present*. Wiley, New York
- Luo MR, Cui G, Rigg B (2001) The development of the CIE 2000 color-difference formula: CIEDE2000. *Color Res Appl* 26:340–350
- Luo MR, Cui GH, Li CJ et al (2006) Uniform colour spaces based on CIECAM02 colour appearance model. *Color Res Appl* 31:320–330
- MacAdam DL (1942) Visual sensitivities to color differences in daylight. *J Opt Soc Am* 32(5):247–273
- McLaren K (1980) CIELAB Hue-Angle Anomalies at Low Tristimulus Ratios. *Color Research & Application* 5(3):139–143
- Munsell AH (1905) *A color notation*. Ellis, Boston
- Narendran N, Deng L (2002) Color rendering properties of LED light sources. In: *Solid state lighting II: proceedings of SPIE, Seattle, WA, vol 4776*. pp 61–67
- Newhall SM, Nickerson D, Judd DB (1943) Final report of the OSA subcommittee on the spacing of the munsell colors. *J Opt Soc Am* 33(7):385–411
- Newton I (1704) *Opticks: or, a treatise of the reflections, refractions, inflections, and colors of light*. London
- Nickerson D (1958) Measurement and specification of color rendition properties of light sources. *Illum Eng* 53:77–90
- Nickerson D, Stultz KF (1944) Color tolerance specification. *J Opt Soc Am* 34:550–570
- Ouweltjes WJ (1960) The specification of colour rendering properties of fluorescent lamps. *Die Farbe* 9:207–246
- Rea MS, Freyssinier JP (2010) Color rendering: beyond pride and prejudice. *Color Res Appl* 35(6):401–409
- Robertson AR (1990) Historical development of CIE recommended color difference equations. *Color Res Appl* 15(3):167–170
- Sandahl LJ, Gilbride TL, Ledbetter MR et al (2006) Compact fluorescent lighting in America: lessons learned on the way to market. Pacific Northwest National Laboratory, Richland
- Seim T (1985) In search of an improved method for assessing the colour rendering properties of light sources. *Light Res Tech* 17(1):12–22
- Smet KA, Ryckaert WR, Pointer MR et al (2010) Memory colours and colour quality evaluation of conventional and solid-state lamps. *Opt Express* 18(25):26229–26244
- Stiles WS, Burch JM (1959) NPL colour-matching investigation: final report (1958). *J Mod Optics* 6(1):1–26
- Troland LT (1922) Report of committee on colorimetry for 1920–21. *J Opt Soc Am* 6(6):527–591
- Vos JJ (1978) Colorimetric and photometric properties of a 2 fundamental observer. *Color Res Appl* 3(3):125–128
- Wright WD (1928) A re-determination of the trichromatic coefficients of the spectral colours. *Trans Opt Soc* 30(4):141–164
- Wyszecki G (1963) Proposal for a new color-difference formula. *J Opt Soc Am* 53:1318

Color Rendering Metrics: Status, Methods, and Future Development

A. Žukauskas and Michael S. Shur

Contents

Introduction	800
Color Rendering Index	801
Early Work Beyond Color Fidelity	803
Criticism and Refinement of the Color Rendering Index Metric	806
Color Rendition Metrics with High Numbers of Test Color Samples	813
Color Rendition Engineering	816
Summary	821
References	825

Abstract

Color rendition metrics, which assess light sources in terms of the color quality of illuminated objects, are advancing with the development of lighting technology and with the increasing needs of lighting users. This chapter reviews different metrics of assessing the color quality of light sources. We show that the traditional measures of the color fidelity, such as the standard color rendering index (CRI) and its single-figure-of-merit refinements, fail to correctly assess the color rendition properties of illumination, especially for the light sources having spectral power distributions composed of narrow-band components, such as polychromatic light-emitting diode clusters. These metrics (based on the estimation of color shifts for a small number of test color samples) do not account for the ability of the light sources to increase or decrease the chromatic contrast (color saturating or dulling) and clash with the subjective preferences to the color quality of

A. Žukauskas (✉)

Institute of Applied Research, Vilnius University, Vilnius, Lithuania
e-mail: arturas.zukauskas@ff.vu.lt

M.S. Shur

Department of Electrical, Computer, and System Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA

illumination. Supplementing these conventional measures with additional figures of merit accounting for the gamut area of a small number of the test color samples does not completely address this issue. The color rendition vector approach to the color shifts allows for a much more comprehensive assessment of the color rendition properties. In particular, many issues of the color rendition problem can be resolved using the statistical approach based on the color rendition vector sorting for a large number of the test color samples. However despite the availability of advanced measures of color rendition for experts, the need for an improved color rendition metric that could substitute for the outdated CRI still exists. An alternative approach to rating the light sources in terms of color quality is the color rendition engineering. The color rendition engineering allows for the development of light sources having requested or even tunable (and traded off) color rendition properties (color rendition engines). Such color rendition engines can meet individual and group needs in color quality of illumination. When supplemented with the additional functionalities offered by information and communication technology, the color rendition engines could become the preferred tools of the smart lighting revolution.

Introduction

The quality of how colors of illuminated objects are rendered has been a permanent concern of lighting scientists, engineers, manufacturers, and consumers ever since the wide spread of efficient sources of white light having highly structured spectra, such as fluorescent lamps (Nickerson 1960). The first widely accepted standard (metric) in this field, color rendering index (CRI), was introduced by the Commission Internationale de l'Éclairage (CIE) in 1965 (Nickerson and Jerome 1965; CIE 1965). It measures the degree to which the perceived colors of objects illuminated by the source under test conform to those under a standard source, i.e., it is a measure of the color fidelity. However, the color fidelity is not a single measure of color quality of light. Certain applications might require light that renders colors with increased or reduced chroma and improved or deteriorated color discrimination. Also, light sources have to account for individual or group preferences to the visual impression from colored objects. To that end, light sources must be assessed in broader terms of *color rendition*, which refers to different color quality properties of light sources rather than to solely the color fidelity.

The development of high-brightness blue light-emitting diodes (LEDs) (Nakamura and Fasol 1997) enabled the emergence of the solid-state lighting technology. White phosphor-converted LEDs and clusters of colored LEDs (Žukauskas et al. 2002a) greatly outperform conventional lamps in efficiency. The versatility of the solid-state lighting technology enables unprecedented diversity of the spectral power distributions (SPDs) of light sources and the development of light sources with instantaneously controlled SPD and color quality of illumination (Žukauskas et al. 2012a). This makes the problem of color rendition even more important.

This chapter deals with the problem of color rendition metrics. Section “[Color Rendering Index](#)” is devoted to a brief description of the CRI. Section “[Early Work Beyond Color Fidelity](#)” introduces early discussions of the color rendition properties that are beyond the color fidelity. In section “[Criticism and Refinement of the Color Rendering Index Metric](#),” we discuss criticism of the CRI and recent approaches to its appending and refining, such as the use of the two-metric system, color appearance models, and color quality scale. Section “[Color Rendition Metrics with High Numbers of Test Color Samples](#)” deals with the new approaches to color rendition using a very large number of test color samples, such as the color rendition icon and statistical method. In section “[Color Rendition Engineering](#),” concepts for color rendition engineering are discussed. Section “[Summary](#)” provides the summary of different approaches and conclusions.

Color Rendering Index

CIE 1995 presents the latest editorial update of the CRI metric. The CRI is based on calculating the color differences for 14 specified color samples, when the reference source is replaced by the source under assessment. The color differences, ΔE_i , are estimated within the three-dimensional $U^*V^*W^*$ color space (1964 CIE) as follows:

$$\Delta E_i = \sqrt{(\Delta U_i^*)^2 + (\Delta V_i^*)^2 + (\Delta W_i^*)^2}, \quad i = 1, 2, \dots, 14, \quad (1)$$

where ΔU_i^* and ΔV_i^* are the shifts of chromaticness indices and ΔW_i^* is the shift of the lightness index.

These color differences are used for the estimation of 14 respective special color rendering indices:

$$R_i = 100 - 4.6\Delta E_i, \quad i = 1, 2, \dots, 14. \quad (2)$$

The average of the first eight test color samples is called the general color rendering index:

$$R_a = \frac{1}{8} \sum_{i=1}^8 R_i. \quad (3)$$

As seen from Eqs. 1 and 2, the general CRI has a maximal value of 100 when the eight color shifts are zeros. The scaling parameter 4.6 in Eq. 2 is selected to obtain the value of $R_a = 51$ for the warm white halophosphate fluorescence lamp.

The initial data for calculating CRI are the SPDs of the source under assessment $S_t(\lambda)$ and the reference source $S_r(\lambda)$ and the reflectivity spectra of 14 test color samples $\rho_i(\lambda)$. The reference source must have the same correlated color temperature (CCT) as that under assessment and is either a blackbody radiator (for CCTs below

5000 K) or a CIE daylight-phase illuminant (for CCTs above 5000 K) (CIE 2004a; Davis and Ohno 2010). The chromaticity difference of the two sources must not exceed a specified limit. The test color samples are selected from the Munsell palette and have specified hue, chroma, and lightness. The first eight test color samples used for the estimation of the general CRI have different hues, moderate chroma (from 4 to 8), and equal lightness of 6. The samples 9–12 have increased chroma (from 8 to 13) and the remaining two samples (13 and 14) mimic the color of human complexion and leaf green, respectively.

The 1995 CIE procedure starts from calculating the tristimulus values X_t , Y_t , and Z_t and X_r , Y_r , and Z_r for the SPDs of the two sources $S_t(\lambda)$ and $S_r(\lambda)$, respectively, as follows:

$$\begin{aligned} X &= k \int_{380\text{nm}}^{780\text{nm}} \bar{x}(\lambda) S(\lambda) d\lambda, \\ Y &= k \int_{380\text{nm}}^{780\text{nm}} \bar{y}(\lambda) S(\lambda) d\lambda, \\ Z &= k \int_{380\text{nm}}^{780\text{nm}} \bar{z}(\lambda) S(\lambda) d\lambda, \end{aligned} \quad (4)$$

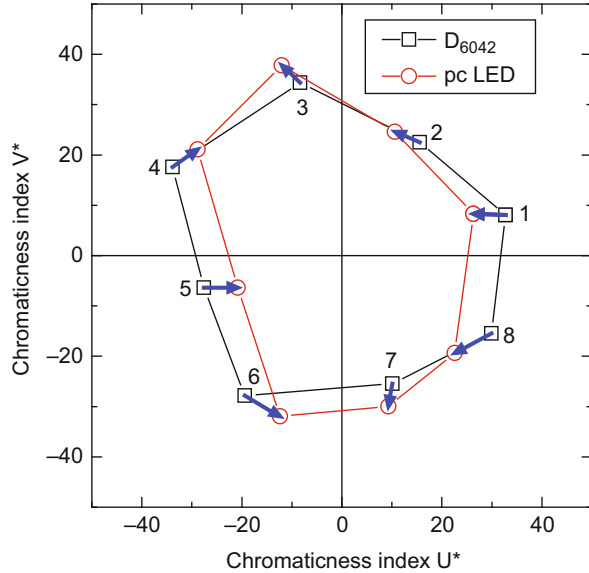
where $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ are the 1931 CIE 2° standard observer color-matching functions (Wyszecki and Stiles 2000) and k is the normalizing coefficient selected for each illuminant in such a way that their tristimulus values Y equal 100. Then the SPDs of the light reflected from each of the 14 test color samples are calculated for each of two illuminants and the tristimulus values for these SPDs are calculated accordingly.

Basing on the obtained tristimulus values, the 1960 CIE uniform chromaticity scale (UCS) coordinates (u_r, v_r) , (u_t, v_t) , (u_{ri}, v_{ri}) , and (u_{ti}, v_{ti}) for the reference and assessed source and for the light of the two sources reflected from the test color samples, respectively, are calculated using the transformations:

$$\begin{aligned} u &= \frac{4X}{X + 15Y + 3Z}, \\ v &= \frac{6Y}{X + 15Y + 3Z}. \end{aligned} \quad (5)$$

In addition, the UCS coordinates of the test color samples (u_{ki}, v_{ki}) under the assessed source are shifted to (u'_{ti}, v'_{ti}) to account for chromatic adaptation, which occurs when the chromaticity of the assessed source does not exactly match that of the reference source. The transformation of the chromaticity coordinates is based on the von Kries hypothesis, which assumes that each of the three retinal cone responses adapt the gain to achieve the color constancy (in this case, of reference and assessed sources).

Fig. 1 Chromaticnesses of the eight color test samples used in the CRI metric as points in the $U^* - V^*$ plane of the 1964 CIE color space. The reference light source and the source under assessment are the 6042 K daylight-phase illuminant and daylight phosphor-converted LED, respectively. The *arrows* show the chromaticness shifts used in the CRI calculation and the lines delineate the gamut areas of the two sets of chromaticnesses (Žukauskas et al. 2010a)



Finally, the shifts of the chromaticness indices (ΔU_i^* and ΔV_i^*) and lightness index (ΔW_i^*) that are used in Eq. 1 are estimated as follows:

$$\begin{aligned}
 \Delta U_i^* &= 13 [W_{ti}^* (u'_{ti} - u_r) - W_{ri}^* (u_{ri} - u_r)], \\
 \Delta V_i^* &= 13 [W_{ti}^* (v'_{ti} - v_r) - W_{ri}^* (v_{ri} - v_r)], \\
 \Delta W_i^* &= W_{ti}^* - W_{ri}^*.
 \end{aligned}
 \tag{6}$$

Here the lightness indices for all SPDs of light reflected from the test color samples are obtained from the respective tristimulus values Y using the following relation:

$$W^* = 25Y^{1/3} - 17.
 \tag{7}$$

Figure 1 illustrates the CRI concept. In the $U^* - V^*$ plane of the 1964 CIE color space, the chromaticness of the eight test color samples under the daylight phosphor-converted LED (CCT = 6042 K, $R_a = 70$) are compared with those of the same samples under daylight-phase illuminant with the same CCT. The arrows show the chromaticity shifts used in the CRI calculation.

Early Work Beyond Color Fidelity

Already at the early stages of the use of CRI metric, many approaches attempted the assessment of the color rendition properties of light sources other than the color fidelity. The main reason was that some light sources were found to increase color saturation.

Such sources improved the subjective rating of color appearance of the illuminated objects due to subjective preferences and/or enhancing the color-discrimination ability of human vision. The approaches worth mentioning are the flattery index (Judd 1967), color-preference index (Thornton 1974), gamut area (color-discrimination index) (Thornton 1972), visual clarity (Aston and Bellchambers 1969; Hashimoto and Nayatani 1994), and color rendering capacity (Xu 1983). Basically, these color rendition concepts are related to the variation of chroma of the object colors under different illuminants.

For instance, Judd's flattery index (Judd 1967) was devised to measure the degree, to which a light source succeeds in flattering subjects by promoting "an optimistic viewpoint." The index rates the reference (high-fidelity) source by a value of 90 and increases up to 100, when the chromaticities of the 10 color test samples from the CRI metric (1–10, 13, and 14) approach particular "preferred" chromaticities, such as complexion, cosmetics, butter, foliage, and grass (Fig. 2). Most of the "preferred" chromaticities have somewhat higher chroma than those of the test samples. Further refinement of the flattery index resulted in the introduction and psychophysical validation of the color-preference index (CPI) (Thornton 1974), which was defined basing on the average vector lengths of eight test color samples (i.e., similarly to the general CRI). For the CIE standard illuminant D_{65} , the color-preference index is set to 100, whereas its value for the "ideal" Judd's illuminant with the same CCT attains the value of 156.

In still another approach, the area G of the octagon formed in the 1960 CIE color space by the chromaticity coordinates of the first eight CRI test color samples has been claimed to serve as a measure of color discrimination (Thornton 1972). The light sources that result in a higher chroma of the test color samples have larger values of G . When normalized to the gamut area of the CIE standard illuminant C (rated at 100 arbitrary units), the standard illuminants with lower CCT, as well as

Fig. 2 Chromaticities of the 10 test color samples in Judd's flattery index. The *points* show the actual chromaticity coordinates in the 1960 CIE color space under a reference illuminant (D_{65}) and the heads of the *arrows* indicate the "preferred" chromaticities (Judd 1967)

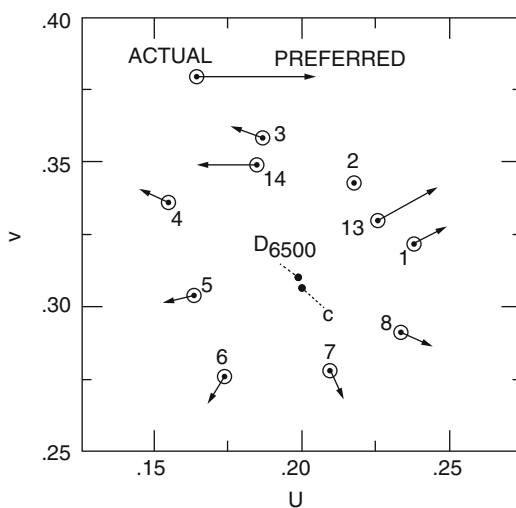
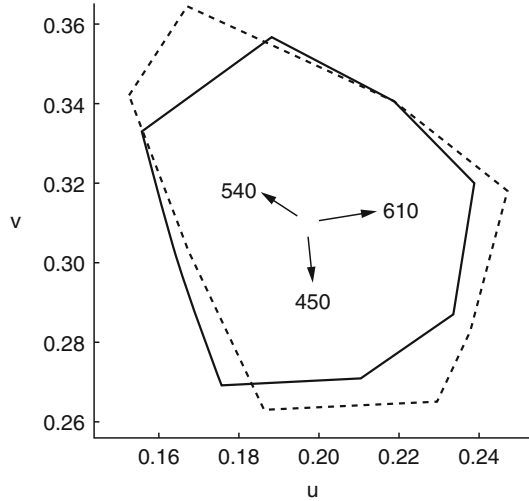


Fig. 3 Gamut areas of the CIE standard illuminant C (solid lines) and of metameric 450/540/610 nm illuminant (dashed lines) in the 1960 CIE color space (Thornton 1972)



most fluorescent lamps, have reduced values of G . However, Thornton discovered that the light sources with the gamut area in excess of 100 are feasible. In particular, a hypothetical lamp with the SPD composed of three narrow-band components peaking at 450 nm, 540 nm, and 610 nm and metameric to the CIE standard illuminant C was shown to have G of 123 (Fig. 3). Also, illuminants having a larger gamut area were found to be subjectively preferred compared to those with a lower gamut despite similar or even somewhat lower values of R_a .

One more color rendition concept beyond color fidelity is the visual clarity, which is the feeling of clear distinction between the surface colors of various objects under a particular illuminant. Initially, the visual clarity has been linked to only color rendering (fidelity) (Aston and Bellchambers 1969). However, later, the general CRI was found to be unable to predict the effect of the visual clarity under different illuminants (Hashimoto and Nayatani 1994). Moreover, this effect was attributed to the feeling of contrast under illumination specified by the four-color component gamut area in a brightness and colorfulness space. Explicitly, the feeling of contrast for the two-color combination was expressed as a linear combination of the color difference and the maximum chroma of the two colors. Also it has been established that lightness distortions are better tolerated than hue distortions, and the distortions increasing chromatic saturation might be even preferred.

Xu (1983) introduced the color rendering capacity (CRC), which is a measure of the maximum possible number of different colors that can be resolved under the given illuminant. This measure is based on estimating the maximum chromaticity ranges (i.e., gamut areas) of rendered colors for different luminances. CRC was proposed to assess the SPDs of illuminants in terms of making a given chromatic environment appear more colorful and brighter. Also, a correlation between CRC and visual clarity, which is due partially to increased chroma of object colors, has been traced.

In addition to defining the color rendition properties due to the chromatic and gamut area effects, the early work on color rendition beyond color fidelity also introduced more advanced color rendition concepts based on the improved vector approach, which takes into account all directions of the color shifts. Pointer (1986) analyzed color-shift vectors in the 1976 CIE UCS color space for 18 colored patches of the so-called Macbeth color checker chart and introduced the hue, chroma, and lightness indices for each of four hue groups (red, yellow, green, and blue) and mean hue, chroma, and lightness indices for the entire set of colored patches. The weighted sum of the three mean indices has been proposed as an integral measure of the appearance of colors under a given illuminant.

Worthey (2004) showed that a matrix formulation could approximately describe a set of color-shift vectors representing the replacement of an illuminant by another one. A single 3×3 “color rendering matrix” \mathbf{P} can be used for the conversion of the tristimulus values expressed in the opponent color system. The diagonal matrix elements P_{11} , P_{22} , and P_{33} represent the gain or loss of lightness, redness or greenness, and yellowness or blueness, respectively, and the off-diagonal elements represent the effect of the cross-translation of the stimuli. The method has been validated with the sets of 8 and 36 Munsell color samples that show systematic patterns of the color-shift vector distributions.

Criticism and Refinement of the Color Rendering Index Metric

Since the very introduction, numerous drawbacks of the CRI metric have been pointed out (CIE 1995; van Trigt 1999; Schanda 2002; Worthey 2003; Bodrogi 2004; Guo and Houser 2004; Davis and Ohno 2005; Sándor and Schanda 2006; CIE 2007; Žukauskas et al. 2009; van der Burgt and van Kemenade 2010). The major group of these drawbacks are related to the limited accuracy of the color rendering indices: small number of test color samples, the use of only samples with moderate chroma in the general CRI, the lack of uniformity of the color space used in respect of the perceived color differences, inaccuracy of the chromatic adaptation transform, and the doubts about the top ranking of the reference illuminants used (especially at extreme values of CCT). Also, the CRI has been criticized for the arithmetic averaging of the color differences that are of very different magnitude and for using confusing scaling, which for large color differences might yield negative values.

The limited accuracy due to the small number of test color samples used in the calculation of the general CRI can be somewhat mitigated by considering some special color rendering indices. In particular, the general CRI is known to noticeably lack relevance in respect of the red color (Bodrogi et al. 2004). Therefore in addition to R_a , many researchers and manufacturers of LEDs specify the special color rendering index for “strong red” test color sample (R_9).

However, probably, the key drawback of the CRI metric is that the direction of the color shifts is disregarded with only the magnitude of the shifts being accounted for. As discussed in section “Early Work Beyond Color Fidelity,” color shifts that

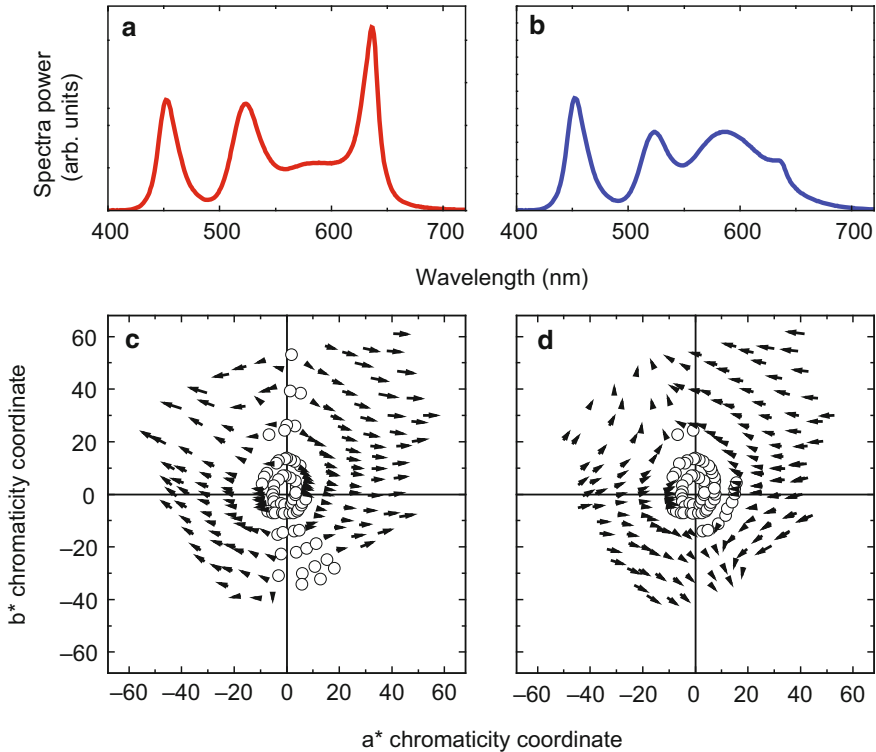


Fig. 4 (a, b) SPDs of two different RAGB blends (CCT = 4500 K) with $R_a = 80$. (c, d) corresponding distributions of the color-shift vectors for 218 Munsell samples of value/6 in the $a^* - b^*$ chromaticity plane of the CIELAB color space. The *arrows* schematically show perceptually noticeable color shifts of the samples and the *open circles* show the samples that have very small color shifts (within three-step MacAdam ellipses) (Žukauskas et al. 2013a AIC)

increase the gamut area of a set of test color samples might have the preferential impact for the visual impression of observers, whereas those color shifts that decrease the gamut area might result in the opposite effect. This causes the ambiguity of the CRI metric in that two sources having the same value of a color rendering index below 100 might render some colors in completely opposite way: the same colors appear more saturated or less saturated in these two cases, respectively. Figure 4 illustrates this ambiguity for two metameric SPDs of a tetrachromatic red-amber-green-blue (RAGB) solid-state lamp, both having an R_a of 80. The spectra differ just in the ratio of the relative partial radiant fluxes of the red and amber components (cf. Fig. 4a, b). However, the patterns of the distributions of the color-shift vectors are completely different: in Fig. 4c the majority of the vectors are directed outward showing the increased color saturation of the samples, whereas in Fig. 4d the majority of the vectors indicate on the decreased saturation.

The ambiguity of the CRI metric has resulted in the confusion while rating different light sources, especially when red-green-blue (RGB) LED clusters were

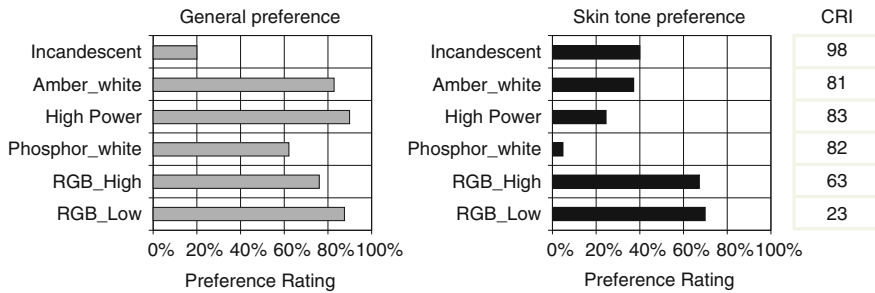


Fig. 5 Results of subjective evaluation of light sources with the different general CRI for general preference and skin tone preference (Narendran and Deng 2002)

compared to the fluorescent lamps and phosphor-converted LEDs (Narendran and Deng 2002; Shakir and Narendran 2002; Bodrogi et al. 2004; Nakano et al. 2005; Sándor and Schanda 2006; CIE 2007). Typically, the RGB LED clusters, which have low R_a but high color-saturating ability (Žukauskas et al. 2010b), were subjectively rated higher than the fluorescent lamps and phosphor-converted LEDs, which have higher values of R_a but lack the color-saturating ability or even dull colors (Žukauskas et al. 2009; Žukauskas et al. 2012b). Figure 5 illustrates the effect of such subjective “misrating” observed by Narendran and Deng (2002).

Eventually, the CIE concluded that “visual experience has shown that the current CRI based ranking of a set of light sources containing white LED light sources contradicts the visual ranking” and recommended the development of a new color rendering index or a set of indices (CIE 2007).

Some of the drawbacks of the CRI metric were attempted to mitigate by the CIE by updating the set of test color samples, using a more equidistant color space CIELAB, a finite set of reference illuminants and advanced transform for chromatic adaptation, and improving scaling (Schanda 1999). However, these mitigations resulted in no new standard due partially to the lack of consensus between researchers and manufacturers.

The color appearance models, which specify additional viewing conditions such as illuminance, background color, and surround, might considerably improve the accuracy of evaluating the color difference of the test color samples under different illuminants. For instance, CIECAM02 color appearance model (CIE 2004b) uses the advanced chromatic adaptation transform (CIECAT02) and specifies the color appearance of an object in terms of lightness, brightness, colorfulness, chroma, saturation, hue composition, and hue angle. Basing on the highly uniform color space of lightness (J), colorfulness (M), and hue angle (h) used in CIECAM02, an accurate calculation of the color difference and a corresponding color rendering index (CRI-CAM02UCS) have been elaborated (Luo 2011). Smet et al. 2013 proposed a further refinement of this approach (CRI2012 color-fidelity index) based on 17 thoroughly selected color test samples, the use of CIE 1964 10° standard observer color-matching functions, and improved averaging and scaling.

The color harmony rendering index (HRI) is an alternative approach of the color difference calculation (Bodrogi et al. 2004). This index sums the differences of the distances between all possible pairs of the samples instead of summing the color differences for each color sample illuminated by the source under assessment and a reference illuminant (see Eqs. 1, 2, and 3). Such summing is based on the assumption that the subjective appraisal of color quality of illumination is due to the “internal color harmony” between the color test samples.

The ambiguous treatment of chromatic distortions, as well as the recently established limitations in respect to rapidly evolving field of solid-state lighting, stimulated the search for more refinements and amendments of this metric. The simplest approach has been proposed by Žukauskas et al. (2013a). Within this approach, the ambiguity in ranking a set of light sources in color rendition quality is mitigated by the supplementation of the general CRI score by an additional symbol that indicates on the dominant type of chromatic distortion (saturating or dulling). For instance, the SPDs shown in Fig. 4a, b, where the dominant chromatic distortions are the increased and reduced saturation of colors, could be characterized by modified indices 80S and 80D, respectively. According to this approach, many phosphor-converted LEDs that are based on yellow-blue blends would be designated as color dulling (“D”), whereas most of RGB LED clusters would be attributed to color-saturating (“S”) light sources. It is to be noted that this approach does not provide a single order of ranking for a set of light sources. Color-saturating and color-dulling light sources are to be considered separately, with the understanding that the color-saturating sources receive more subjective preferences than the color-dulling sources.

Still another approach was introduced by Rea and Freyssinier-Nova (2008). It is based on the used two metrics, CRI for color fidelity (i.e., how “natural” objects appear) and the gamut area index (GAI) for chromatic effects (i.e., how “vivid” objects appear). Similarly to the gamut area introduced by Thornton (1972), GAI measures the relative change in the area of the polygon defined by the chromaticities of the eight CRI test color samples. However for simplicity, GAI uses the $U^* - V^*$ chromaticity plane of the 1964 CIE color space, which is also used in the CRI metric. Besides, the single reference illuminant used in the calculation of GAI has a flat SPD (CIE standard illuminant E). Just like for the gamut area G metric, the sources with a lower CCT have lower values of GAI. While GAI might be useful for comparing color rendition by different sources (Rea and Freyssinier-Nova 2008), its drawbacks are similar to those of the general CRI (such as using a small number of the test color samples and no accounting for the nonuniformity of the color space). Moreover, GAI might deceptively compensate the increased chroma of some test color samples by the decreased chroma of some other samples, distorting the effect on the net gamut area, since GAI integrates both positive and negative contributions to the gamut variation. For example, for the daylight phosphor-converted LED (Fig. 1), the gamut area under the LED is only 85 % of that under the reference illuminant, even though three of the eight test color samples used (3, 6, and 7) noticeably gain in saturation.

One of the most advanced refinements of the CRI metric is the color quality scale (CQS) developed at the National Institute of Standards and Technology (NIST) by Davis and Ohno (2005, 2010). CQS goal is to mitigate the main drawbacks of CRI (Davis and Ohno 2005), even though it uses the same principle of comparing the appearance of test color samples illuminated by a light source under assessment and the reference illuminant. CQS uses the same reference illuminants as in the CRI metric. However, CQS contains numerous important updates. First, the accuracy is increased by employing a set of 15 test color samples with a larger chroma (about/11 in average) than for the samples used to determine the general CRI. Second, the color differences are estimated in the CIELAB color space (CIE 2004a), which is much more equidistant in respect of perceivable color differences than the $U^*V^*W^*$ color space used in the CRI metric. The respective lightness L_i^* and chromaticity coordinates a_i^* and b_i^* of the test color samples in the CIELAB color space are obtained from the corresponding tristimulus values (X_i , Y_i , and Z_i) (Eq. 4) as follows:

$$\begin{aligned} L_i^* &= 116 \left(\frac{Y_i}{Y_w} \right)^{1/3} - 16, \\ a_i^* &= 500 \left[\left(\frac{X_i}{X_w} \right)^{1/3} - \left(\frac{Y_i}{Y_w} \right)^{1/3} \right], \\ b_i^* &= 500 \left[\left(\frac{Y_i}{Y_w} \right)^{1/3} - \left(\frac{Z_i}{Z_w} \right)^{1/3} \right], \end{aligned} \quad (8)$$

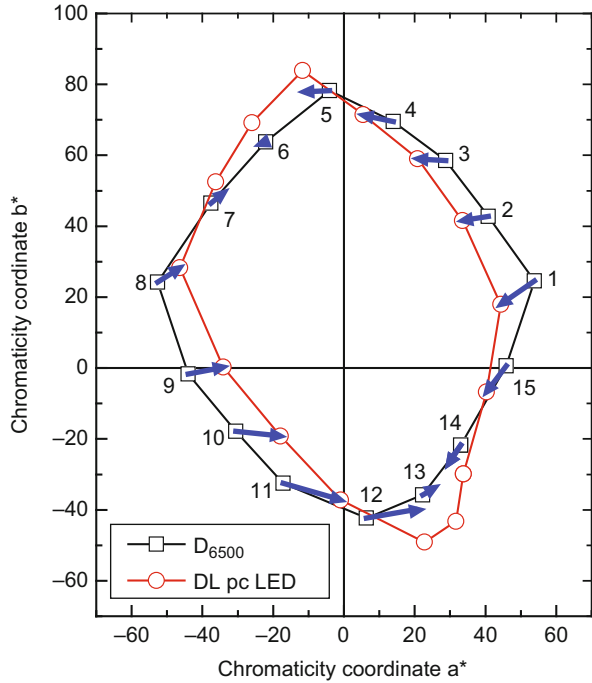
where X_w , Y_w , and Z_w are the tristimulus values of either of the two illuminants. When the chromaticities of the light source under assessment and reference illuminant do not exactly match, the tristimulus values of the assessed light source are transformed to the “corresponding” values using the CMCCAT2000 chromatic adaptation transform at a high luminance provided by both illuminants (Li et al. 2002; Davis and Ohno 2010).

The initial idea of the CQS metric (Davis and Ohno 2005) is to use an integral figure of merit to account for the subjective color saturation preferences by quantifying the ability of a lighting source to render colors with high fidelity and with an increased chromatic saturation. To this end, CQS introduces a reduced color difference for each test color sample, ΔE_i^* , which excludes the increase of chromatic saturation:

$$\Delta E_i^* = \left[(\Delta L_i^*)^2 + \left(\operatorname{Re} \sqrt{-\Delta C_i^*} \right)^2 + (\Delta H_i^*)^2 \right]^{1/2}. \quad (9)$$

Here ΔL_i^* , ΔC_i^* , and ΔH_i^* are the shifts in lightness chroma, and hue, respectively, between the reference source and the source under assessment. These shifts are calculated from the corresponding values of the lightness and chromaticity coordinates as follows:

Fig. 6 Chromaticity of the 15 color test samples used in the CQS metric in the $a^* - b^*$ plane of the CIELAB color space. *Squares* and *circles* show the chromaticities under the reference source (6500 K daylight illuminant) and daylight phosphor-converted LED, respectively. The *lines* delineate the gamut areas of the two sets of chromaticities and the *arrows* show the reduced chromaticity shifts (with positive increase in chromatic shifts excluded) used in the scoring (Žukauskas et al. 2010a)



$$\begin{aligned}
 \Delta L_i^* &= L_i^* - L_{ri}^* \\
 \Delta C_i &= \left[(a_{ti}^*)^2 + (b_{ti}^*)^2 \right]^{1/2} - \left[(a_{ri}^*)^2 + (b_{ri}^*)^2 \right]^{1/2}, \\
 \Delta H_i &= \arctan(b_{ti}^*/a_{ti}^*) - \arctan(b_{ri}^*/a_{ri}^*).
 \end{aligned}
 \tag{10}$$

As seen in Fig. 6, the reduced chromaticity shifts (arrows) with negative chromatic shifts are accounted for (samples 1, 2, 3, 4, 8, 9, 10, 11, and 15) and those with positive chromatic shifts are excluded (samples 5, 6, 7, 12, 13, and 14).

Furthermore, the CQS metric introduces the root mean square (RMS) of the reduced color differences:

$$\Delta E_{\text{RMS}}^* = \sqrt{\frac{1}{15} \sum_{i=1}^{15} (\Delta E_i^*)^2}.
 \tag{11}$$

This averaging ensures that a large color difference for any sample has a higher influence on the overall rating, i.e., the strong distortion of the color of a single test color sample cannot be compensated by low reduced color differences of the rest of the samples.

The final steps are a nonlinear conversion to a 0–100 rating scale and accounting for low CCTs, which result in noticeably reduced gamut areas of the test color samples. As a result, the general CQS is defined as:

$$Q_a = M_{\text{CCT}} \times 10 \ln \left\{ \exp \left[\frac{(100 - 3.01 \Delta E_{\text{RMS}}^*)}{10} \right] + 1 \right\}. \quad (12)$$

Here 3.01 is the scaling factor selected by equating the average CQS score for the standard set of fluorescent lamp spectra to the average general CRI for these lamps and M_{CCT} is the CCT factor, which penalizes light sources with CCT below 3500 K. The same scaling formula as Eq. 12 can be used for estimating 15 special CQS indices.

However, the general CQS metric is unable to completely distinguish between the light sources with different ability in saturating colors. For instance, the sources that render the color of a sample with increased ($\Delta C_i > 0$) and without increased ($\Delta C_i = 0$) chroma, respectively, can have the same Q_a score, i.e., this metric is still ambiguous. In order to mitigate this ambiguity of the integral figure of merit, additional scales have been proposed (Davis and Ohno 2010). One such scale is the color-fidelity scale, Q_f , which is defined similarly to the general CQS with the difference in that the root mean square is applied to the color differences without reduction (i.e., ΔC_i is taken into account regardless of the sign) and the scaling factor is changed to 2.93:

$$Q_f = M_{\text{CCT}} \times 10 \ln \left\{ \exp \left[\frac{(100 - 2.93 \Delta E_{\text{RMS}})}{10} \right] + 1 \right\}. \quad (13)$$

Another important figure of merit is the color gamut scale, Q_g , which is defined similarly as GAI but with the area of a polygon formed by the chromaticities of the 15 CQS samples illuminated by the light source under assessment being calculated in the CIELAB color space and using the normalization area under the CIE standard illuminant D_{65} . Q_f and Q_g are the advanced successors of the general CRI and GAI, respectively, with some of their drawbacks inherited (such as the ambiguity of color fidelity due to the disregard of the directions of the chromatic shifts and the limitations of gamut area due to the compensation of positive and negative chromatic shifts in different test samples; see Fig. 3). However, these two scales have distinct meaning and, when used together, could comprise an advanced two-metric system similar to that introduced by Rea and Freysinnier-Nova (2008). This has been confirmed by the analysis of 22 color rendition metrics showing that the measures of color fidelity and gamut area contain most of the information on the color rendition of the light sources (Houser et al. 2013).

In addition, Davis and Ohno (2010) introduced the color-preference scale, Q_p , which differs from the general CQS by additionally rewarding the light sources for increased saturation of the test color samples. However, this scale has not been psychophysically validated and might cause deceptive overrating of the light sources that not only increase color saturation but also severely distort hues, such as some RGB LED sources (see section “Color Rendition Engineering.”)

Color Rendition Metrics with High Numbers of Test Color Samples

An alternative way compared to CRI and its refinements is the development of more radical concepts for assessing the color quality of the light sources using a large number of the test color samples having very different chroma. In this case, analyzing and sorting the color-shift distributions replaces summing color differences (which is difficult to apply because of their different magnitudes and significance). One of such methods is the categorical color rendering based on subjective sorting of a large number of the color samples under different illuminants into the color categories specified by their color names (Boynton et al. 1990). Yaguchi et al. (2001) applied this method for grouping 292 Munsell color test samples into 11 color categories (red, green, yellow, blue, orange, pink, purple, brown, white, gray, black) under 14 illuminants. The color categories were specified within the color space of the CIECAM97 color appearance model. The calculation of the categorical color rendering index (CCRI) was based on the shift of the boundaries of the color name regions in respect of the D_{65} illuminant. A disagreement between the CCRI and the general CRI was found for some of the illuminants.

Another large sample number approach involves the analysis and graphical representation of color rendition vectors (CRVs) (van der Burgt and van Kemenade 2010). A CRV connects the color point of a test color sample illuminated by the reference source to the color point corresponding to the source under assessment. The CRV contains information on both the magnitude and direction of the color shift. Figure 7a shows the distribution of the CRVs in the CIELAB color space for a standard cool-white halophosphate fluorescent lamp. The arrows pointed outward and inward indicate the increased and decreased saturation of colors, respectively. The tangential components of the arrows indicate the hue distortion. The analysis of such vector distributions for various light sources shows that the distributions of the CRV hue and chroma components have characteristic periodic patterns in respect of the hue angle (Fig. 7b). (The lightness components only marginally contribute to the CRVs magnitude.) Such patterns show that a small number of test color samples used in the general CRI (8) and even in CQS (15) are insufficient for the accurate assessment of the color rendition properties of light sources.

Figure 7c shows the snapshot of the information on the hue and saturation shifts for the same lamp in 36 hue segments for 5600 colors cumulated from six different data sets. (Different test color sample sets were used in order to minimize the dependence of the output on the color set.) In different hue segments, the magnitude and direction of the vectors indicate the dominant color distortion and the radial size of the colored icon is proportional to color fidelity. Such a graphical representation of color rendition properties of light sources can be described in full detail. For simplicity, a single figure of merit can be derived from the area of the icon. However such a figure of merit would lack information on the dominant directions of the CRVs and, therefore, would be unable to unambiguously describe the appearance of the illuminated colored objects.

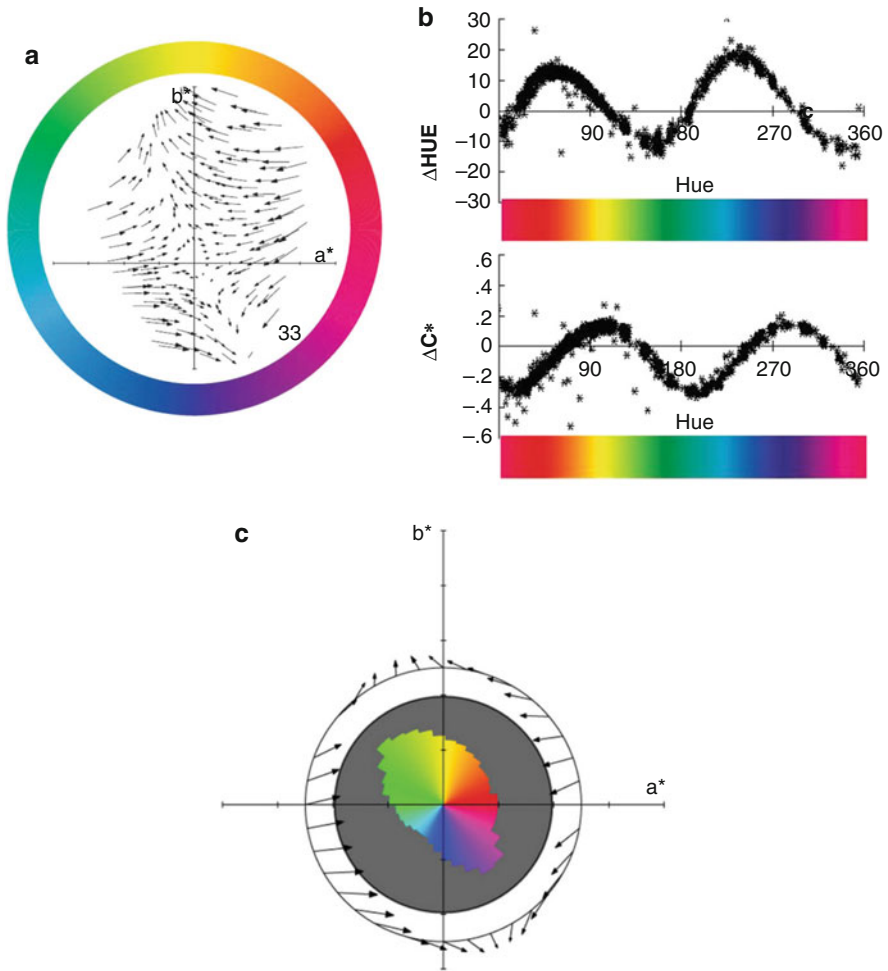


Fig. 7 Color rendition vectors of 215 test color samples (a), hue, and chroma components of the CRVs as functions of hue angle for 1240 test color samples (b), and color rendition icon with vectors showing the average hue and saturation shifts in 36 hue segments for 5600 colors from six data sets (c) for a standard cool-white halophosphate fluorescent lamp in respect of the metameric blackbody radiator (van der Burgt and van Kemenade 2010)

One more recent approach uses the computational sorting of the CRVs into several groups depending on the type of color distortion and statistical analysis of the sorting results (Žukauskas et al. 2009). This “statistical” metric is based on defining the number of differently rendered colors out of 1269 spectrophotometrically characterized Munsell samples (University of Eastern Finland, Spectral Color Research Group). Color rendition of each sample is examined via the behavior of the CRV in respect to an individual color-discrimination shape, which is built using the experimental data on just perceivable chromaticity and luminance differences.

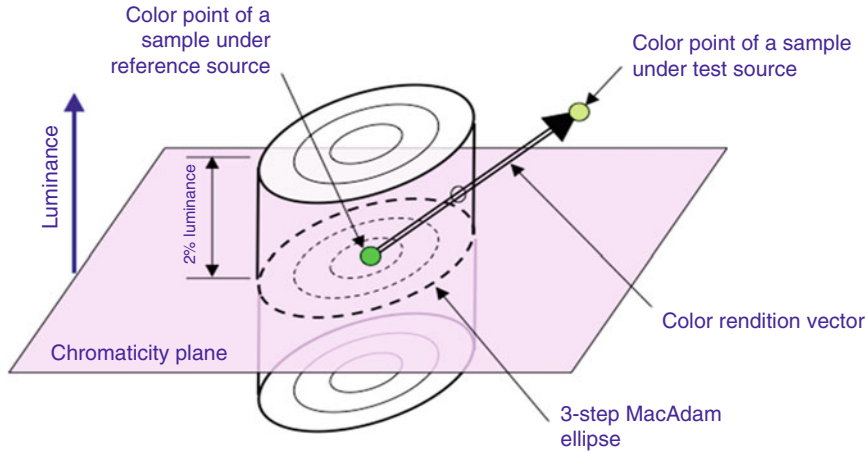


Fig. 8 Color discrimination shape (elliptical cylinder) used in the statistical approach to the assessment of color quality of lighting. The cross-section of the cylinder is the three-step MacAdam ellipse and the half-height is 2 % of luminance (Žukauskas et al. 2009)

Figure 8 shows such a color-discrimination shape (elliptical cylinder) with the center corresponding to the test sample color under reference illuminant (the reference illuminants as well as the chromatic adaptation transform are the same as in the CRI metric). The cross-section of the cylinder is identical to a triple-sized region of the just perceivable chromaticity and the half-height equals to three times the just perceivable luminance difference. The regions of the just perceivable chromaticity are the nonlinearly interpolated MacAdam ellipses defined for the constant luminance of $\sim 48 \text{ cd/m}^2$ (MacAdam 1942; Žukauskas et al. 2009), and the just perceivable luminance difference is 0.7 % (Wyszecki and Stiles 2000). If the CRV for a test color sample remains within the shape, the color is considered as rendered with high fidelity (indistinguishably from the reference source) and is scored to the color-fidelity index (CFI). If the CRV escapes from the shape, the sample is scored to one or several of the color-distortion indexes, depending on the direction of the vector escaping the shape. When the projection of a CRV directed toward increased or decreased saturation exceeds the size of the ellipse, the sample scores to the color saturation index (CSI) or color-dulling index (CDI), respectively. CRVs with the projections exceeding the size of the ellipse in the direction of different hue and/or larger than the half-height of the shape score to the hue distortion index (HDI) and/or to the luminance distortion index (LDI). Eventually, each statistical index is calculated as the percentage of the scoring samples in respect of the total number of the samples (1269).

The statistical metric has several important advantages in that the behavior of the CRV in respect of the color-discrimination shapes does not depend on the nonuniformity of the color space. Any sufficiently large set of the test color samples (including those obtained from the different data sets) can be used, and the results are comprehensive and easy to understand. In addition, a full graphical representation of

the color rendition characteristics of a light source can be obtained from a color quality chart within an appropriate color space (e.g., CIELAB); see Fig. 7a. The limitation of the statistical metric is treating the magnitudes of CRVs in a simplified way: either staying within the color-discrimination shape or stretching out of that. An appropriate weighting of the color-distortion scores can mitigate this drawback.

Color fidelity of light sources has also been assessed using different large sets of large number of test color samples (David 2014). Many sources, especially those having structured and truncated SPDs, were found to have lower color fidelity than that predicted by the CRI metric.

Color Rendition Engineering

The purpose of color rendition engineering is the tailoring of SPDs of light sources that have requested color rendition properties. Such engineering can be implemented within two approaches. Within the first approach, the fixed SPD of a light source is optimized in order to attain the maximal value of a particular color rendition index under defined constraints. The second approach is the development of light sources with tunable color rendition characteristics that can be selected depending on the needs and preferences of a user.

The formulation of the optimization problem starts from the selection of the number n of spectral components that constitute the composite SPD of a light source. The composite SPD of the light source under optimization is:

$$S(\lambda) = \sum_{i=1}^n c_i S_i(\lambda), \quad (14)$$

where $S_i(\lambda)$ and c_i ($i = 1, 2, \dots, n$) are the individual SPDs and partial radiant fluxes of the spectral components, respectively.

The spectral components can be provided by a variety of emitters such as different phosphors in fluorescent lamps and phosphor-converted LEDs, additive metals in metal-halide lamps, colored direct-emission and phosphor-converted LEDs in polychromatic LED arrays, etc. In many cases, the individual SPDs of the spectral components comprise single band having the peak wavelengths λ_i . Examples of single-peak components used in the optimization problems are Gaussian bands on the wavelength scale (Thornton 1971; Žukauskas et al. 2002b; Žukauskas et al. 2008a; Žukauskas et al. 2008b) and photon-energy scale (Lehmann 1963; Žukauskas et al. 2013b), and multi-Gaussian polynomials (Ohno 2005; He and Zheng 2010).

The relative partial fluxes of the components must satisfy the equations that follow from the principle of additive color mixing (Wyszecki and Stiles 2000). For both composite and component SPDs normalized to unit power, the color-mixing equations are as follows (Žukauskas and Vaicekauskas 2011):

$$\begin{cases} \sum_{i=1}^n c_i X_i = x \sum_{i=1}^n c_i (X_i + Y_i + Z_i), \\ \sum_{i=1}^n c_i Y_i = y \sum_{i=1}^n c_i (X_i + Y_i + Z_i), \\ \sum_{i=1}^n c_i = 1. \end{cases} \quad (15)$$

Here X_i , Y_i , and Z_i are the tristimulus values of the normalized SPD of the i -th colored component and x and y are the 1931 CIE chromaticity coordinates of the composite source.

For the selected set of individual model SPDs of single-band spectral components, the n peak wavelengths and n partial radiant fluxes are found through the maximization of the objective function F , which is a color rendition index (figure of merit) M (Žukauskas and Vaicekauskas 2011):

$$\begin{aligned} F(\lambda_1, \lambda_2, \dots, \lambda_n; c_1, c_2, \dots, c_n) &= M, \\ \text{subject to:} & \\ \{M_1 \in [M_{1\min}, M_{1\max}], \dots, M_p \in [M_{p\min}, M_{p\max}]\} & \end{aligned} \quad (16)$$

The objective function is usually subjected to a set of t constraints, which determine the feasible intervals of other figures of merit (M_1, \dots, M_p). The constraining figures of merit can be other color rendition indices (e.g., statistical hue distortion index HDI) and/or such parameters as the luminous efficacy of radiation (LER) or luminous efficacy of the source, and the bandwidth of the spectral components. An alternative approach to the introduction of the constraints is the use of the objective function that is a linear combination of several figures of merit, e.g., the general CRI and LER (Thornton 1971; Žukauskas et al. 2002b).

Generally for the SPD containing n spectral components, the optimization is to be performed within the $2n$ -dimensional parametric space of the peak wavelengths and relative partial radiant fluxes that are bound by the three color-mixing equations. Therefore, the objective function is maximized in the optimization domain, which is the parametric space with $2n-3$ degrees of freedom. For $n = 2$, the optimization domain has just one degree of freedom, i.e., the color rendition properties depend only on one variable (e.g., the peak wavelength of the short-wavelength component; the peak wavelength of the long-wavelength component and the two partial radiant fluxes are found from the color-mixing equations). For $n = 3$, two variables (e.g., the peak wavelengths of two components) can be independently adjusted.

In some problems, the set of spectral components (peak wavelengths) is already established and the optimization domain is the parametric space with $n-3$ degrees of freedom (for $n \leq 3$, complementary sets of variables are to be used and no optimization can be performed). In this case, the optimization of color rendition properties requires at least four spectral components. For instance, for $n = 4$, the optimization problem can be solved within a 1-dimensional parametric space (e.g., one relative

radiant flux; the rest ones are found from the three color-mixing equations). For $n = 5$, the relative radiant fluxes of two spectral components can be independently adjusted.

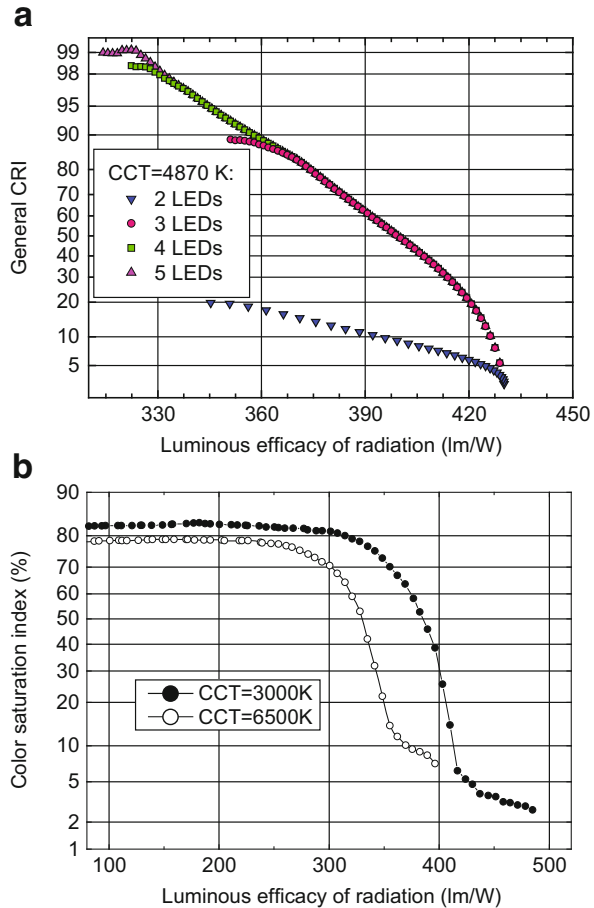
For a small number of the degrees of freedom (one or two), maximizing the objective function can be performed using simple computer routines, which perform exhaustive searching on a low-dimensional surface. For a larger number of degrees of freedom, the optimization requires sophisticated approaches of nonlinear programming that increase the operating speed of the searching routine. Examples of such approaches are stochastic moving of the parametric vector on the $2n-3$ -dimensional hypersurface (Žukauskas et al. 2002b), iterative method (Ohno 2005), and fast Pareto genetic algorithm (Zhong et al. 2012).

Using the CRI metric, Koedam and Opstelten (1971), Walter (1971), and Thornton (1971) optimized two-component and three-component SPDs of light sources. In particular for three-component SPDs, the peak values of 610, 540, and 450 nm were found to provide high values of both R_a and LER, whereas the peak wavelengths around 500 and 580 nm were found to be avoided (Thornton 1971). Four narrow-band components with the peak wavelengths of about 620, 580, 530, and 460 nm yielded the R_a values in excess of 90 (Walter 1978). Žukauskas et al. (2002b) optimized the general CRI for the model polychromatic LED clusters containing from 2 to 5 spectral components and established trade-offs (Pareto boundaries) between R_a and LER (Fig. 9a). Ries et al. (2004) established similar trade-offs between R_a and luminous output for clusters of practical LEDs. However, many SPDs having high values of the general CRI suffer from the deficiency in rendering red colors, which is indicated by the low values of the special color rendering index R_9 (Ohno 2005). This drawback of the maximization of the general CRI can be mitigated by using a constraint on the minimal value of R_9 (Zhong et al. 2012).

The optimization of SPDs of light sources using the color rendition indices different from the general CRI has been also performed, typically for solid-state devices. Using the CQS metric, Davis and Ohno (2005) have shown that an improved color fidelity in respect of R_a can be attained by shifting the red component of the RGB LED cluster to longer wavelengths. A linear combination of the general CQS and luminous efficacy was used as an objective function for the optimization of a pentachromatic LED cluster (Chien and Tien 2012). Berns (2011) applied the CQS approach for the optimization of LED clusters having the highest ability to saturate and desaturate the colors of illuminated objects.

The optimization of polychromatic LED clusters using the statistical metric also has resulted in the longer peak wavelengths of the Gaussian components in respect of those obtained by the maximization of the general CRI (Žukauskas et al. 2008a). The statistical approach has been also applied for the optimization of color-saturating (Žukauskas et al. 2010b) and color-dulling (Žukauskas et al. 2012b) abilities of the solid-state light sources. Similarly to the color fidelity, the color saturation ability of trichromatic LED clusters has been shown to be within a trade-off with LER (Fig. 9b). The statistical metric has been used for the optimization of SPDs of phosphor-converted LEDs with the of phosphor bands having Gaussian shapes on

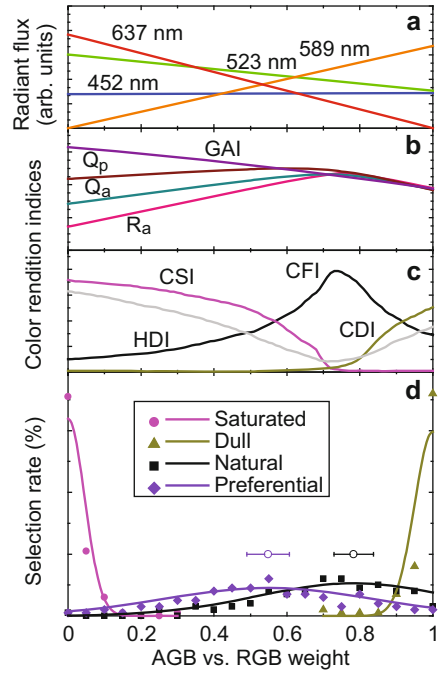
Fig. 9 Pareto boundaries in the phase space of the general CRI versus LER for polychromatic LED clusters containing 2-, 3-, 4-, and 5 30-nm wide spectral components (a) and in the phase space of statistical CSI and LER for 3 30-nm wide spectral components (After Žukauskas et al. 2002b, 2010b, respectively)



both wavelength scale (Žukauskas et al. 2008b) and photon-energy scale (Žukauskas et al. 2013b).

The distinction between different color rendition properties of light sources provided by the statistical metric and the progress in smart lighting technology led to the appearance of a “color rendition engine,” which is a light source with variable metameric SPD and, therefore, tunable color rendition properties. The color rendition engine contains a polychromatic cluster of at least four different LEDs with microcontroller driven multichannel current source. The simplest algorithm for the control of the tetrachromatic color rendition engine composed of red, green, amber, and blue LEDs is the dynamic formation of a composite SPD $S_{\Gamma}(\lambda)$ as the weighted sum of metameric (having the same CCT) color-saturating and color-dulling SPDs of the RGB and AGB subclusters, $S_{\text{RGB}}(\lambda)$ and $S_{\text{AGB}}(\lambda)$, respectively (Žukauskas et al. 2012a):

Fig. 10 (a) Variation of parameters of the tetrachromatic color rendition engine with AGB versus RGB weight at the CCT of 4500 K: radiant fluxes of the four colored LED groups (b) the general CRI, GAI, and CQS Q_a and Q_p (c) statistical indices CFI, CSI, CDI, and HDI (d) percentage of the subjective selections of the weight parameter for illumination characterized as “most saturated,” “most dull,” “most natural,” and “preferential” (After Žukauskas et al. 2012a)



$$S_I(\lambda) = \sigma S_{AGB} + (1 - \sigma) S_{RGB}. \tag{17}$$

Here the control parameter σ defines the weight of the AGB subcluster SPD in the overall SPD. At the color-saturating ($\sigma = 0$) and color-dulling ($\sigma = 1$) end points, the color rendition engine has the highest ability to saturate and dull the colors of illuminated objects, respectively. When the control parameter is varied within the end points, an infinite number of metameric SPDs with gradually varying color rendition properties are obtained, including those with the highest color fidelity or any individual preference.

Figure 10 shows the variation of the component radiant fluxes (a), color rendition indices (b) and (c), and subjective selection rates for the estimation of the color appearance of a simple scene containing familiar vegetables, fruits, and soft-drink cans (d) as functions of the control parameter. The results of subjective assessment (100 subjects) show that subjects succeeded in confidently finding the highly color-saturating (high GAI and high CSI) and highly color-dulling (low GAI and high CDI) end points defined as providing the “most saturating” and “most dull” appearance, respectively. Also, the RAGB blends having the highest color fidelity (high R_a , Q_a and CFI) were clearly identified with the “most natural” appearance. However, all these three figures of merit are ambiguous as the measures of the color saturating/desaturating ability due to nonmonotonous variation with the weight parameter. The “most preferential” color rendition conditions have been identified as those somewhat shifted from the high-fidelity blend toward the color-saturating end point,

i.e., when gamut area (color saturation) indices are increased at some expense of color-fidelity indices. However, the highest selection rate for the preferential color quality does not coincide with the peak of the proposed CQS color-preference scale, which shows that the latter color rendition property still lacks confident quantification.

As seen from Fig. 10, the color rendition engines allow for dynamically finding individual and object-specific preferences to color rendition without knowing indices, scores, and scales of color rendition metrics. Such engines can be easily adjusted to a desired chromaticity and can be supplemented by additional functionalities, such as instantaneous dimming for maintaining constant photochemical damage of the artwork pigments (Tuzikas et al. 2014) or for the control of circadian action. These engines can also meet specific visual needs, such as those of the elderly and color-deficient people.

Summary

Table 1 presents color rendition indices of standard illuminants and some practical light sources assessed using the above-considered two-metric system, color quality scale, and statistical metric (Lebedenko and Vaicekauskas 2014). Light sources can be classified into several qualitative categories, such as high-fidelity, color-saturating, color-dulling, and color-preference light sources. The first three categories rely on color rendition indices that have the highest or the lowest values (Žukauskas et al. 2010a), i.e., show the prevailing character of color rendition, and the fourth (color-preference) category can be introduced by analyzing the interplay of color-fidelity and color-saturating properties (Žukauskas et al. 2013b):

- High-fidelity light sources (typically, $R_a > 90$; $Q_f >$; $CFI > 50$) have SPDs covering the entire spectrum with the balanced spectral power in the wavelength ranges of both 530–610 nm and beyond 610 nm. Examples of such high-fidelity sources are non-filtered incandescent lamps, multiband fluorescent lamps, triphosphor fluorescent lamps (with some limitations), trichromatic diphosphor-converted RGB LEDs, RAGB LED clusters with particular blends, and some complementary clusters of white and colored LEDs (Žukauskas et al. 2010a).
- Color-saturating light sources (typically $Q_g > 120$; $CSI > 50$) usually significantly lack power in the 530–610 nm wavelength range. Examples of such color-saturating sources are RGB or red-cyan-blue (RCB) direct-emission LED clusters and diphosphor RCB LEDs with deep red and blue spectral components. (Some color-saturating light sources, such as RGB clusters of LEDs, have also high values of HDI; therefore, they do not necessarily provide the highest color discrimination due to distorted hues.)
- Color-dulling light sources ($CDI > 50$) have SPDs that lack power for wavelengths longer than 610 nm. Examples of such color-dulling sources are

Table 1 Color rendition indices of standard illuminants and light sources defined using different color rendition metrics

Light source	CCT (K)	LER (lm/W)	Two-metrics		Color quality scale				Statistical metrics					
			R_a	GAI	\bar{Q}_a	\bar{Q}_f	\bar{Q}_g	\bar{Q}_p	CFI	CSI	CDI	HDI	LDI	
CIE standard illuminant A	2856	155	100	56	98	98	98	98	98	100	0	0	0	0
CIE standard illuminant D ₆₅	6504	184	100	98	100	100	100	100	100	100	0	0	0	0
CIE standard illuminant E	5455	155	95	100	97	95	102	101	86	2	0	0	0	4
Fluorescent halophosphate	2778	310	52	44	54	55	76	54	12	1	72	46	58	
	6321	284	76	80	75	76	86	73	26	1	53	38	36	
Fluorescent triphosphor	2790	322	82	59	76	73	101	80	46	12	20	24	31	
	6114	274	80	94	76	75	94	78	38	2	31	30	35	
Fluorescent multiband	3024	278	92	67	89	88	100	92	74	7	8	8	15	
	7102	237	96	103	97	96	102	98	91	2	0	2	3	
Neodymium-glass filtered incandescent ^a	2763	220	77	68	87	83	110	98	37	52	0	15	27	
High-pressure mercury	3264	304	52	57	53	51	85	56	10	6	66	53	55	
High-pressure sodium	1714	265	13	15	31	32	39	31	9	0	83	27	67	
Metal halide	3966	285	70	55	72	73	81	66	33	0	55	30	27	
Phosphor-converted LED (diphosphor)	3359	295	90	65	88	88	96	88	61	0	25	10	12	
Phosphor-converted LED (monophosphor)	5000	348	65	71	68	68	90	70	16	5	58	52	49	
Phosphor-converted LED (monophosphor)	6042	325	71	85	70	69	90	73	17	4	53	50	45	

RGB LED cluster (637–523–452 nm)	3000	327	41	75	60	50	123	79	11	77	1	53	61
	4499	317	49	112	65	54	128	86	13	68	1	56	59
	6500	297	54	135	69	56	129	91	12	67	2	58	55
pcAGB LED cluster ^b (589–523–452 nm)	3000	446	28	33	35	38	54	26	12	1	67	50	62
	4499	399	52	55	48	50	67	39	21	1	57	47	51
	6500	355	64	70	54	56	72	47	24	0	51	43	46
RpcAGB LED cluster (452–523–589–637 nm) high-fidelity blend ^{b,c}	3000	344	95	63	91	89	103	95	87	1	4	5	4
	4500	333	93	86	93	90	106	97	76	2	4	9	5
	6500	310	92	104	91	88	107	97	66	9	4	11	5
RpcAGB LED cluster (452–523–589–637 nm) preferential blend ^{b,c}	3000	325	80	69	87	82	112	96	39	45	1	16	27
	4500	313	80	101	88	81	115	100	37	43	1	23	29
	6500	295	77	119	86	77	118	100	30	47	2	31	33

^aBouma (1938)

^bContains phosphor-converted amber LED

^cZukauskas et al. (2012b)

halophosphate fluorescent lamps, dichromatic daylight mono-phosphor-converted LEDs and other yellow-blue (YB) emitters, and amber-green-blue (AGB) LED clusters (Žukauskas et al. 2012b);

- Color preference light sources ($Q_f > 80$ and $Q_g > 120$; CFI \sim CSI > 30) have SPDs that moderately lack power in the 530–610 nm wavelength range and are rich in blue and red. Examples of such color-preference sources are neodymium-glass-filtered incandescent lamps, special fluorescent lamps (Thornton 1973), RAGB LED clusters with particular blends, and diphosphor-converted RCB LEDs.

From the data presented in Table 1, one can conclude that different color-fidelity metrics converge when the score is approaching 100 (see also Fig. 10). However, when a light source has color-fidelity indices well below 100, very different color rendition properties can account for this decrease (increased color saturating or dulling or a certain preferential color rendition).

To conclude, color rendition metrics follow advances in the lighting technology and the increasing of the sophistication and diversification of the needs of lighting users. Different approaches based on precise estimation of color difference and gamut area and color-shift vector analysis are presently available for experts, although the lighting industry is still waiting for an improved and comprehensive metric that could substitute for the outdated CRI. Of several color rendition properties, the most developed approaches have been elaborated for estimating color fidelity. Also, it is clear that maximizing color rendition indices related to color fidelity (also to gamut area) is in conflict with LER and luminous efficacy of light sources. Therefore, practical lamps should be developed within the trade-off between color rendition and energy saving.

A single figure of merit describing color fidelity is completely insufficient to assess the color quality of light sources. None of several candidates introduced for meeting this need are good enough: the general color quality scale suffers from ambiguity, the gamut area does not account for hue distortions that decrease color discrimination, and the subjectively established color quality preferences lack reliable quantification. Therefore, further work on the development of the second figure of merit for assessing the color rendition properties of light sources beyond color fidelity is required. An alternative to the second figure of merit is supplementing the color-fidelity index by qualitative descriptors based on the light source categorization, such as presented above.

The color rendition engines with tunable SPD, which can meet individual needs in color quality of illumination, provide a viable technological alternative for the further sophistication of the computational assessment of color rendition properties of light sources.

Acknowledgment The work at RPI was partially supported by the National Science Foundation (NSF) Smart Lighting Engineering Research Center (# EEC-0812056).

References

- Aston SM, Bellchambers HE (1969) Illumination, color rendering, and visual clarity. *Light Res Technol* 1:259–261
- Berns RS (2011) Designing white light LED lighting for the display of art: a feasibility study. *Color Res Appl* 36:324–334
- Bouma PJ (1938) The color reproduction of incandescent lamps and “Philiphon” glass. *Philips’ Tech Rev* 3:47–49
- Bodrogi P (2004) Colour rendering: past, present (2004), and future. In: Proceedings of the CIE expert symposium on LED light sources: physical measurement and visual and photobiological assessment, CIE publication no x026, pp 12–15
- Bodrogi P, Csuti P, Horváth P, Schanda J (2004) Why does the CIE color rendering index fail for white RGB LED light sources? In: Proceedings of the CIE expert symposium on LED light sources: physical measurement and visual and photobiological assessment, CIE publication no x026, pp 24–27
- Boynton RM, Fargo L, Collins BL (1990) Categorical color rendering of four common light sources. *Color Res Appl* 15:222–230
- Chien M-C, Tien C-H (2012) Multispectral mixing scheme for LED clusters with extended operational temperature window. *Opt Express* 20:A245–A254
- CIE (1965) Method of measuring and specifying colour rendering properties of light sources. CIE publication no 13
- CIE (1995) Method of measuring and specifying colour rendering properties of light sources. CIE publication no 13.3
- CIE (2004a) Colorimetry. CIE publication no 15
- CIE (2004b) A colour appearance model for colour management systems: CIECAM02, CIE publication no 159
- CIE (2007) Color rendering of white LED sources. CIE publication no 177
- David A (2014) Color fidelity of light sources evaluated over large sets of reflectance samples. *Leukos* 10:59–75
- Davis W, Ohno Y (2005) Toward an improved color rendering metric. *Proc SPIE* 5941:59411G
- Davis W, Ohno Y (2010) Color quality scale. *Opt Eng* 49:033602
- Guo X, Houser KW (2004) A review of color rendering indices and their application to commercial light sources. *Light Res Technol* 36:183–189
- Hashimoto K, Nayatani Y (1994) Visual clarity and feeling of contrast. *Color Res Appl* 19:171–185
- He G, Zheng L (2010) Color temperature tunable white-light light-emitting diode clusters with high color rendering index. *Appl Opt* 49:4670–4676
- Houser KW, Wei M, David A, Krames MR, Shen XS (2013) Review of measures for light-source color rendition and considerations for a two-measure system for characterizing color rendition. *Opt Express* 21:10393–10411
- Judd DB (1967) A flattery index for artificial illuminants. *Illum Eng* 62:593–598
- Koedam M, Opstelten JJ (1971) Measurement and computer-aided optimization of spectral power distributions. *Light Res Technol* 3:205–210
- Lebedenko D, Vaicekaskas D (2014) Light source assessment. Vilnius University Lighting Group, Vilnius. <http://demo.lrg.projektas.vu.lt/lcq/en/>
- Lehmann W (1963) Emission spectra of (Zn, Cd)S phosphors. *J Electrochem Soc* 110:754–758
- Li C, Luo MR, Rigg B, Hunt RWG (2002) CMC 2000 chromatic adaptation transform: CMCCAT2000. *Color Res Appl* 27:49–58
- Luo MR (2011) The quality of light sources. *Color Technol* 127:75–87
- MacAdam DL (1942) Visual sensitivities to color differences in daylight. *J Opt Soc Am* 32:247–274

- Nakamura S, Fasol G (1997) *The blue laser diode: GaN based light emitters and lasers*. Springer, Berlin
- Nakano Y, Tahara H, Suehara H, Kohda J, Yano T (2005) Application of multispectral camera to color rendering simulator. In: Nieves JL, Andres JH (eds) *Proceedings of the 10th Congress of the International Colour Association – AIC Colour 05*, pp 1625–1628
- Narendran N, Deng L (2002) Color rendering properties of LED light sources. *Proc SPIE* 4776:61–67
- Nickerson D (1960) Light sources and color rendering. *J Opt Soc Am* 50:57–69
- Nickerson D, Jerome CW (1965) Color rendering of light sources: CIE method of specification and its application. *Illum Eng* 60:262–271
- Ohno Y (2005) Spectral design considerations for white LED color rendering. *Opt Eng* 44:111302
- Pointer MR (1986) Measuring colour rendering – a new approach. *Light Res Technol* 18:175–184
- Rea MS, Freyssinier-Nova JP (2008) Color rendering: a tale of two metrics. *Color Res Appl* 33:192–202
- Ries H, Leike I, Muschaweck J (2004) Optimized additive mixing of colored light-emitting diode sources. *Opt Eng* 43:1531–1536
- Sándor N, Schanda J (2006) Visual color rendering based on color difference evaluations. *Light Res Technol* 38:225–239
- Schanda J (1999) Colour rendering, CIE TC 1-33 closing remarks. In: *CIE collection 1999. Vision and colour. Physical measurement of light and radiation*. CIE publication no 135, pp 10–17
- Schanda J (2002) The concept of color rendering revisited. In: *Proceedings of the 1st European conference on colour in graphics, image, and vision*. Poitiers, France, pp 37–41
- Shakir I, Narendran N (2002) Evaluating white LEDs for outdoor landscape lighting application. *Proc SPIE* 4776:162–170
- Smet KAG, Schanda J, Whitehead L, Luo RM (2013) CRI 2012: a proposal for updating the CIE colour rendering index. *Light Res Technol* 45:689–709
- Thornton WA (1971) Luminosity and color-rendering capability of white light. *J Opt Soc Am* 61:1155–1163
- Thornton WA (1972) Color-discrimination index. *J Opt Soc Am* 62:191–194
- Thornton WA (1973) Fluorescent lamps with high color-discrimination capability. *J Illum Eng Soc* 3:61–64
- Thornton WA (1974) A validation of the color-preference index. *J Illum Eng Soc* 4:48–52
- Tuzikas A, Žukauskas A, Vaitiekuskas R, Petrulevičius A, Vitta P, Shur M (2014) Artwork visualization using a solid-state lighting engine with controlled photochemical safety. *Opt Express* 22:16802–16818
- University of Eastern Finland, Spectral Color Research Group. <http://www.uef.fi/spectral/spectral-database>
- van der Burgt P, van Kemenade J (2010) About color rendition of light sources: the balance between simplicity and accuracy. *Color Res Appl* 35:85–93
- van Trigt C (1999) Color rendering, a reassessment. *Color Res Appl* 24:197–206
- Walter W (1971) Optimum phosphor blends for fluorescent lamps. *Appl Opt* 10:1108–1113
- Walter W (1978) Optimum lamp spectra. *J Illum Eng Soc* 7:66–73
- Worthington JA (2003) Color rendering: asking the question. *Color Res Appl* 28:403–412
- Worthington JA (2004) Color rendering: a calculation that estimates colorimetric shifts. *Color Res Appl* 29:43–56
- Wyszecki G, Stiles WS (2000) *Color science: concepts and methods, quantitative data and formulae*. Wiley, New York
- Xu H (1983) Color-rendering capacity of illumination. *J Opt Soc Am* 73:1709–1713
- Yaguchi H, Takahashi Y, Shioiri S (2001) A proposal of color rendering index based on categorical color names. In: *Proceedings of the International Lighting Congress, vol II*. Istanbul, 12–14 Sept 2001, pp 421–426

- Zhong P, He G, Zhang M (2012) Spectral optimization of the color temperature tunable white light-emitting diode (LED) cluster consisting of direct-emission blue and red LEDs and a diphosphor conversion LED. *Opt Express* 20:A684–A693
- Žukauskas A, Shur MS, Gaska R (2002a) Introduction to solid-state lighting. Wiley, New York
- Žukauskas A, Vaicekauskas R, Ivanauskas F, Gaska R, Shur MS (2002b) Optimization of white polychromatic semiconductor lamps. *Appl Phys Lett* 80:234–236
- Žukauskas A, Vaicekauskas R, Ivanauskas F, Vaitkevičius H, Shur MS (2008a) Rendering a color palette by light-emitting diodes. *Appl Phys Lett* 93:021109
- Žukauskas A, Vaicekauskas R, Ivanauskas F, Vaitkevičius H, Vitta P, Shur MS (2008b) Spectral optimization of phosphor-conversion light-emitting diodes for ultimate color rendering. *Appl Phys Lett* 93:051115
- Žukauskas A, Vaicekauskas R, Ivanauskas F, Vaitkevičius H, Vitta P, Shur MS (2009) Statistical approach to color quality of solid-state lamps. *IEEE J Sel Top Quantum Electron* 15:1753–1762
- Žukauskas A, Vaicekauskas R, Shur MS (2010a) Colour-rendition properties of solid-state lamps. *J Phys D Appl Phys* 43:354006
- Žukauskas A, Vaicekauskas R, Shur M (2010b) Solid-state lamps with optimized color saturation ability. *Opt Express* 18:2287–22951
- Žukauskas A, Vaicekauskas R (2011) LEDs in lighting with tailored color quality. *Int J High Speed Electron Syst* 20:287–301
- Žukauskas A, Vaicekauskas R, Vitta P, Tuzikas A, Petrusis A, Shur M (2012a) Color rendition engine. *Opt Express* 20:5356–5367
- Žukauskas A, Vaicekauskas R, Shur M (2012b) Color-dulling solid-state sources of light. *Opt Express* 20:9755–9762
- Žukauskas A, Vaicekauskas R, Vitta P, Shur M (2013a) Resolving the ambiguity of color fidelity indices. In: MacDonald L, Westland S, Wuerger S (eds) *Proceedings of the 12th congress of the International Colour Association – AIC Colour 2013*, vol 3, pp 1129–1132
- Žukauskas A, Vaicekauskas R, Vitta P, Zabaliūtė A, Petrusis A, Shur M (2013b) Color rendition engineering of phosphor-converted light-emitting diodes. *Opt. Express* 21:26642–26656

Photoreception for Human Circadian and Neurobehavioral Regulation

George C. Brainard and John P. Hanifin

Contents

Introduction	830
Circadian, Neuroendocrine, and Neurobehavioral Regulation by Light Wavelength	830
Phototransduction for Circadian, Neuroendocrine, and Neurobehavioral Regulation	832
Neural Pathways for Circadian and Neurophysiological Regulation	832
Circadian, Neuroendocrine, and Neurobehavioral Regulation by Light Intensity	834
Ocular and Physiological Elements that Mediate the Biological Effects of Light	836
Clinical and Nonclinical Applications of Solid State Light Therapy	838
Light Measurement for Biological and Behavioral Regulation	841
Conclusion	843
References	844

Abstract

Two convergent developments are transforming architectural lighting: (1) the advance of solid state lighting technologies and (2) the confirmation that light regulates human circadian, neuroendocrine, and neurobehavioral physiology, thereby influencing health and well-being. Analytic action spectra studies have shown peak sensitivity in the short-wavelength portion of the visible spectrum from 447 to 484 nm for the biological and behavioral effects of light in humans and other mammalian species. These studies led to the discovery of intrinsically photosensitive retinal ganglion cells (ipRGCs) that contain a photopigment named melanopsin. The ipRGCs interconnect with the classical visual rod and cone photoreceptors. Together, all retinal photoreceptors provide input to the retinohypothalamic tract (RHT). The RHT transmits information about environmental light to the central circadian pacemaker as well as many other nonvisual centers in the nervous system. This chapter reviews the fundamental

G.C. Brainard (✉) • J.P. Hanifin
Department of Neurology, Thomas Jefferson University, Philadelphia, PA, USA
e-mail: george.brainard@jefferson.edu

neurophysiology, the clinical and nonclinical therapeutic uses of light, as well as selected examples of published data on the effects of solid state light on human biology and behavior. Both the basic and applied science related to these discoveries are in a nascent stage. As new lighting technologies and applications are developed with the intent to improve human health and well-being, empirical evidence is critically needed to ensure the safety and efficacy of these advances. Collaboration between scientists and engineers across the fields of physics, biomedicine, lighting, and architecture will guide the best use of light for the benefit of humanity.

Introduction

There are two convergent developments which will ultimately transform architectural lighting: (1) the advance of solid state technologies for indoor and outdoor illumination and (2) the confirmation that light regulates human circadian, neuroendocrine, and neurobehavioral physiology, thereby influencing health and well-being. The aim of this chapter is to review the empirical basis of the biological, behavioral, and therapeutic effects of light with selected examples of published data from studies on light emitted from solid state luminaires.

Traditionally, the principal goals of architectural lighting have been to provide light that (1) is optimum for visual performance, (2) is visually comfortable, (3) permits aesthetic appreciation of the space, and (4) conserves energy (DiLaura et al. 2011). Over the past three decades, empirical evidence has demonstrated that light stimuli received by the human eye can have potent biological, behavioral, and therapeutic effects that are relatively separate from the defined physiology for visual support (Lam and Levitt 1999; JBR 2005; Lucas et al. 2014). The professional communities of lighting designers, manufacturers, and architectural engineers have opened the door to the development of appropriate applications that might result from these emergent neurophysiological discoveries (CIE 2004; IESNA 2008; DiLaura et al. 2011). Ultimately, lighting based on the classical design objectives will be evolved to integrate the recent discoveries about the role of light in human health and well-being with what is known about lighting factors for supporting vision and environment.

Circadian, Neuroendocrine, and Neurobehavioral Regulation by Light Wavelength

Melatonin, a hormone secreted by the pineal gland in humans and other mammals, has roles in regulating circadian, neuroendocrine, and photoperiodic physiology (JBR 2005). Environmental light stimuli entrain the circadian rhythm of melatonin

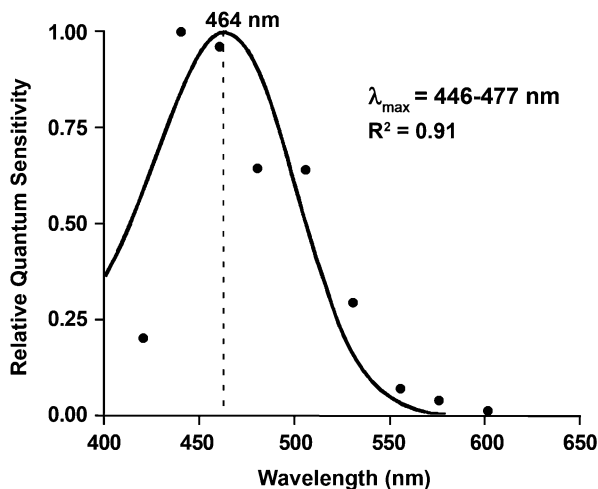


Fig. 1 This graph illustrates the action spectrum for percent control-adjusted melatonin suppression in 72 healthy human subjects. The filled circles represent the half-saturation constants of nine wavelengths from 420 to 600 nm that were normalized to the maximum response and plotted as log relative sensitivity. The solid curve portrays the best-fit template for vitamin A₁ retinaldehyde photopigments, which predicts a calculated maximal spectral absorbance (λ_{\max}) of 464 nm, with a λ_{\max} range from 446 to 477 nm \pm 1 SD. There is a high coefficient of correlation ($R^2 = 0.91$) for fitting this opsin template to the melatonin suppression data (Brainard et al. 2001b, 2008)

production as well as acutely suppress nocturnal melatonin secretion (Lewy et al. 1980; Czeisler et al. 1986; Burke et al. 2013; Wright et al. 2013). Melatonin studies have led to a major shift in the understanding of photoreceptive input to the circadian and neuroendocrine systems of humans. Specifically, in 2001, a study confirmed that the three-cone system that mediates human photopic vision is not the primary photoreceptor system that transduces light stimuli for acute melatonin suppression in healthy human subjects (Brainard et al. 2001a). Later that year, two analytical action spectra in healthy human subjects identified 446 nm to 477 nm as the most potent wavelength region for melatonin regulation (Brainard et al. 2001b; Thapan et al. 2001). Consistent with earlier studies on rodents, those action spectra strongly indicated that the human eye contains a novel photosensory system that is distinct from the visual rods and cones and is responsible for regulating human physiology (Foster and Hankins 2007). Figure 1 illustrates data from one of these action spectra (Brainard et al. 2001b, 2008). To date, ten analytic action spectra and many investigations based on selected wavelength comparisons of such responses in humans, nonhuman primates, and rodents support a peak sensitivity in the short-wavelength portion of the visible spectrum from 447 to 484 nm distinctly divergent from that predicted by the standard observer (V_{λ}) that has a peak sensitivity of 555 nm (Brainard and Hanifin 2005; Lucas et al. 2014, for reviews).

Phototransduction for Circadian, Neuroendocrine, and Neurobehavioral Regulation

Extensive animal and human studies have elucidated the neuroanatomy and neurophysiology of the photosensory system that provides input for circadian neuroendocrine and neurobehavioral regulation (Lucas et al. 2014). A photopigment named melanopsin has been localized both in the retinas of rodents and humans (Provencio et al. 2000). Melanopsin is found in a subtype of intrinsically photoreceptive retinal ganglion cells or ipRGCs (Berson et al. 2002; Hattar et al. 2002; Gooley et al. 2003). Using immunohistochemistry and double labeling studies in mice, ipRGC anatomy has been described (Ecker et al. 2010) and five types of ipRGCs, which have differing locations, dendritic processes, and cell bodies, have been identified.

In humans, these ipRGCs are found within the inner retina and have an extensive dendritic tree that forms a photoreceptive network (Rollag et al. 2003). IpRGCs comprise approximately 0.2 % to 0.8 % of all ganglion cells present in the human retina (1–3 % in rodents), and their dispersion of dendritic processes seems to encompass the entire retinal area as observed in earlier rodent studies (Hankins et al. 2008). These seminal discoveries, and further clarification of the biochemistry, anatomy, and physiology of melanopsin and the ipRGCs, have been a crowning achievement in neuroscience (Lucas et al. 2014).

Abundant evidence shows that the melanopsin ipRGCs are anatomically and functionally interconnected with the rods and cones that support vision. Light-induced physiological responses reflect input from all of the retinal photoreceptor classes, with the relative importance of each being highly variable within and between response types. This complex interconnection is represented clearly in Fig. 2 from Lucas and colleagues (2014).

Neural Pathways for Circadian and Neurophysiological Regulation

Projections from the ipRGCs form the origin of the retinohypothalamic tract (RHT). The mammalian RHT is the primary neural projection to the circadian oscillator for entraining circadian rhythms to environmental light–dark cycles (Klein et al. 1991). This pathway acts to convey information about external light conditions from the retina to several areas of the hypothalamus, including the suprachiasmatic nuclei (SCN), the primary site of the biological clock, or central timekeeper in the brain. The RHT terminates primarily in the ventrolateral aspect of the SCN. In turn, the SCN transmit information about lighting and circadian time to a diversity of major control regions of the nervous system including the pineal gland where the hormone melatonin is synthesized. Thus, the RHT has been studied extensively for its role in synchronizing the endogenous oscillator in the SCN with environmental light cues and mediating systemic circadian physiology (Klein et al. 1991; Gooley et al. 2003; Hattar et al. 2006).

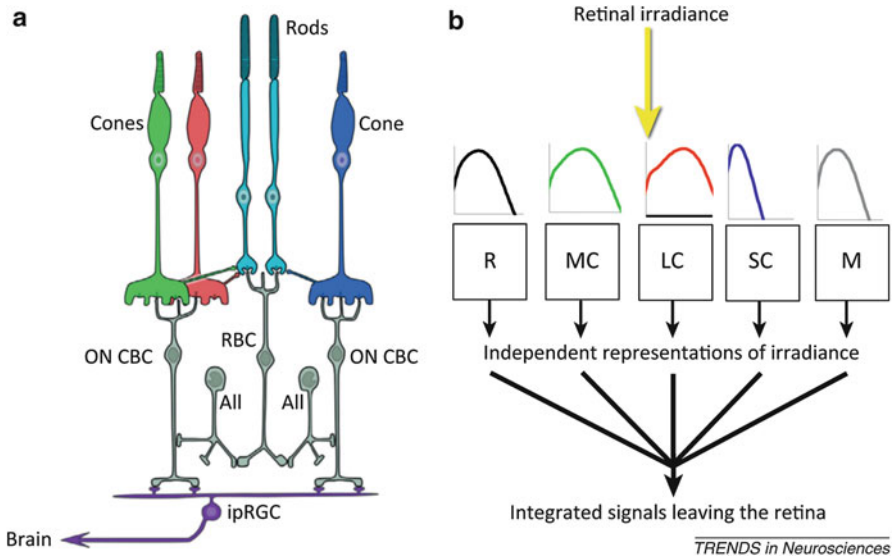


Fig. 2 (a) Schematic of the relevant retinal circuitry in humans. On cone bipolar cells (on *CBCs*), amacrine cells (*All*), and rod bipolar cells (*RBC*); (b) human photoreceptors R (rod opsin), M (melanopsin), SC (S-cone opsin), MC (m-cone opsin), LC (l-cone opsin) with plots of log sensitivity against wavelength to generate a distinct measure of illuminance (Lucas et al. 2014) (Figure reprinted with permission from Elsevier)

Although the RHT has its densest projection in or around the SCN, this pathway has diffuse projections to other areas including the preoptic nuclei, anterior and lateral hypothalamic areas, retrochiasmatic areas, dorsal hypothalamic nuclei, the intergeniculate leaflets, and the midbrain periaqueductal gray (Klein et al. 1991; Hattar et al. 2006). Some of the functional roles of the other nonvisual projections from the retina have been elucidated. The ventrolateral preoptic nucleus (VLPO) is known to be integral in sleep/arousal state. Additionally, projections to the intergeniculate leaflet (IGL) from the retina are involved in regulation of circadian phase shifting and other integration of photoperiodic information. Projections to the ventral subparaventricular zone (vSPZ) are thought to be involved in the circadian and photic modulation of sleep and locomotor activity. The pretectal area (PTA) receives projections from the RHT and which contributes to the control of the pupillary light. In summary, these projections, which are relatively separate from areas of the brain that are involved in forming vision, are thought to form an irradiance detection system providing photic information to several brain regions controlling numerous functions separate from areas of the brain that are involved in forming vision (Gooley et al. 2003). Figure 3 provides a simplified illustration of the neural anatomy that supports vision and circadian, neuroendocrine, and neurobehavioral responses (Gooley et al. 2003; Brainard and Hanifin 2014).

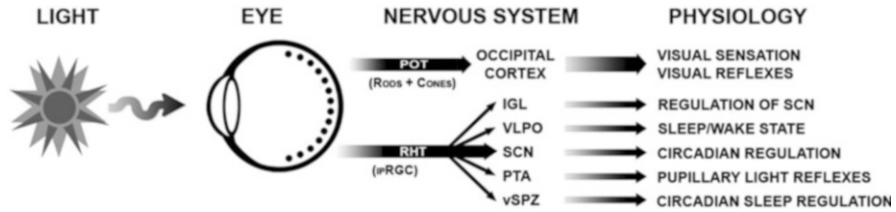


Fig. 3 The diagram above is a highly simplified schematic of the neuroanatomy responsible for mediating both the sensory capacity of the visual system and the nonvisual regulation of circadian, neuroendocrine, and neurobehavioral functions. Abbreviations: *POT* primary optic tract, *RHT* retinohypothalamic tract, *ipRGCs* intrinsically photosensitive retinal ganglion cells, *IGL* intergeniculate leaflets, *VLPO* ventrolateral preoptic nuclei, *SCN* suprachiasmatic nuclei, *PTA* pretectal areas, *vSPZ* ventral subparaventricular zones (This figure is modeled after an illustration in Gooley et al. (2003) and adapted from a diagram in Brainard and Hanifin (2014))

Circadian, Neuroendocrine, and Neurobehavioral Regulation by Light Intensity

The human visual system is exquisitely sensitive to light. Photopic daytime vision detects light down to about 40 lx of indoor light, mesopic vision operates in the 12 to 0.004 lx range, and scotopic vision can detect light as low as 0.000004 lx (DiLaura et al. 2011). In contrast, considerably higher levels of illumination are required to evoke circadian, neuroendocrine, and neurobehavioral responses in humans. The earliest demonstrations of light eliciting melatonin suppression required illuminances of 2,500 lx, and circadian phase shifts required 7,000 to 12,000 lx in healthy humans (Lewy et al. 1980; Czeisler et al. 1986). Those early studies employed broad spectrum white light as the experimental stimulus. Later studies that also used broad spectrum white light showed that when exposure is carefully controlled in a laboratory setting, illuminances as low as 100 lx of white light can suppress nocturnal melatonin, and 119 lx of white light can phase shift melatonin rhythms in humans (Brainard et al. 1997; Zeitzer et al. 2000). Further, studies employing 460 nm monochromatic light in controlled laboratory exposures showed that as little as 1.3 lx can suppress melatonin and 5 lx can phase shift the melatonin rhythm and elicit alerting responses (Brainard et al. 2001b; Lockley et al. 2003, 2006).

A recent within-subjects study with healthy men and women was the first to characterize a full-range dose–response curve for melatonin suppression using solid state light. That experiment employed a 122×122 cm² panel of 5,776 light-emitting diodes (LEDs) with a lens diffuser that emitted a relatively even distribution of polychromatic but narrow bandwidth blue-appearing light (λ_{\max} 469 nm, half peak bandwidth of 26 nm). Irradiance emitted from the light panel was adjusted using a rheostat, at times combined with the use of acrylic neutral density filter panels fitted in front of the diffusing panel. Intensities studied included eight different corneal irradiances ranging from 0.1 to 600 $\mu\text{W}/\text{cm}^2$ or 0.09 to 562 lx (West et al. 2011).

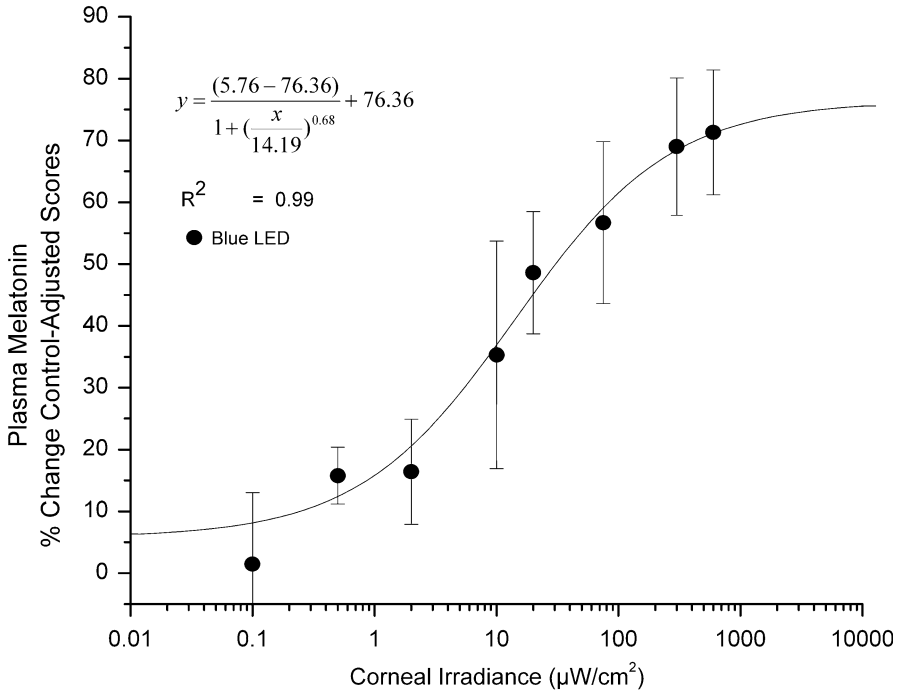


Fig. 4 This graph presents the fluence–response curve derived from the plasma melatonin percent change control-adjusted scores from eight healthy men and women relative to corneal irradiances from the blue LED exposure panel. Each data point represents one group mean \pm SEM. The formula for the sigmoidal curve fit is inset at the top left of the graph along with coefficient of correlation for the fitted line to the empirical data. This figure first appeared in West et al. (2011)

Subjects came in one night a week, with at least 1 week between study nights. Volunteers arrived at the laboratory by 11:45 PM on each study night, and blindfolds were placed over subjects’ eyes at midnight. Subjects then remained awake and seated in a constant upright posture. At 2:00 AM, while still blindfolded, a blood sample was taken by venipuncture. The blindfolds were then removed, and the subjects began a 90-min light exposure, during which they remained seated upright and with fixed gaze at a specified point on the light panel at a distance of 35 cm, ensuring a full retinal field exposure. Volunteers’ pupils were freely reactive during the light exposures. At 3:30 AM, a second blood sample was taken while subjects continued to gaze at the light. On control nights, the 90-min light exposure was replaced with continued dark exposure, and both blood draws were taken while the subject was blindfolded (West et al. 2011). As shown in Fig. 4, the results demonstrate that increasing irradiances of narrowband blue-appearing LED light elicit increasing plasma melatonin suppression in healthy subjects ($P < 0.0001$).

Figure 4 shows the first published full-range fluence–response curve for a human biological response to LED light. These data illustrate that it is feasible to quantify how a specific solid state luminaire impacts human physiology. Specifically, the

calculated half-saturation constant for this LED luminaire was $14.1 \mu\text{W}/\text{cm}^2$ or 13.2 lx. This photobiological fluence–response relationship illustrates the potential for light to effect changes in human biology in a similar fashion to drug or pharmaceutical agents. Classically pharmaceuticals elicit sigmoidal dose–response curves similar to the curve shown in Fig. 4 – the higher the drug dose, the greater the physiological change, often with saturating limits.

It is important to note that for the data illustrated in Fig. 4, the light exposure technique involved constant 90 min, full retinal field exposures. Such exposures are optimum for controlled laboratory studies that permit comparisons between different light sources (West et al. 2011; Brainard et al. 2015). When used in daily applications for environments such as homes, schools, hospital, and workplaces, however, exposure conditions rarely would be full retinal field or continuous. Studies in a variety of architecturally illuminated spaces are needed to define the optimum irradiances and illuminances that support vision as well as the biological and behavioral efficacy of built-in lighting.

Ocular and Physiological Elements that Mediate the Biological Effects of Light

The basic science of visual psychophysics and its application to electrical lighting design and architecture has matured over more than a century (DiLaura et al. 2011). With the more recent understanding that light regulates circadian, neuroendocrine, and neurobehavioral responses in humans, similar psychophysical studies need to be carried out to elucidate the unique effects of light on biological and behavioral systems. There are two general categories of physical and biological elements involved in light regulation of these photic effects: (1) physical/biological stimulus processing and (2) sensory/neural signal processing. Specifically, the elements for physical and biological stimulus processing involve the light source physics, conscious and reflex behavior relative to the light source, and the transduction of light to the retina through the pupil and ocular media. In turn, sensory/neural signal processing is initiated as photons are absorbed by retinal photopigments and neural signals are generated in the RHT. Factors influencing this physiology include (1) the wavelength sensitivity of the operative photoreceptors, (2) the distribution of the operative photoreceptors, (3) the state of photoreceptor adaptation, and (4) the ability of the central nervous system to integrate photic stimuli temporally and spatially. Each of these elements determines the effectiveness of an environmental photic stimulus for regulating the circadian, neuroendocrine, and neurobehavioral systems. A number of laboratories have worked on clarifying specific ocular and neural elements that mediate the biological and behavioral effects of light, but this is still an emergent science especially in determining the interdependence and variability of these elements (Brainard et al. 1997; Brainard and Hanifin 2005; Lucas et al. 2014 for reviews).

It is beyond the scope of this chapter to fully describe the literature on psychophysics related to the circadian, neuroendocrine, and neurobehavioral effects of

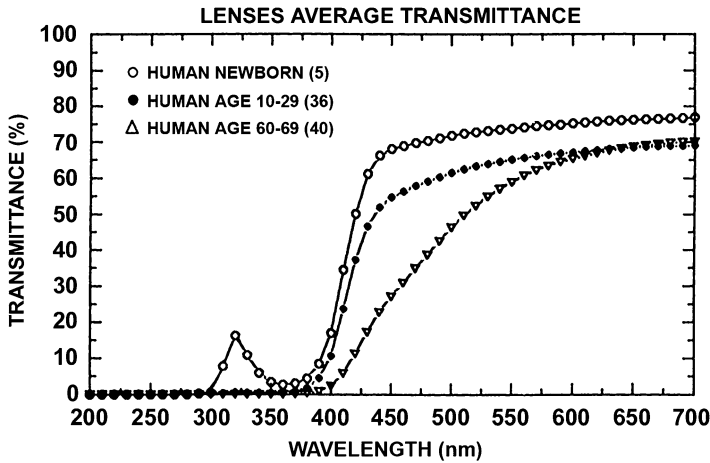


Fig. 5 The mean transmittance of visible and ultraviolet wavelengths of postmortem human lenses from newborn donors, donors aged 20 to 29, and donors aged 60 to 69 (Reprinted from Brainard et al. 1997 with permission)

light, but the example of human lens transmittance can illustrate the relevance of this basic science to lighting applications. In human eyes, the cornea, aqueous humor, and vitreous humor are clear tissues that transmit nearly 100 % of visible and ultraviolet wavelengths down to 300 nm to the retina with little age-related change in the transmission characteristics. By contrast, the crystalline lens in the human eye develops a yellow pigmentation with age that acts as a filter that significantly reduces the total transmission of radiant energy to the retina, particularly in the shorter wavelength portion of the spectrum (Pokorny et al. 1987; Brainard et al. 1997). The data illustrated in Fig. 5 show human lenticular transmission from three different age groups. Transmission of longer wavelength visible light is not substantially different between these three groups. The lenses of newborn children and young adults transmit some ultraviolet radiation to the retina. In contrast, the lenses of adult humans aged 60–69 do not transmit ultraviolet radiation to the human retina, and there is a significant reduction of wavelengths transmitted in the violet, indigo, blue, and green portions of the visible spectrum. Hence, with aging, the ocular lens modifies the total quantity of light as well as the balance of wavelengths that reach the retinal photoreceptors. Such changes have the potential for a significant impact on systemic neurophysiological processes.

It is clear that age-related changes in lens transmission can have significant impact on human vision (DiLaura et al. 2011). Given the high sensitivity of the ipRGCs to short-wavelength visible light, the age-related changes in lens transmission appears also to impact the neurophysiological effects of light. A controlled laboratory study compared melatonin suppression in young versus older women (mean ages 24 versus 57 years) with exposure to short-wavelength monochromatic light at 456 nm. The results demonstrated a significant reduction of melatonin suppression in the older subjects, which the authors interpreted to be a consequence

of changes in lens transmission (Herljevic et al. 2005). Similarly, an epidemiological study on 970 randomly selected individuals aged 30–60 years showed a significant increase in sleep disturbances associated with decreased transmission of short-wavelength light in the lenses of older adults. The authors concluded that filtration of blue light by the aging lens resulted in disturbance of circadian rhythm photoentrainment increasing risk of sleep disturbances (Kessel et al. 2011). In terms of lighting applications, this suggests that lighting levels need to be specified relative to the age of the occupants in the designed environment to optimize both vision and circadian, neuroendocrine, and neurobehavioral regulation. Simply put, the quantity and spectral quality of light that is best for an elementary school will be different than the lighting needed for a retirement community.

Clinical and Nonclinical Applications of Solid State Light Therapy

Following the discovery that bright white light exposure at 2,500 lx suppressed melatonin secretion in healthy humans, researchers quickly determined that light could be used therapeutically to treat seasonal affective disorder (SAD or winter depression) and to phase shift human circadian rhythms (Lewy et al. 1980; Rosenthal et al. 1984; Czeisler et al. 1986). Since then, light therapy has proven to be an effective therapeutic intervention for SAD patients and its subclinical variant, sSAD. A variety of light treatment devices have been tested for treating these affective disorders, including light boxes, dawn simulators, and head-mounted delivery systems (e.g., light visors and light masks). The standard practice has been for patients to try a daily trial of 10,000 lx white light for 30–60 min in the morning upon awakening (Lam and Levitt 1999; Golden et al. 2005). As with the treatment of many medical disorders, patients vary in their responsiveness to light therapy. Although a majority of clinical trials employing light therapy have been concerned with the treatment of SAD, additional clinical applications have been explored including light treatment of nonseasonal depression, various sleep disorders, menstrual cycle problems, bulimia nervosa, and problems associated with senile dementia (Lam and Levitt 1999; Golden et al. 2005; Brainard and Hanifin 2014). With the advent of solid state technology, light therapy devices are now being produced with both broad and narrow bandwidths of light emitted by LEDs. This advance in solid state technology has enabled light therapy equipment to be produced in conventionally sized light therapy panels as well as relatively small, portable devices. This has also facilitated the design of clinical trials that test different narrow bandwidth wavelengths of light in the treatment of SAD.

In one such study, prototype light panels with arrays of LEDs were tested in a randomized, blinded phase I clinical trial for efficacy in treating SAD (Glickman et al. 2006). The LED panels emitted either brighter narrow-band blue light (468 nm, at 607 $\mu\text{W}/\text{cm}^2$, or about 400 lx) or dimmer narrow-band red light (654 nm, at 34 $\mu\text{W}/\text{cm}^2$, or about 25 lx). The red light condition was an intended placebo light, but patients were blind to this, not knowing which color light exposure was the intended treatment. SAD patients completed a 3-week outpatient treatment with light

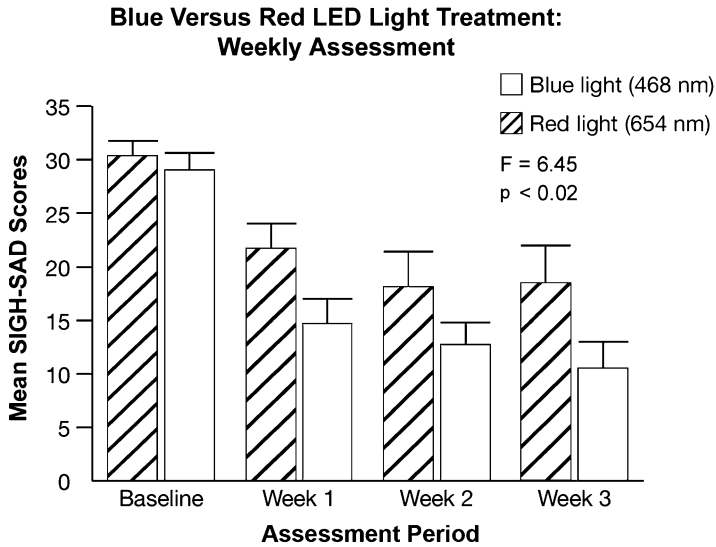


Fig. 6 This figure shows a comparison between mean (+SEM) SIGH-SAD scores at each week in patients who were treated with the 468 nm blue light panel versus those treated with the 654 nm dim red light placebo device. While there was no difference between baseline scores across conditions, SIGH-SAD scores of patients treated with the blue LED light were significantly lower than SIGH-SAD scores of patients treated with the blue LED light ($p < 0.02$)

exposure scheduled for 45 min daily between 6:00 and 8:00 AM. They were instructed to sit directly in front of the light panel approximately 50 cm from their eyes and to glance at the light for a few seconds every minute. SIGH-SAD scores were assessed weekly for each patient by the same rater, who was blind to the subject's condition.

Study results shown in Fig. 6 revealed that symptom improvement was significantly better in the group treated with the blue LED light compared to those treated with the red LED light. Further, the remission rates of the patients treated with the blue LED panel were comparable to the remission rates typically reported in patients utilizing current standard bright light treatment, even though the blue light panel was at a much lower intensity/illuminance (Glickman et al. 2006).

The majority of studies on light therapy have been concerned with treating SAD or other clinical disorders. In addition, light therapy has been evaluated for healthy individuals who experience problems associated with intercontinental jet travel, shift work, and space flight (Lam and Levitt 1999; JBR 2005; IESNA 2008; Barger et al. 2014; Lucas et al. 2014; Brainard and Hanifin 2014). Ongoing work is testing light emitted by LED luminaires for supporting vision and assessing neuroendocrine, circadian, neurobehavioral, and sleep effects in both NASA ground crew and astronauts in flight (Barger et al. 2012; Brainard et al. 2013). In terms of emergent application of light therapy to manned space flight, solid state lighting assemblies (SSLAs) composed of LEDs will be replacing the general luminaire assemblies,

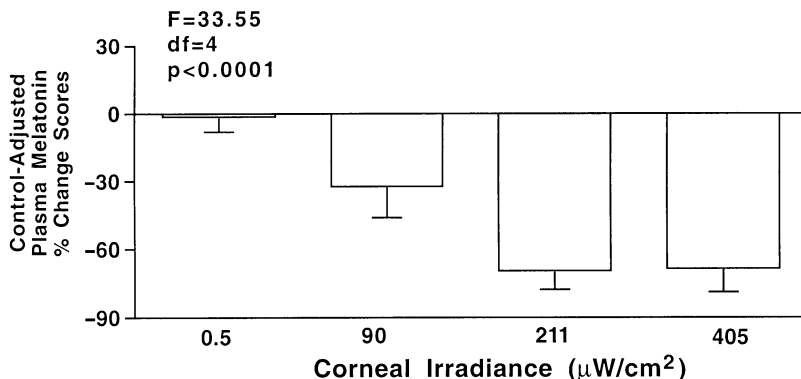


Fig. 7 This graph illustrates percent control-adjusted melatonin suppression in eight healthy male and female subjects exposed to four different irradiances of white 4,800 K light emitted by a solid state light source being developed for installation of the ISS. The range of 0.5–405 $\mu\text{W}/\text{cm}^2$ irradiances correspond to illuminances of 1.3–1270 lx (Brainard et al. 2013)

which contain fluorescent lights on the International Space Station (ISS). The advantages of LEDs over conventional fluorescent light sources related to space flight include lower up-mass, power consumption, and heat generation, as well as fewer toxic materials, greater resistance to damage, and long lamp life (NASA 2011). An initial ground-based study was performed testing acute plasma melatonin suppression as well as visual performance and color discrimination in cohorts of healthy, human subjects under different SSLA light exposure conditions within a high-fidelity replica of the ISS Crew Quarters (Brainard et al. 2013).

Figure 7 shows the percent control-adjusted melatonin change scores for healthy male and female volunteers exposed to four different intensities of light emitted by a prototype SSLA that was developed for ISS. Analysis of variance showed that there was a significant effect of light irradiance on melatonin suppression ($F = 33.55$, $df = 4$, $p < 0.0001$). Increasing corneal irradiances of 4,800 K SSLA light exposure evoked progressively larger melatonin suppressions. Specifically, the Fisher PLSD test showed that the highest corneal irradiances of 4,800 K SSLA light (405 and 211 $\mu\text{W}/\text{cm}^2$) elicited significantly stronger melatonin suppression than the lower irradiances (90 and 0.5 $\mu\text{W}/\text{cm}^2$). The corneal irradiance 90 $\mu\text{W}/\text{cm}^2$ of 4,800 K SSLA light elicited significantly stronger melatonin suppression than the lowest irradiance (0.5 $\mu\text{W}/\text{cm}^2$).

In addition, visual tests were done under indirect daylight at 201 lx, fluorescent room light at 531 lx, and 4,800 K SSLA light at 1266 lx inside the replica of the ISS crew quarters. Visual performance was assessed with numerical verification tests (NVT). NVT data show that there are no significant differences in score ($F = 0.73$, $p = 0.48$) or time ($F = 0.14$, $p = 0.87$) for subjects performing five contrast tests (10 to 100 %). Color discrimination was assessed with Farnsworth-Munsell 100 hue tests (FM-100). The FM-100 data showed no significant differences ($F = 0.01$, $p = 0.99$) in color discrimination for indirect daylight, fluorescent room light, and

4800 K SSLA light in the crew quarters (Brainard et al. 2013). Risk factors for the health and safety of astronauts include disturbed circadian rhythms and altered sleep–wake patterns (Whitmire et al. 2010; Barger et al. 2014). These studies will help determine how SSLA lighting can be used both to support astronaut vision and serve as an in-flight countermeasure for circadian desynchrony, sleep disruption, and cognitive performance deficits in astronauts on the ISS.

Light Measurement for Biological and Behavioral Regulation

Initial human studies examining the effects of light on melatonin suppression, circadian phase shifting, and winter depression measured light levels in lux, a unit based on the responsiveness of the human visual system (DiLaura et al. 2011). This implies that these responses are mediated by the three-cone photopic visual system. Numerous human studies show that this assumption is not true (Brainard and Hanifin 2005; Lucas et al. 2014, for reviews). Several groups have developed light measurement models based on data from their own studies or analysis of other published data (Lucas et al. 2014). These models have sought to develop functions to quantify the biological and behavioral effects of light mediated by melanopsin within the ipRGCs in terms of a single measurement unit. IpRGCs not only are capable of photon detection independently, they also receive input from classic cone and rod photoreceptors. As illustrated in Fig. 2, there is anatomical and physiological connectivity of the classic photoreceptors and melanopsin containing ipRGCs. This morphological and functional connectivity makes it challenging to identify the specific contribution of each photoreceptor type. Light-induced physiological responses reflect input from all of the retinal photoreceptor classes, with the relative importance of each being highly labile within and between response types. For example, a study on healthy human subjects showed that the relative contributions of cone photoreceptors to melatonin suppression and circadian phase shifting varied across different light intensities (Gooley et al. 2010).

It has been problematic that there has not been consistency in the methods used for quantifying light between different laboratories studying photic regulation of biological and behavioral responses. Often, this can make it challenging to replicate experimental conditions or to compare data across studies. Recently, a consensus position was developed by many of the laboratories that have studied wavelength regulation of the biological and behavioral effects of light in rodents, humans, and other species. That consensus defines the best practices for measuring and reporting experimental light stimuli (Lucas et al. 2014). In addition, a freely available web-based toolbox was provided that permits calculation of the effective irradiance experienced by each of the ipRGC, cone, and rod photoreceptors that, in turn, drive circadian, neuroendocrine, and neurobehavioral effects. Table 1 provides an example of human retinal photopigment-weighted illuminances calculated by the toolbox for two LED light sources.

The light sources in Table 1 were used in studies discussed above (West et al. 2011; Brainard et al. 2013). For this analysis, the LED spectral power

Table 1 Calculated irradiances, photopic illuminances $v(\lambda)$, and human photopigment illuminances relative to one broad spectrum and one narrowband LED light source

	Radiometric and photometric values (380–780 nm inclusive)			Retinal photopigment-weighted illuminances (α -opic lux)				
	Photon flux (cm^2/s)	Irradiance ($\mu\text{W}/\text{cm}^2$)	Photopic illuminance (lux)	S Cone	Melanopsin ipRGC	Rod	M Cone	L Cone
4,800 K broadband LED	$9.16\text{E} + 14$	324	1,019	686	815	861	963	992
Narrowband 469 nm LED	$9.11\text{E} + 14$	376	435	1,816	2,723	2,023	1,118	618

distributions were measured at the photon flux shown in the first column of Table 1. Currently, calculation of the biological potency of light for nonvisual photoreception with a single metric remains elusive. Characterizing and collecting experimental light source descriptions on a single platform will facilitate comparisons of study results from different laboratories. Ultimately, the generation and compilation of such data will permit the development of testable hypotheses that predict physiological responses to LED and other light sources.

Conclusion

Light is miraculous. It provides the basis of the food chain with the absorption of photons by organic molecules within plants that convert the energy of light into chemical energy in sugars. These sugars are the primary fuel for life. Light enables us to visualize the world in exquisite detail bringing us awareness of brightness, color, shape, motion, and image. As with the food chain, vision begins with the absorption of photons by organic molecules in the retinal photoreceptors that ultimately trigger a cascade of neural events throughout the nervous system. Through visual physiology, light can dominate our awareness. Consequently, philosophers and scientists from Plato to Goethe and from Newton to Einstein enthusiastically explored the physics of light and its relationship to vision and the human eye (Zajonc 1993).

Below consciousness, light powerfully controls human hormones, levels of alertness, and seasonal and daily rhythms of physiology. Important steps have been made to craft this capacity of light into strategies for treating circadian disruption, remedying affective disorders, improving astronaut health during space travel, as well as promoting the health and well-being of occupants in buildings such as schools, healthcare facilities, work environments, and homes. Although light therapy for treating winter depression has been accepted into modern medicine, many of the clinical and nonclinical applications of light therapy are presently in a nascent stage. As new lighting technologies and applications are developed with the intent to improve human health and well-being, empirical evidence is critically needed to ensure the safety and efficacy of these advances. Collaboration between scientists and engineers across the fields of physics, biomedicine, lighting, and architecture will guide the best use light for the benefit of humanity.

Acknowledgments The authors gratefully acknowledge the dedicated support of Samar Jasser M.D. for the editorial review and of Benjamin Warfield for formatting all figures and the creative development of Fig. 3. Figure 3 and portions of this manuscript were adapted and updated from an earlier publication (Brainard and Hanifin 2014) with permission from the Commission Internationale de l'Eclairage (CIE). Figures 2, 4, and 5 were reprinted with permission from Elsevier publications (Lucas et al. 2014; Glickman et al. 2006; Brainard et al. 2013, respectively). Figure 6 was originally published in a book chapter of Brainard et al. 1994, *Advances in Pineal Research: 8*, M Møller and P Pévet, eds, pp 415–432, John Libbey & Company Ltd., London and later in Brainard et al. 1997 cited here. The publishers have given permission for it to be reprinted here. The work was supported, in part, by grants from the Smart Lighting ERC under NSF

EEC-0812056; NSBRI under NASA Cooperative Agreement NCC 9–58; NIH ROINS36590; NIMH1R43, Apollo Health, Philips Healthcare, The Institute for Integrative Health, and the Philadelphia Section of the Illuminating Engineering Society.

References

- (2005) Special issue: human circadian rhythms: regulation and impact. *J Biol Rhythms* 20:279–386
- Barger LK, Sullivan JP, Vincent AS, Fiedler ER, McKenna LM, Flynn-Evans EE, Gilliland K, Sipes WE, Smith PH, Brainard GC, Lockley SW (2012) Learning to live on a Mars day: fatigue countermeasures during the Phoenix Mars Lander mission. *Sleep* 35:1423–1435
- Barger LK, Flynn-Evans EE, Kubey A, Walsh L, Ronda JM, Wang W, Wright KP, Czeisler CA (2014) Prevalence of sleep deficiency and use of hypnotic drugs in astronauts before, during, and after spaceflight: an observational study. *Lancet Neurol* 13:904–912
- Berson DM, Dunn FA, Takao M (2002) Phototransduction by retinal ganglion cells that set the circadian clock. *Science* 295:1070–1073
- Brainard GC, Hanifin JP (2005) Photons, clocks and consciousness. *J Biol Rhythms* 20:314–325
- Brainard GC, Hanifin JP (2014) Exploring the power of light: from photons to human health. In: Proceedings of CIE 2014 lighting quality & energy efficiency, CIE x039:2014, pp 19–31
- Brainard GC, Rollag MD, Hanifin JP (1997) Photic regulation of melatonin in humans: ocular and neural signal transduction. *J Biol Rhythms* 12:537–546
- Brainard GC, Hanifin JP, Rollag MD, Greeson J, Byrne B, Glickman G, Gerner E, Sanford B (2001a) Human melatonin regulation is not mediated by the three cone photopic visual system. *J Clin Endocrinol Metab* 86:433–436
- Brainard GC, Hanifin JP, Greeson JM, Byrne B, Glickman G, Gerner E, Rollag MD (2001b) Action spectrum for melatonin regulation in humans: evidence for a novel circadian photoreceptor. *J Neurosci* 21:6405–6412
- Brainard GC, Sliney D, Hanifin JP, Glickman G, Byrne B, Greeson JM, Jasser S, Gerner E, Rollag MD (2008) Sensitivity of the human circadian system to short wavelength (420 nm) light. *J Biol Rhythms* 23:379–386
- Brainard GC, Coyle W, Ayers M, Kemp J, Warfield B, Maida J, Bowen C, Bernecker C, Lockley SW, Hanifin JP (2013) Solid-state lighting for the International Space Station: tests of visual performance and melatonin regulation. *Acta Astronaut* 92:21–28
- Brainard G, Hanifin J, Warfield B, Stone M, James M, Ayers M, Kubey A, Byrne B, Rollag M (2015) Short wavelength enrichment of polychromatic light enhances human melatonin suppression potency. *J Pineal Res.* doi:10.1111/jpi.12221
- Burke TM, Markwald RR, Chinoy ED, Snider JA, Bessman SC, Jung CM, Wright KP (2013) Combination of light and melatonin time cues for phase advancing the human circadian clock. *Sleep* 36:1617–1624
- Commission Internationale de l’Eclairage (2004) Ocular lighting effects on human physiology and behaviour. Technical Report #158, Commission Internationale de l’Eclairage, Vienna
- Czeisler CA, Allan JS, Strogatz SH, Ronda JM, Sanchez R, Rios CD, Freitag WO, Richardson GS, Kronauer RE (1986) Bright light resets the human circadian pacemaker independent of the timing of the sleep-wake cycle. *Science* 233:667–671
- DiLaura DL, Houser KW, Mistrick RG, Steffy GR (eds) (2011) *Lighting handbook: reference and application*, 10th edn. Illuminating Engineering Society of North America, New York
- Ecker JL, Dumitrescu ON, Wong KY, Alam NM, Chen S, Legates T, Renna JM, Prusky GT, Berson DM, Hattar S (2010) Melanopsin-expressing retinal ganglion-cell photoreceptors: cellular diversity and role in pattern vision. *Neuron* 67:49–60
- Foster RG, Hankins MW (2007) Circadian vision. *Curr Biol* 17:R746–R751

- Glickman G, Byrne B, Pineda C, Hauck WW, Brainard GC (2006) Light therapy for seasonal affective disorder with blue narrow-band light-emitting diodes (LED). *Biol Psychiatry* 59:502–507
- Golden RN, Gaynes BN, Ekstrom RD, Hamer RM, Jacobsen FM, Suppes T, Wisner KL, Nemeroff CB (2005) The efficacy of light therapy in the treatment of mood disorders: a review and meta-analysis of the evidence. *Am J Psychiatry* 162:656–662
- Gooley JJ, Lu J, Fischer D, Saper CB (2003) A broad role for melatonin in nonvisual photoreception. *J Neurosci* 23:7093–7106
- Gooley JJ, Rajaratnam SM, Brainard GC, Kronauer RE, Czeisler CA, Lockley SW (2010) Spectral responses of the human circadian system depend on the irradiance and duration of exposure to light. *Sci Transl Med* 2:31ra33
- Hankins MW, Peirson SN, Foster RG (2008) Melanopsin: an exciting photopigment. *Trends Neurosci* 31:27–36
- Hattar S, Liao H-W, Takao M, Berson DM, Yau K-W (2002) Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* 295:1065–1070
- Hattar S, Kumar M, Park A, Tong P, Tung J, Yau K-W, Berson DM (2006) Central projections of melanopsin-expressing retinal ganglion cells in the mouse. *J Comp Neurol* 497:326–349
- Herljevic M, Middleton B, Thapan K, Skene DJ (2005) Light-induced melatonin suppression: age-related reduction in response to short wavelength light. *Exp Gerontol* 40:237–242
- Illuminating Engineering Society of North America (2008) Light and human health: an overview of the impact of optical radiation on visual, circadian, neuroendocrine, and neurobehavioral responses, IES TM-18-08. Illuminating Engineering Society of North America, New York
- Kessel L, Siganos G, Jorgensen T, Larsen M (2011) Sleep disturbances are related to decreased transmission of blue light to the retina caused by lens yellowing. *Sleep* 34:1215–1219
- Klein DC, Moore RY, Reppert SM (eds) (1991) *Suprachiasmatic nucleus: the mind's clock*. Oxford University Press, Oxford
- Lam RW, Levitt AJ (eds) (1999) *Canadian consensus guidelines for the treatment of seasonal affective disorder*. Clinical and Academic Publishing, Vancouver
- Lewy AJ, Wehr TA, Goodwin FK, Newsome DA, Markey SP (1980) Light suppresses melatonin secretion in humans. *Science* 210:1267–1269
- Lockley SW, Brainard GC, Czeisler CA (2003) High sensitivity of the human circadian melatonin rhythm to resetting by short wavelength light. *J Clin Endocrinol Metab* 88:4502–4505
- Lockley SW, Evans EE, Scheer FA, Brainard GC, Czeisler CA, Aeschbach D (2006) Short-wavelength sensitivity for the direct effects of light on alertness, vigilance, and the waking electroencephalogram in humans. *Sleep* 29:161–168
- Lucas RJ, Peirson SN, Berson DM, Brown TM, Cooper HM, Czeisler CA, Figueiro MG, Gamlin PD, Lockley SW, O'Hagan JB, Price LLA, Provencio I, Skene DJ, Brainard GC (2014) Measuring and using light in the melanopsin age. *Trends Neurosci* 37:1–9
- National Aeronautics and Space Administration (2011) *ISS Interior Solid State Lighting Assembly (SSLA) Specification, Revision A, July 2011, S684-13489*, Johnson Space Center, Houston, pp 1–60
- Pokorny J, Smith VC, Lutze M (1987) Aging of the human lens. *Appl Optics* 26:1437–1440
- Provencio I, Rodriguez IR, Jiang G, Hayes WP, Moreira EF, Rollag MD (2000) A novel human opsin in the inner retina. *J Neurosci* 20:600–605
- Rollag MD, Berson DM, Provencio I (2003) Melanopsin, ganglion-cell photoreceptors, and mammalian photoentrainment. *J Biol Rhythms* 18:227–234
- Rosenthal NE, Sack DA, Gillin JC, Lewy AJ, Goodwin FK, Davenport Y, Mueller PS, Newsome DA, Wehr TA (1984) Seasonal affective disorder. A description of the syndrome and preliminary findings with light therapy. *Arch Gen Psychiatry* 41:72–80
- Thapan K, Arendt J, Skene DJ (2001) An action spectrum for melatonin suppression: evidence for a novel non-rod, non-cone photoreceptor system in humans. *J Physiol* 535:261–267
- West KE, Jablonski MR, Warfield B, Cecil KS, James M, Thiessen MA, Maida J, Bowen C, Sliney DH, Rollag MD, Hanifin JP, Brainard GC (2011) Blue light from light-emitting diodes (LEDs) elicits a dose-dependent suppression of melatonin in humans. *J Appl Physiol* 110:619–626

- Whitmire AM, Leveton LB, Barger L, Brainard G, Dinges DF, Klerman E, Shea C (2010) Risk of performance errors due to sleep loss, circadian desynchronization, fatigue, and work overload. In: McPhee JC, Charles JB (eds) Human health and performance risks of space exploration missions. NASA/Johnson Space Center, Houston, pp 85–116
- Wright KP, McHill AW, Birks BR, Griffin B, Rusterholz T, Chinoy ED (2013) Entrainment of the human circadian clock to the natural light–dark cycle. *Curr Biol* 23:1554–1558
- Zajonc A (1993) *Catching the light: the Entwined history of light and mind*. Oxford University Press, New York
- Zeitler JM, Dijk D-J, Kronauer RE, Brown EN, Czeisler CA (2000) Sensitivity of the human circadian pacemaker to nocturnal light: melatonin phase resetting and suppression. *J Physiol* 526:695–702

Lighting and the Elderly

Eunice Noell-Waggoner

Contents

Lighting for Older Adults	848
Lighting Needs of the Aging Eye: Quantity and Quality	849
Strategies to Design Appropriate Lighting for the Elderly	851
Area Specific Lighting	852
Daylight: The Devil Is in the Details	853
Senior Care Facilities	856
Alzheimer’s Disease and Other Dementia: Impact on Vision	858
Circadian Rhythm Support	858
Importance of Vitamin D ₃	860
Call to Action	861
References	861

Abstract

Good lighting is perhaps the most important, and least understood, design element required to provide supportive environments for all older adults. It is essential to maximize independence (abilities), quality of life, health, wellness and safety. It is critical that homes and public buildings, especially hospitality, medical, and care facilities address not only the impact of normal age-related changes to vision, but also plus the added disability of eye diseases for some, and the important role that light and the visual environment plays in the lives of older people. As people age, they become more dependent on their environment to compensate for their sensory loss, increasing frailty and reduced mobility.

The special lighting needs for older adults are not limited to vision, but also include the biological effects of light on personal health. The non-visual or photobiological effects of light include both light entering the eyes, which

E. Noell-Waggoner (✉)
Center of Design for an Aging Society, Portland, OR, USA
e-mail: eunice@centerofdesign.org

impacts circadian rhythm (sleep/wake cycle), and light falling on the skin (vitamin D synthesis so that calcium can be absorbed by bones and tissue). Because of the dramatic growth of the 65+ population, we all need to understand the needs of older adults and provide environments designed to meet their (our) needs.

Lighting for Older Adults

In 2011, the first of the Baby Boomers turned 65. Boomers made up approximately 25 % of the total US population, as of July 2011, and every day 10,000 more Americans turn 65. Unlike their parent's generation, only 11 % of Boomers plan to retire completely at age 65. Some may go for an encore career, and others may volunteer in the community to fulfill their desire to make a difference. You will not find this population group signing up for retirement housing right away. But, they may move to a condo or an apartment where there is less physical demand on them than in a typical single-family home. They will be active in all areas of public and private life for many years. This means that the lighting in all public areas, offices, hospitality/lodging, houses of worship, theaters/assembly spaces, and especially healthcare facilities must serve the vision needs of older adults.

As we all know, aging is a process that brings changes throughout life. Sensory loss is the most common age-related change that we *all* will experience. The normal age-related changes to the visual system, which contribute to inevitable vision impairment, begin long before the age of 65. For older people, the stages of aging have been characterized as the “go-go” group, “go-slow” group, and “no-go” group. An increase in the prevalence of various eye diseases begins in the “go-slow” group. A study of over 20,000 Medicare recipients reported that half of those studied developed at least one of three eye diseases: diabetic retinopathy, glaucoma, or macular degeneration, over a 9-year span after 65 (CDC 2010). Physical abilities and health status also influence to which of these groups an individual belongs. Physical, sensory, mental, and perception changes all alter how an older person interacts with the world around them. Reduced vision among mature adults has been shown to result in social isolation, increased risk of falling and resultant hip fractures, depression, family stress, and ultimately a greater tendency to be disabled or die prematurely (Vitale et al. 2006). Appropriate lighting that supports better vision can help older adults achieve a dramatic reduction of these negative effects.

The importance of lighting for seniors is not just limited to vision. Experiencing the brightness of daylight is necessary to maintain a healthy circadian rhythm, and direct sunlight on the skin to help synthesize vitamin D is requisite for bone health. Direct daylight needs to be incorporated into the daily routine for all seniors. As mobility decreases, getting outdoors may become more challenging, so creative solutions are needed.

Understanding the changing lighting needs of the aging population requires architects, landscape architects, interior designers, and lighting designers to learn the importance of *quality* and *quantity* of light and to incorporate features into the built environment to support aging vision and health.

Lighting Needs of the Aging Eye: Quantity and Quality

The human visual system is made up of three parts: the eyes that gather information in the form of reflected light, the visual pathway (optic nerve) which transmits the signals from the eyes to the brain, and the visual cortex of the brain, which processes the information (Sekuler and Blake 1994). Older people experience both physiological and neurological deterioration in the eye and the brain. Consequently, the lighting needs of older adults differ from the younger population. In some cases, younger people experience vision impairment comparable to that of older people, resulting in low vision. “Low vision” is defined as vision with best-corrected visual acuity less than 6/12 ($<20/40$) in the better-seeing eye (Fig. 1).

The appropriate *quantity* and *quality* of the light can help to minimize the impact of aging vision and eye diseases and improve the quality of life for older people. Almost 20 years ago, a research study found that when lighting in an older person’s home was modified to provide higher lighting levels without glare, their quality of life improved in the following ways: increased feeling of self confidence, regained activities that had been given up, successful managing on one’s own, and a general feeling of well-being (Sorensen and Brunnstrom 1995; ANSI/IES RP-28-07). Older adults require higher levels of light, or quantity of light, than younger people to compensate for dramatically reduced light receptivity. Two-thirds less light reaches their retinas, due to a smaller pupil that is almost fixed in size and increased light absorption all along the visual pathway. The recommended quantity of light for adequate vision is separated into categories of ambient lighting and task lighting (see Table 1). The values are adjusted for the nature of the visual task and for day/night conditions.

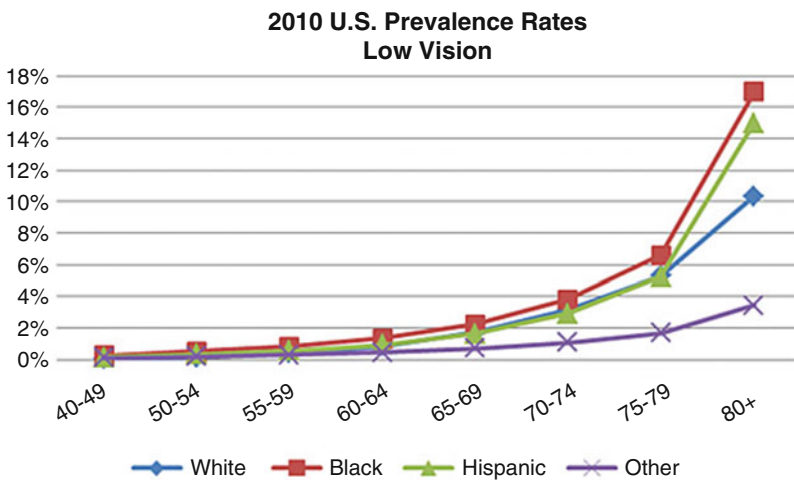


Fig. 1 2010 US age-specific prevalence rates for low vision by age and race/ethnicity (NEI Low Vision 2010)

Quality lighting includes the following characteristics: even and consistent light levels, light that is free of direct or reflected glare, light that provides good modeling of faces, light that does not flicker, spectral power distribution appropriate for day/night conditions, high color rendering, and diffused light that does not create strong shadows on floors or stairs. In addition to appropriate lighting, characteristics

Table 1 Minimum maintained average illuminance. Updated Table (2016) can be found on Springerlink

Areas	Ambient light in		Task light ^c in	
	Lux	foot-candles	Lux	foot-candles
Exterior entrance (night)	100	10		
Interior entry (day) ^d	1000	100		
Interior entry (night)	100	10		
Exit stairways and landings	300	30		
Elevator interiors	300	30		
Exterior walkways	50	5		
Administration (active)	300	30	500	50
Active areas (day only)	300	30	500	50
Visitor waiting (day)	300	30		
Visitor waiting (night)	100	10		
Resident room				
Entrance	300	30		
Living room	300	30	750	75
Bedroom	300	30	750	75
Wardrobe/closet	300	30		
Bathroom	300	30		
Makeup/shaving area	300	30	600	60
Shower/bathing rooms	300	30		
Kitchen area	300	30	500	50
Barber/beautician (day)	500	50		
Chapel or quiet area (active)	300	30		
Hallways (active hours)	300	30		
Hallways (sleeping hours)	100	10		
Dining (active hours)	500	50		
Medicine preparation	300	30	1000	100
Nurses' station (day)	300	30	500	50
Nurses' station (night)	100	10	500	50
Physical therapy area (active hours)	300	30	500	50
Occupational therapy (active hours)	300	30	500	50
Examination room (dedicated)	300	30	1000	100
Janitor's closet	300	30		
Laundry (active hours)	300	30	500	50
Clean/soiled utility	300	30		
Commercial kitchen	500	50	1000	100

(continued)

Table 1 (continued)

Areas	Ambient light in		Task light ^c in	
	Lux	foot-candles	Lux	foot-candles
Food storage (non-refrigerated)	300	30		
Staff toilet areas	200	20	600	60

Reprinted with permission from *Lighting and the Visual Environment for Senior Living (RP-28-07)* published by the Illuminating Engineering Society of North America

^aOlder Adults include persons aged 60 years and older, and people of all ages with some form of visual impairment

^bNote: Ambient light levels are minimum averages measured at 76 cm (30 in.) above the floor in a horizontal plane. Task light levels are minimums taken on the visual task. For make-up/shaving, the measurement is to be taken on the face in a vertical position

^cNote: It should be understood that the values are *minimums*. The optimum solution for task lighting is to give users control over the intensity and positioning of the light source to meet their individual needs

^dUtilization of daylight is encouraged in entryways to provide a transition between outside and interior illumination levels

of the building’s physical environment can improve the visual environment to promote greater independence. Providing a good contrast between objects and their background and/or adjacent surfaces will help to compensate for the loss of contrast sensitivity. Highly polished floor and wall surfaces must be avoided to prevent reflected glare from daylight and light fixtures.

Some problems of sensory loss can be interrelated, i.e., many people with a hearing loss depend upon lip reading or facial expressions to help them understand what is being said. However, if the lighting does not illuminate faces adequately, this coping strategy cannot be utilized successfully.

Strategies to Design Appropriate Lighting for the Elderly

“The foundation of lighting design is ensuring that people have enough light to safely, efficiently, and accurately perform predominate visual tasks,” as defined in IES DG-18 2008. The purpose of this chapter is to provide an overview of the special needs of older adults and identify areas of concern that should influence lighting design to serve them.

When we take into consideration the age-related changes to the visual system, plus the increase of eye diseases in the older population, we understand why lighting and the visual environment need to be different to serve older adults than what might be acceptable for the majority of younger people. Due to the growth of the 65+ demographic, which has already exceeded 25 % of the US population, it is important to design for the “optimum” rather than the “minimum” lighting solution so that older people, as well as younger people with low vision, have equal access within the built environment. A coordinated team approach to design, including the architect, landscape architect, interior designer, and lighting designer, is needed to truly address all of these issues. Each design discipline has an important role to play to maximize the visual capabilities of older people to keep them safe and independent.

Vision problems experienced by older people, discussed in more detail earlier in this chapter, include perception of reduced brightness, increased sensitivity to glare, slower adaptation to change in light levels, loss of contrast sensitivity, and reduced ability to distinguish colors, all of which contribute to limited mobility.

Area Specific Lighting

The combined use of daylight and electric lighting is strongly encouraged. Daylight is so important for older people that an entire section of this chapter is devoted to it specifically (see section “[Daylight: The Devil Is in the Details](#),” below).

Ambient Light: To best serve seniors, ambient lighting should be uniform. Direct/indirect lighting is the best way to provide higher light levels while avoiding direct or reflected glare. The diffused light provides light onto the walls, floor, and ceiling without creating shadows.

Task Visibility: To be comfortable in their surroundings, seniors need adequate light to find their way to and around all public buildings. Seeing faces is extremely important for a senior’s sense of security. Signage must be easy to locate, receive sufficient light, and have letters of sufficient height and contrast to be read easily. Medical facilities may pose the greatest visual challenge because of the complex nature of the facility coupled with the added stress experienced during a medical crisis.

Theaters and museums must be designed to provide adequate time for seniors’ eyes to adapt to the change in brightness when entering the hall. Finding a seat in a dark theater, for instance, will be aided significantly by providing a small amount of light and high contrast letters to identify the aisle number/letter. To ensure safety, steps must be illuminated and have a strong value contrast on the edge of the step.

Almost all activities of daily living (ADL) need a higher level of task lighting, including dressing, selecting clothing, grooming, cooking, dining, reading, writing, paying bills, identifying medication (by size, shape, and color), reading directions, and doing light house work, i.e., laundry (finding soiled areas on clothing). Older adults need to use the toilet frequently during the night. Night lights should be amber in color to light their path from the bed to the bathroom and back again. Low level amber light will not suppress the flow of melatonin and, therefore, will avoid disruption of the circadian system (see chapter “[► Photoreception for Human Circadian and Neurobehavioral Regulation](#)”) and will preserve night vision, thus making it easier to find the way back to bed. Because retired seniors have more leisure time to enjoy dining out, travel, and staying in hotels, the lighting in these settings should support the same sensitivities and visual tasks required at home.

Vanity lighting typically denotes a decorative light over the top of a mirror above a sink, which provides general room light and task light for grooming. This typical solution does not serve the senior population well, for the following reasons: (1) Light from above creates shadows under the nose and under the chin, making it difficult to apply makeup evenly or shave (white whiskers on white skin are hard to see). (2) Moving closer to the mirror makes it easier to see ones reflection.

Many older people have bad backs with limited range of motion, making it difficult to lean forward, so the vanity cabinet or sink acts as a barrier. A grooming station without the obstruction allows the user to get close to the mirror while standing without bending or twisting. Vertical lights on each side of the mirror illuminate the entire face, including under the nose and chin. A full-height, well-lighted grooming station provides a place to check appearance before leaving the house, and it can help with more critical functions, such as wound care or checking the skin for abnormal conditions.

Everyone should select their own task light at their private desk. Task lights are available in a wide range of color temperatures ranging from warm to cool. Everyone's eyes are different, depending upon many factors, including eye diseases. Some may find that a light with a cooler color temperature makes the print on a page crisp and easier to read, while others may find that the cooler color is bothersome, contributing to reflected glare.

What to Avoid

- Visibility of the light source (lamp) when standing, sitting, or reclining
- Pools of light on floors or stairs
- Scalloped patterns on walls
- Contrast ratios greater than 3:1 (ANSI/IES RP-28-2007)
- Flicker/strobe effect

Qualities of Appropriate Light Sources for Older Adults

- Sufficient lumen output.
- High color rendering index.
- The color temperature of the light source is appropriate for the time of day/night when the light is to be used or has the ability to change with the day/night cycle of light.

For appropriate quantity of light, please refer to the RP-28- 2007, Table 1.

Daylight: The Devil Is in the Details

Daylight:

The biological effects of light are an indispensable health factor. Optimum light exposure ought to be as uncontroversial an aim of future health policy as best-possible nutrition. (Ehrenstein 1995)

Providing access to daylight both inside and outside buildings is critical for people in later life. Older people experience less daylight exposure than younger people, due to decreased independence and mobility. Studies have shown that the average daily exposure above 2000 lx was 90 min for younger people aged 21–42 years (Savides et al. 1986), compared to 59 min for older people aged 55–81 years



Fig. 2 Clear glazing in skylight allows direct sunlight into the space creating glare and confusing patterns on the floor of a Memory Care unit. (Photographer: Eunice Noell-Waggoner)

(Campbell et al. 1988.) When people are placed in a nursing home, their light exposure diminishes significantly from that of people living in the community in regard to intensity, duration, and spectrum. People living in the community are free to increase the daylight in their homes and change their personal daily activities to enjoy the benefits of daylight both indoors and outdoors for improved vision, health, and well-being. However, residents of care facilities do not have the ability to modify their lighting or, for many, their daily routine. For that reason, a separate section of this article is devoted to specific recommendations for senior residential care facilities, attention to solutions for the common circadian rhythm disturbances experienced by residents, and the need of sunlight on the skin (the hands, face, and forearms) for vitamin D synthesis.

Increasing daylight within interior spaces can be very helpful for older people by not only raising the level of overall light but also providing the changing color of light required to maintain a healthy circadian rhythm (Fig. 2). However, direct beams of sunlight create glare and shadows, along with dramatic differences of light levels (Fig. 2). To be a positive source of light, the daylight needs to be controlled so that the ambient light has an even distribution throughout the space and is free of glare and shadows. There are various techniques to diffuse and balance the daylight in a space. One of the most common is to provide windows that bring in daylight from more than one direction or from different elevations, i.e., clerestory windows above the view windows, or skylights for ambient daylight across the space (Fig. 2) (ANSI/IES RP-28-07).

There are various strategies to increase the daylight available from upper windows or clerestory windows. A light shelf is a horizontal projection which separates the view window below and the daylighting section above. When the light shelf is on the outside of the building, it will provide shade for the view window below and reflect the sunlight through the upper window onto the ceiling, providing diffused lighting within the space. When the light shelf is on the inside, it reflects the low-angle sun to the ceiling and blocks the view of the bright sky. In seniors'

Fig. 3 Skylight helps to balance the brightness of daylight from the windows. Woven shades filter the brightness of the sky (Photographer: Eunice Noell-Waggoner)



homes, it is common to see a portion of the shades dropped down to block the excessively bright view of the heavens, while retaining some view of the outdoors. View windows are important to connect seniors to nature, the time of day, and seasons of the year. Tinted films or glazing are not recommended, because they alter the color perception of the outside as being gloomy and can increase feelings of depression experienced by many seniors (Fig. 3) (Bulow-Hube, 1995; [ANSI/IES RP-28-07](#)).

Glazing materials of skylights must be diffused to prevent direct beams of sunlight. The two negative effects of allowing direct sunlight to enter a space are glare and shadows. Strong shadow patterns may create optical illusions on walkways. Windows, particularly those with east or west orientation, must have shades to control excessive brightness caused by the rising or setting sun. The color/value of the shade material will have different effects. Shade materials with dark colors have the following attributes: lower surface brightness (less potential for glare), lower light transmittance (less light entering the space), better views to the outdoors, and low reflectance (higher heat gain). Shade materials with lighter colors have the following attributes: exterior views are reduced, higher surface brightness when back-lit by sunlight (glare source), higher light transmittance, and low heat gain. A dual shade system that responds to the changing daylight condition may be the optimal solution, if landscape plantings or other exterior shading options are not available (Fig. 4) ([ANSI/IES RP-28-07](#)).

Older eyes adapt much more slowly to changes in brightness levels, such as transitioning from outdoors to interior space during the day and vice versa at night. Exterior entrances, entries, and lobbies must provide a gradual change of light level so that the eyes of older people have time to make the transition and adapt to the changing light condition. Utilizing the daylight techniques described above will help with the transition during the day. At night, a gradual lowering of interior light levels will help aging eyes adapt to nighttime lighting conditions when they go outside.

Fig. 4 Brightness of white shade back-lit by the sun is perceived as glare
(Photographer: Eunice Noell-Waggoner)



Exterior night lighting should be directed downward to light the walking surface while avoiding light toward the eyes.

Senior Care Facilities

Is light deprivation an unintended consequence of entering a senior care facility? There is strong evidence that lighting in most senior care facilities does not support aging vision. In addition, there exists a strong correlation between the known photobiological effects of light and many common physical and psychological problems experienced by residents of care facilities where insufficient light exposure is the norm.

Inadequate Lighting Regulations Coupled with Greater Vision Impairment.

Retirement housing, assisted living, nursing homes, and memory care facilities all come under the umbrella of *senior care facilities*. The average age of people living in these facilities is in the range of 83–86. These facilities offer different levels of care dependent upon the individual's physical and mental status, not his or her age. Since all these residents are in the upper age range, their visual needs are greater than the younger “go-go” group described earlier in this chapter. One study reported that residents of nursing homes experience 13–15 times greater visual impairment than age-matched people living in the community (Tielsch et al. 1995; ANSI/IES RP-28-07).

Given the severity of vision impairment documented by many studies, it is difficult to understand why the Center for Medicare & Medicaid Services (CMS) has no definitive lighting regulations governing these facilities, except for vague terms requiring “adequate and comfortable light levels” (CMS 2009). Since we

know that normal age-related changes to vision and eye disease increase with age, the question is, “adequate and comfortable” for whom? CMS is the only agency that could have a national impact, given that as of 2004, 70 % of nursing home care is paid by Medicaid (Wodchrist et al. 2007). Each state creates their own regulations for senior care facilities. Without CMS setting a standard which could be used by all states, there is great variation in lighting regulations from one state to another. Some states reference wattage of an incandescent light source, rather than illuminance levels. A look at the nursing home regulations for each state, which can be viewed at the University of Minnesota’s website, <http://www.hpm.umn.edu/NHRegsPlus>, will quickly illustrate the wide disparity of requirements. The site allows you to search by *Regulations by State* or *Regulations by Topic*. You will find Lighting, Noise, Temperature, HVAC, and Odors as subtopics under Physical Environment, Facility Wide.

One survey of 53 nursing homes in four states found that most facilities were dimly lit. When compared to the recommended illuminance values found in ANSI/IES RP-28-07, the lighting was rated as barely adequate in 45 % of the hallways, 17 % of activity areas, and 51 % of the resident rooms (Sloan et al. 2000).

In 2008, Creating Home in the Nursing Home: A National Symposium on Culture Change and the Environmental Requirements was held in Washington, DC, sponsored by CMS. Dr. Lois Cutler reported on studies of nursing home environments she had conducted in the last decade. In large part, her work focused on a study of the physical environments of 1,988 residents in 40 nursing homes in five states (Cutler 2008). The table above, part of Dr. Cutler’s report, illustrates not only the extent of suboptimal lighting but also the variation in light levels throughout the facilities. Considerable variation was found in corridor readings, especially in corridors without windows. The highest level was taken directly under the brightest fixture (or daylight); the lowest reading was often in a corner. When there is a great difference between a very high *Maximum* and very low *Minimum*, as shown in the highlighted rows below, this difference suggests that daylight/sunlight was present and therefore the *Mean* does not represent a uniform light level during the day. This also calls into question the quantity of light present at night. A comparison of the remaining *Mean* recorded light levels with the recommended *minimum* illuminance levels found in ANSI/IES RP-28- 2007 Table 2 (light blue columns to the left) demonstrates that existing light levels in these nursing homes are far below the recommended minimums at almost every measurement location. Nurse’s stations had consistently higher light levels than resident areas, which is true in almost all facilities.

CMS has done little in the years since this report to require or provide a more appropriate and definitive lighting regulation for senior care facilities (CMS 2009).

In addition to the specific locations within the nursing homes measured above, night lights are of particular concern because of their direct impact on quality sleep. Almost all states require night lights, but the requirement is vague, allowing the general ceiling lights to qualify as night lights. Only 52 % of the 1988 resident rooms that were surveyed had a separate night light (Cutler 2008).

Table 2 Minimum Recommended Illuminance

Research Study Light Measurements				RP-28 -07 Minimum	
Light Levels Location	Mean	Minimum	Maximum	Ambient	Task
Head of bed	37 fc	4 fc	95 fc	30 fc	75 fc
Bathroom at sink	25 fc	1 fc	75 fc	30 fc	60 fc
Bathroom at commode	13 fc	1 fc	48 fc	30 fc	--
Highest level tub/shower room	83 fc	7 fc	505 fc		
Lowest level tub/shower room	17 fc	2 fc	85 fc	30 fc	--
Highest level nurse's station	93 fc	10 fc	410 fc		
Lowest level nurse's station	33 fc	5 fc	140 fc	30 fc	50 fc
Highest level unit corridor	109 fc	10 fc	3200 fc		
Lowest level unit corridor	15 fc	1 fc	82 fc	30 fc	---
Highest level unit lounge	292 fc	15 fc	3100 fc		
Lowest level unit lounge	25 fc	2 fc	132 fc	30 fc	75 fc
Highest level unit dining room	428 fc	2 fc	10,500 fc		
Lowest level unit dining room	25 fc	3 fc	130 fc	50 fc	50 fc

Alzheimer's Disease and Other Dementia: Impact on Vision

Advanced age is a risk factor for Alzheimer's disease (AD). After age 85, the risk reaches nearly 50 % of that age group (Alzheimer's Association 2014). People living with AD will experience the same age-related changes to vision and eye disease as other aging people. However, those with AD can also suffer from visual disturbances caused by changes in the brain. Their problems occur in the areas of visual field loss (Trick et al. 1995) and perception of motion, depth, color and contrast (Solomons 2005). Because of these factors, dementia is considered an independent risk for falling. Residents of nursing homes with dementia fall twice as often as those without dementia (Van Doorn et al. 2003). Appropriate lighting and increased value contrast can abate the number of falls in the AD population, as well as improve the residents' overall perception of the world around them.

Circadian Rhythm Support

Circadian Rhythm Disturbances are associated with increased sleep disorders (Van Someren 2000) and depression in the elderly (Lieverse et al. 2011). Epidemiological studies indicate that 40–70 % of the elderly population suffers from chronic sleep disturbances. Many aspects of the body's circadian regulation show changes in the elderly age group (Van Someren 2000; ANSI/IES RP-28-07), and reduced bright light exposure increases the problem.

Appropriate light to support healthy circadian rhythms is not currently included in national design guidelines or regulations for senior care facilities in the United States published by CMS. There is no recognition of the importance of experiencing bright light during the day and darkness at night. Therefore, residents of nursing homes and long-term care facilities have significantly more sleep

Fig. 5 Memory Care Home with very little daylight and low light levels from incandescent sources. The light will not provide the color or intensity to entrain circadian rhythms. A greater value contrast is needed between the furniture and the floor to enhance vision. (Photographer: Eunice Noell-Waggoner)



problems than others of the same age group living in the community (Clapin-French 1986; Shochat et al. 2000). The condition that residents of nursing homes find themselves in has been described as “spending their final years in a twilight state, rarely fully awake or fully asleep, and physiologically in the dark” (Ancoli-Israel et al. 1991; Fig. 5).

Studies indicate that nursing home residents experience only minimal bright light exposure during the day (Noell-Waggoner 2006). In fact, one study reported an average of only 1 min per day of bright light exposure for residents with dementia (Ancoli-Israel et al. 1997). Increasing light intensity in the day rooms where residents with dementia spend most of their time has proven to slow down the rate of cognitive decline, reduce agitation, improve depressive mood, and increase sleep time when compared with other residents with dementia living under normal memory care facility lighting regimens (Riemersma-van der Lek et al. 2008). Contrary to treatment with hypnotics, treatment with increased light intensity has been shown to improve sleep without adverse effects, and it even results in an improvement of performance, daytime energy, and quality of life (Sorensen and Brunnstrom 1995; Riemersma-van der Lek et al. 2008).

There is a growing concern by CMS and the Canadian Health Ministry about the overprescription of antipsychotic drugs for those with dementia-related behaviors. These commonly prescribed antipsychotics can have a sedative effect. Fifteen percent of all Ontario, Canada’s nursing home residents aged 65–79 are taking both an antipsychotic and a sedative at the same time. “Sedation comes at a price – falls, bedsores, blood clots and direct adverse reactions to the drugs themselves, which can sometimes be fatal,” according to Dr. David Juurlink (Bruser and McLean 2014). In other studies, insomnia has been identified as a risk factor for falls among nursing home residents (Avidan et al. 2005). Increased exposure to bright light in the daytime can improve sleep patterns and mood, thereby reducing the need for sedatives and antipsychotics.

The greatest benefit in my opinion is realizing that a life so compromised by an unwelcome and uninvited dementia illness can be improved by something as simple as appropriate lighting. I consider it a disservice to the residents to add excess disabilities by not providing the best setting possible for their 24/7 residence in our communities. Marge Coalman, VP of Wellness & Programs, Touchmark (Coalman 2010)

Importance of Vitamin D₃

Vitamin D₃ has long been associated with bone health. It is also essential to allow the body to maximally utilize calcium and optimize muscle function, which are needed for maintaining skeletal integrity and muscle strength (Wolopwitz and Gilchrist 2006). A significant number of elderly people living in the northern regions of the northern hemisphere are vitamin D deficient. Vitamin D deficiency has been reported in 30–40 % of hip fracture patients in Great Britain and at a Boston hospital (Doppelt et al. 1983). Hip fractures and deficiency of Vitamin D₃ are more common among nursing home residents than among the elderly population living in the community (Nieves and Lindsay 1994).

For residents of care facilities to gain the benefits of vitamin D₃, they need regular direct sunlight exposure on their hands, forearms, and face, without sunscreen. Vitamin D₃ is synthesized in the skin when exposed to the sun's ultraviolet B (UVB) radiation between the wavelengths of 290 and 315 nm (Holick 2004). Since typical glazing materials for window and skylights block UVB, people must spend short periods of time outdoors on a regular basis to benefit from direct exposure to the sun's UVB. Views to the outdoors and ease of access are important factors in motivating the residents to go outside. Well-designed gardens with paved walkways, seating, areas of sun and shade, greenery, and interesting views are all needed to encourage residents to spend time outside (Rodiek 2006).

Unfortunately, outdoor access has been limited in long-term care facilities. A study of 40 nursing homes found that 46 % of the facilities surveyed had no direct outdoor access from the unit (Cutler and Kane 2005). Residents of these nursing homes were surveyed as to how often they get outdoors and whether that amount is what they would prefer. Of 1,068 residents, 32 % responded they went outdoors less than once a month, 13.4 % responded they went outdoors less than once a week, 17 % went out several times a week, and 22 % said they went outdoors every day. Thirty-nine percent of the residents responded that they did not get outside as much as they would like. Based on the limited time nursing home residents are spending outdoors, it is not surprising that hip fractures and vitamin D₃ deficiency are more common among nursing home residents than in the general population of the same age (Tinetti and Speechley 1989).

Spending time outdoors is a natural way to provide the bright light for circadian entrainment and light on the skin for vitamin D₃ synthesis, the two photobiological effects of light. Therefore, regular, scheduled outdoor activities for all residents should be part of a facilities' plan of care. Increasing controlled daylight within

the built environment can also provide the necessary higher light levels, with the appropriate color of light during the day, to entrain circadian rhythm and improve sleep patterns, especially in facilities where the outdoors is not easily accessible. Equally important is controlling light at night to allow residents to sleep in darkness.

New color-turning *light-emitting diodes* (LED) offer new possibilities to change the color and intensity of light throughout the day. Rigorously controlled laboratory research needs to be done before architects, lighting designers, and manufacturers rush into applying imperfect and incomplete data (Wirz-Justice and Fournier 2010). *Measuring and using light in the melanopsin age* (Lucas et al. 2014) provide guidance until there are research studies upon which manufacturers and design professionals can rely.

The positive effects of appropriate light for aging vision, increased daylight in common areas, and direct exposure to sunlight for the elderly population are well understood and should now be included in federal and state regulations. However, additional rigorous research is needed before regulatory agencies can have the confidence to issue directives for specific light color and intensity to coincide with the time of day/night for regulating circadian, hormonal, and behavioral systems in senior care environments.

Call to Action

Research Needed: The design community needs to come together to identify qualified research teams and senior care facilities that are willing to accept the challenge of participating in rigorous studies, upon which new lighting regulation for vision and health can be based.

Building Codes: Many of our building codes are based on the abilities of younger people. The Americans with Disabilities Act does not include requirements to support aging vision or low vision. Volunteers are needed to join the various code/guidelines committees to ensure that older people are given equal access to the built environment.

References

- Alzheimer's Association (2014) 2014 Alzheimer's disease facts and figures. http://www.alz.org/downloads/Facts_Figures_2014.pdf
- Ancoli-Israel S, Jones DW, Hanger MA, Parker L, Klauber MR, Kripke DF (1991) Sleep in the nursing home. In: Kuna ST et al (eds) *Sleep and respiration in aging adults*. Elsevier Science, New York, pp 77–84
- Ancoli-Israel S, Klauber MR, Jones DW, Kripke DF, Martin J, Mason W, Pat-Horenczyk R, Fell R (1997) Variations in circadian rhythms of activity, sleep and light exposure related to dementia in nursing home patients. *Sleep* 20:18–23
- ANSI/IES RP-28 (2007) Recommended practice 28 2007: lighting and the visual environment for senior living. Illuminating Engineering Society of North America, New York

- Avidan YA, Fries BE, James ML, Szafara KL, Wright GT, Chervin RD (2005) Insomnia and hypnotic use, recorded in the minimum data set, as predictors of falls and hip fractures in Michigan nursing homes. *J Am Geriatr Soc* 53(6):955–962
- Bruser D, McLean J (2014) Antipsychotic drugs prescribed to seniors at alarming rates, Providence finds. *Toronto Star*, 21 Apr 2014, thestar.com
- Campbell SS, Kripke DF, Gillin JC, Hrubovak JC (1988) Exposure of light in healthy elderly subjects and Alzheimer's patients. *Physiol Behav* 42:141–144
- CDC (2010) Improving the nations vision health: a coordinated public health approach. Center of Disease Control and Prevention. http://www.cdc.gov/visionhealth/pdf/improving_nations_vision_health.pdf
- Clapin-French E (1986) Sleep patterns of aged people in long-term care facilities. *J Adv Nurs* 11:57–66
- CMS (2009) Manual System, Pub. 100–07 State Operations, 483.15(h)(5) – *Environment*, p 16. <https://www.cms.gov/Regulations-and-Guidance/Guidance/Transmittals/downloads/R48SOMA.pdf>
- Coalman M (2010) Communication to Oregon Department of Human Services, Seniors and People with Disabilities, in support of improved lighting requirements for Memory Care Facilities
- Cutler L (2008) Proceedings: creating home in the nursing home: a national symposium on culture change and the environmental requirements. Center for Medicare & Medicaid Services, Washington, DC
- Cutler LJ, Kane RA (2005) As great as all outdoors: a study of outdoor spaces as a neglected resource for nursing home residents. *J Hous Elder* 19(3/4):29–48
- Doppelt SH, Neer RM, Daly M et al (1983) Vitamin D deficiency and osteomalacia in patients with hip fractures. *Orthop Trans* 7:512–513
- Ehrenstein W (1995) Circadian lighting systems. *Int Light Rev Neth* 2:64–67
- Holick MF (2004) Sunlight and vitamin D for bone health and prevention of autoimmune diseases, cancers, and cardiovascular disease. *Am J Clin Nutr* 80(6):1678S–1688S
- IES DG-18 (2008) Light + design: a guide to designing quality lighting for people and buildings. Quality of the Visual Environment Committee, Illuminating Engineering Society of North America, New York
- Lieveer R, Van Someren EJ, Nielen MM, Uitdehaag BM, Smit JH, Hoogendijk WJ (2011) Bright light treatment in elderly patients with nonseasonal major depressive disorder: a randomized placebo-controlled trial. *Arch Gen Psychiatry* 68(1):61–70
- Lucus RJ, Peirson SN, Berson DM, Brown TM et al (2014) Measuring and using light in the melanopsin age. *Trends Neurosci* 37(1):1–9
- NEI Low Vision (2010) <https://www.nei.nih.gov/eyedata/lowvision.asp>
- Nieves JW, Lindsay R (1994) Vitamin D malnutrition and skeletal health in the nursing home. *Nurs Home Med* 2(8):167–170
- Noell-Waggoner E (2006) Lighting in nursing homes – the Unmet Need. 2nd Expert Symposium on Light and Health, International Commission on Illumination
- Riemersma-van der Lek RF, Schwaab DF, Twisk J, Hol EM, Hoogendijk WJG, Van Someren EJW (2008) Effect of bright light and melatonin on cognitive and noncognitive function in elderly residents of group care facilities: a randomized controlled trial. *JAMA* 299(22):2642–2655
- Rodiek SD (2006) A missing link: can enhanced outdoor spaces improve seniors housing? *Seniors Hous Care J* 14(1):3–19
- Savides TJ, Messin S, Senger C, Kripke DF (1986) Natural light exposure of young adults. *Physiol Behav* 38:571–574
- Sekuler R, Blake R (1994) Perception. McGraw-Hill, New York
- Shochat T, Martin J, Marler M, Ancoli-Israel S (2000) Illumination levels in nursing homes patients: effects on sleep an activity rhythms. *J Sleep Res* 9(4):373–379
- Sloan PD, Mitchell CM, Calkin M, Zimmerman S (2000) Lighting and noise levels in Alzheimer's. *Res Pract Alzheimers Dis* 4:241–249
- Solomons H (2005) Vision and dementia. *Optometry Today*, Clinical Articles, 7 Oct 2005, pp 25–28. http://www.optometry.co.uk/uploads/articles/0b5c132a1f57bdda881aa091b1ac2476_Solomons051007.pdf

- Sorensen S, Brunnstrom G (1995) Quality of light and quality of life: an intervention study among older people. *Int J Light Res Technol* 27(2):113–119
- Tang JY, Fu T, Lau C, Ho DH, Bikle DD, Asgari MM (2012) (Part I & II) Vitamin D in cutaneous carcinogenesis. *J Am Acad Dermatol* 67(5): Part I 803–816, Part II, 817–828
- Tielsch JM, Javits JC, Coleman A, Katz J, Sommer A (1995) The prevalence of blindness and visual impairment among nursing home residents in Baltimore. *N Engl J Med* 332(18):1205–1209
- Tinetti ME, Speechley M (1989) Prevention of falls among the elderly. *N Engl J Med* 320(16):1057–1058
- Trick GL, Trick LR, Morris P, Mitchell W (1995) Visual field loss in senile dementia of the Alzheimer's type. *Neurology* 45:68–74
- Van Doorn C, Gruber-Baldini AL, Zimmerman C, Hebel JR et al (2003) Dementia as a risk factor for falls and fall injuries among nursing home residents. *J Am Geriatr Soc* 51(9):1213–1218
- Van Someren EJW (2000) Circadian and sleep disturbances in the elderly. *Exp Gerontol* 35:1229–1237
- Vitale S, Cotch MF, Sperduto RD (2006) Prevalence of visual impairment in the United States. *JAMA* 295(18):2158–2163
- Wirz-Justice A, Fournier C (2010) Light, health and wellbeing: implications from chronobiology for architectural design. *World Health Design*. www.worldhealthdesign.com
- Wodchist WP, Hirth RA, Fries BE (2007) Effect of medicaid payment on rehabilitation care for nursing home residents. *Health Care Finan Rev* 28(3):117–129
- Wolopwitz D, Gilchrist BA (2006) The vitamin D questions: how much do you need and how should you get it? *J Am Acad Dermatol* 54(2):301–317

Photobiological Safety

Christophe Martinsons

Contents

Introduction	866
The Issues of Glare and Photobiological Safety of SSL Products	867
Effects on the Skin	869
Effects on the Eye	869
Photochemical Retinal Damage	870
The Blue-Light Hazard	872
Regulations on Personal Exposure to Optical Radiations Emitted by Artificial Light Sources	876
Assessment of Personal Exposure to Optical Radiations Emitted by Artificial Light Sources	876
Photobiological Safety Standards for Lighting Products	877
Limitations of the CIE/IEC Blue-Light Hazard Assessment	888
Blue-Light Hazard Exposure Data Concerning LEDs and SSL Products	889
Conclusions	892
References	894

Abstract

Photobiology is a scientific field that involves biology, physics and chemistry in order to study the effects of optical radiations on living organisms. Lighting systems are sources of artificial optical radiations used primarily to provide light to the human eye in order to enable visual processes in the absence of enough daylight. The first photobiological effect of a visible light source is vision itself.

Photobiological safety refers to the undesirable effects of optical radiations on human tissues, especially the skin and the eye. These effects have several possible

C. Martinsons (✉)

Lighting and Electromagnetism Division, Centre Scientifique et Technique du Bâtiment,
St Martin d'Hères, France

e-mail: christophe.martinsons@cstb.fr

causes, according to the exposed tissue, the wavelength of the incident radiation, the intensity of the exposure and the duration of the exposure. According to these parameters, the effects can be temporary (reversible), or permanent in the case of severe exposures.

This chapter presents an overview of the knowledge concerning the photobiological safety of LEDs and products using LEDs such as solid-state lighting (SSL) products.

Introduction

Photobiology is a scientific field that involves biology, physics, and chemistry in order to study the effects of optical radiations on living organisms. Lighting systems are sources of artificial optical radiations used primarily to provide light to the human eye in order to enable visual processes in the absence of enough daylight. The first photobiological effect of a visible light source is vision itself. In humans, light interacts with several ocular tissues and reaches four types of retinal photoreceptors that send signals to a multilayered retinal neural network, directly connected to the brain through the optical nerves.

Apart from vision, one of the most important photobiological effects of visible light on animals is the regulation of circadian rhythms. This effect has been observed by chronobiologists in humans (including some blind subjects) and many other animal species since the 1980s. Light happens to be the most powerful agent to perform the daily synchronization of the biological circadian clock, whose period intrinsically deviates from 24 h by a short delay or a short advance of a few minutes, up to a few tens of minutes. In the absence of light stimuli, the circadian clock would drift and become desynchronized with the daily schedule. The most striking feature of this synchronization mechanism is that it only happens through the eye. The discovery of a new type of photoreceptive cells in the retina in the 1990s provided the physiological basis to explain this phenomenon. A small number of ganglion cells were found to have a photoreception capacity that does not contribute to vision. It has been demonstrated that the optical excitation of these cells is responsible for suppressing the production of melatonin, the sleep hormone, and is also responsible for many other non-visual effects such as pupil constriction, increase of the heart rate and body temperature, etc.

Photobiological safety refers to the undesirable effects of optical radiations on human tissues, especially the skin and the eye. These effects have several possible causes, according to the exposed tissue, the wavelength of the incident radiation, the intensity of the exposure, and the duration of the exposure. According to these parameters, the effects can be temporary (reversible) or permanent in the case of severe exposures.

This chapter presents an overview of the knowledge concerning the photobiological safety of LEDs and products using LEDs such as solid-state lighting (SSL) products.

The Issues of Glare and Photobiological Safety of SSL Products

In a typical LED, the chip that emits light is so small that although the total emitted flux may be moderate, the radiance and luminance levels may be extremely high. For example, luminance values greater than 10^7 cd.m^{-2} and radiance values greater than $50,000$ $\text{W.m}^{-2}.\text{sr}^{-1}$ are common figures for white LED components used in lighting products (ANSES 2010). These values are much higher than the values found in the case of common lamps used in general lighting such as fluorescent lamps (1000 to 10,000 cd.m^{-2}) and halogen lamps (10^5 to 10^6 cd.m^{-2}). Professional high-power lamps such as high-intensity discharge lamps also have very high luminance levels but are not used by the general public.

The fact that most LEDs, even low-power components, have very high luminance levels has raised concerns about glare, which can be responsible for a discomfort (discomfort glare) or a temporary reduction of visual acuity (disability glare). Disability glare appears with high vertical illuminance levels on the eye (corneal illuminance). Light is scattered in the ocular tissues causing a veiling phenomenon, which can be characterized by a veil luminance. By definition, glare phenomena are temporary and reversible as long as no permanent ocular damage is induced.

Glare is a source of indirect hazards, which are not caused by the light itself. For instance, glare can cause accidents at the workplace when machines and tools cannot be safely used. In the everyday life, glare can be the cause of vehicle accidents and falls.

Normalized indices were defined by the CIE to characterize the glare of lighting installations. The unified glare rating (UGR) is widely used in indoor lighting as a measure of the discomfort glare. It is related to the luminance ratio of the light source to the luminance of the background. However, the UGR method cannot be applied to very small light sources, whose solid angular subtense is smaller than 0.0003 sr (CIE 1995). For instance, at a distance of 1 m, the light source must be larger than $1.5 \text{ cm} \times 1.5 \text{ cm}$. Despite this fundamental limitation given by the CIE, lighting manufacturers and designers usually perform UGR calculations on SSL luminaires consisting of multiple small LED sources but incorrectly considering the average luminance over the whole area of the luminaire. This approach is misleading as the resulting UGR is low and does not reflect the physiological perceived glare. Therefore, the use of UGR should be restricted to SSL products with large diffusers, without any visible point sources.

In indoor lighting, luminance classes are often used to define “visually comfortable” luminaires. The luminance classes are not normalized. They correspond to maximal luminance values between 1000 cd.m^{-2} and 5000 cd.m^{-2} , which are relatively low values, only applicable to luminaires fitted with diffusers. It is more accurate to define the visual comfort by using a luminance ratio criterion. For instance, the French standard on visual ergonomics NF X 35–103 (AFNOR 2013) recommends to limit the ratio of the luminaire luminance to the surrounding luminance to a factor between 20 and 80.

Disability glare is often assessed in outdoor lighting, especially for high-power installations such as stadium lighting (glare index GR). For street lighting, disability

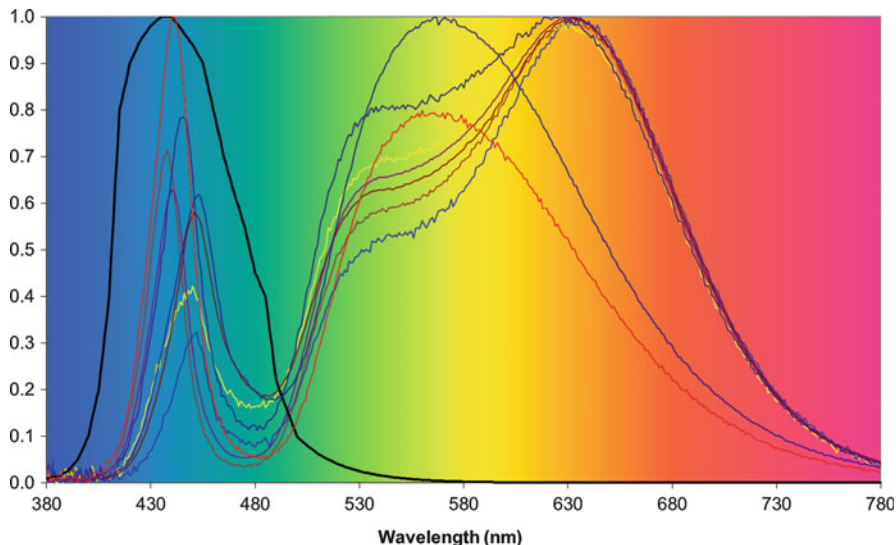


Fig. 1 Action spectrum $B(\lambda)$ of the blue-light hazard (*black curve*) and eight examples of LED emission spectra (*colored curves*), chosen to illustrate the possible coincidence of the short-wavelength emission peak with the spectral range of the blue-light hazard, maximum at around 437 nm (CSTB data)

glare is assessed by using the threshold index (TI) which quantifies the reduction of the visual contrast caused by the veil luminance. The disability glare indices GR and TI are applicable to high-power luminaires and lighting installations located sufficiently far away from the viewer, whatever the lighting technology.

In addition to the high luminance values, another key feature of LEDs has attracted the attention of lighting specialists and ophthalmologists. The vast majority of commercial LEDs producing white light rely on a chip emitting blue light associated with layers of luminophores to produce light at longer wavelengths by fluorescence. As a consequence, the emission spectrum of a white LED consists of a narrow blue primary peak and a large secondary peak in the yellow-orange-red region. The two peaks are separated by a region of low emission in the blue-green part of the spectrum (Behar-Cohen et al. 2011). In many cases, the blue peak lies in the spectral region corresponding to the highest retinal phototoxicity, as shown in Fig. 1 and detailed in the following sections.

The international guidelines concerning the human exposure limits to optical radiations are established and regularly updated by ICNIRP (International Commission for Non-Ionizing Radiation Protection). The exposure to incoherent visible and infrared radiation is addressed in ICNIRP (2013). The exposure to UV radiation is addressed in ICNIRP (2004).

In the case of constant light sources (non-pulsed sources), the effects are summarized in the following sections.

Effects on the Skin

The deleterious effects of light to the skin essentially appear in the UV range (e.g., erythema, carcinogenesis, aging, melanogenesis, etc. (IESNA 2010)). With visible and infrared radiation, burns can be induced with very high irradiances. LEDs used in SSL are currently far from reaching the high irradiance levels required to burn the skin. Therefore, the general population should not be concerned by potential risks to the skin arising from the use of LEDs in lighting. Only a small number of people suffering from photosensitive syndromes might see an aggravation of their preexisting condition triggered by blue light emitted by LEDs. Patients taking photosensitizing drugs should also be aware of a potential risk.

Effects on the Eye

According to the wavelength, optical radiation interacts with different ocular tissues, as Fig. 2 illustrates.

Since UV radiation is mainly absorbed by the cornea and the lens, excessive exposures lead to photokeratitis, photoconjunctivitis, and cataracts. Infrared radiation with wavelengths greater than about 1.4 μm is mainly absorbed by the cornea and may induce corneal burns. Emitting negligible amounts of UV and IR radiation,

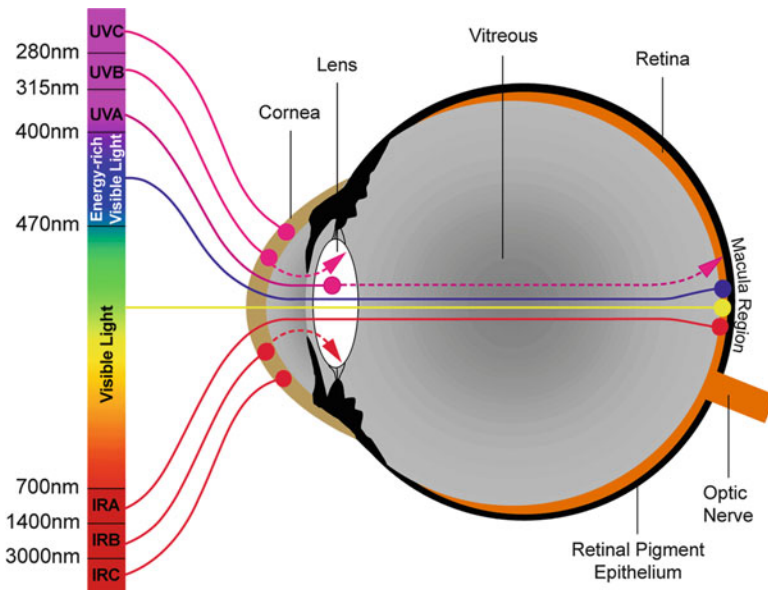


Fig. 2 Adapted from Behar-Cohen et al. (2011). Illustration of the different penetration depths of optical radiation in the eye according to the wavelength

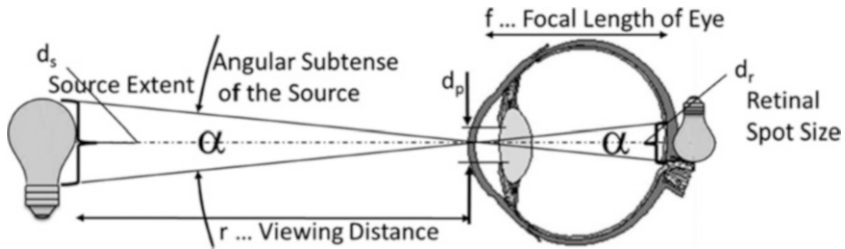


Fig. 3 Reproduced from ICNIRP (2013). Imaging of a light source on the retina showing the retinal image and the angular subtense of the source

LEDs should not be expected to contribute to the apparition of photokeratitis, photoconjunctivitis and cataracts.

Visible light (0.38–0.78 μm) and near-infrared radiation (0.78–1.4 μm) are focused on the retina and may induce retinal injuries with excessive exposures, which can be the result of thermal damage or photochemical damage:

- Thermal damage (thermal retinopathy) appears with a short-time exposure to a very high irradiance level. The exposure levels needed to produce thermal damage on the retina cannot be met with light emitted by LEDs of current technologies.
- Photochemical damage (photochemical retinopathy) appears after a short-time intense exposure or after a prolonged exposure to lower light levels.

It is important to mention that the retinal exposure to a light source is defined by both the exposure time and the retinal irradiance ($\text{W}\cdot\text{m}^{-2}$) where the retinal image of the light source is produced by the optical system formed by the cornea and the crystalline lens. The retinal irradiance is proportional to the radiance of the light source ($\text{W}\cdot\text{m}^{-2}\cdot\text{sr}^{-1}$), the transmittance of the ocular media, and the pupil diameter and inversely related to the effective focal length of the eye (see Fig. 3). The exposure dose ($\text{J}\cdot\text{m}^{-2}$) is the time integral of the retinal irradiance over the exposure duration (ICNIRP 2013).

From a photometric point of view, the retinal irradiance and the source radiance do not depend on the viewing distance. The viewing distance only defines the size of the optical retinal image. However, the real “physiological” retinal image is the result of the spreading of the image caused by the eye movements. The influence of the eye movements on the physiological retinal image is more pronounced for small optical images (remote light sources) than for large images (light sources at close range).

Photochemical Retinal Damage

Visible light falling on the retina interacts with the visual photoreceptors (rods and cones) but also with the retinal pigment epithelium (RPE). The RPE is the outer layer of the retina (Fig. 4). It plays a crucial role in the phagocytosis of photoreceptor outer

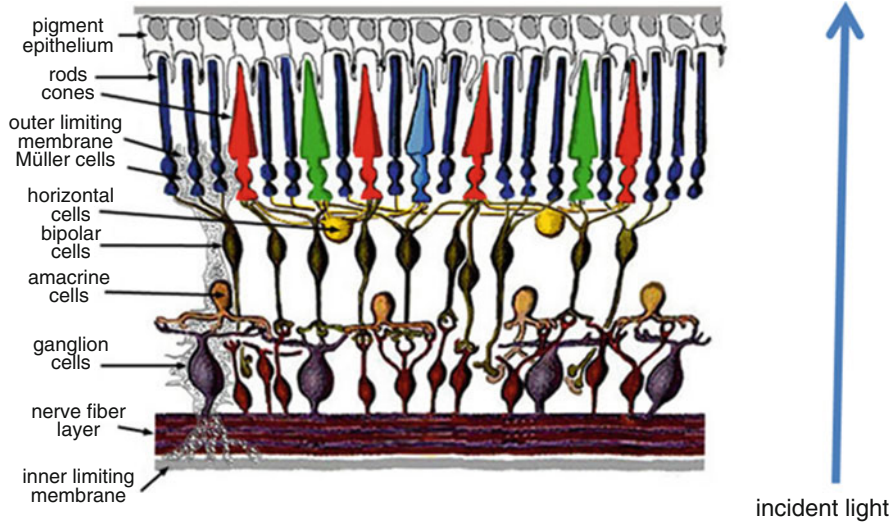


Fig. 4 Cross section of the human retina, adapted from KOLB (2011)

segments and the regeneration of visual pigments. RPE cells contain melanin (a photoprotective pigment) and lipofuscin, a substance which accumulates with age and is associated with some retinal disorders such as age-related macular degeneration (ARMD) (Behar-Cohen et al. 2011).

Research in photobiology has been carried out for more than 50 years on mammalian retinas (rats, mice, monkeys) in order to identify the injuries caused by retinal light exposures that were measured in terms of retinal irradiance dose (in $\text{J}\cdot\text{m}^{-2}$). This body of research reveals that there could be two types of retinal damage processes induced by visible light (ICNIRP 2013):

- Type 1 (Noell et al. 1966): the damage observed after 12 h per day (long exposures) is the bleaching of the retinal photopigments, with a possible toxic buildup in the RPE. The action spectrum of the type 1 damage is very similar to the photopic sensitivity of the eye $V(\lambda)$.
- Type 2 (Ham et al. 1976): the damage is a photoretinopathy caused by phototoxic reactions in the RPE, following an acute exposure to blue light. Blue light excites lipofuscin by producing reactive oxygen species and free radicals, causing an oxidative stress to the RPE cells.

The existence of type 1 retinal damage was questioned by Van Norren in 2011 (Van Norren and Gorgels 2011), following an extensive review of the literature and the lack of reproducible data related to this type of damage. Figure 5 is an excerpt from Van Norren and Gorgels (2011). The graphs summarize the dose obtained for retinal damage as a function of wavelength.

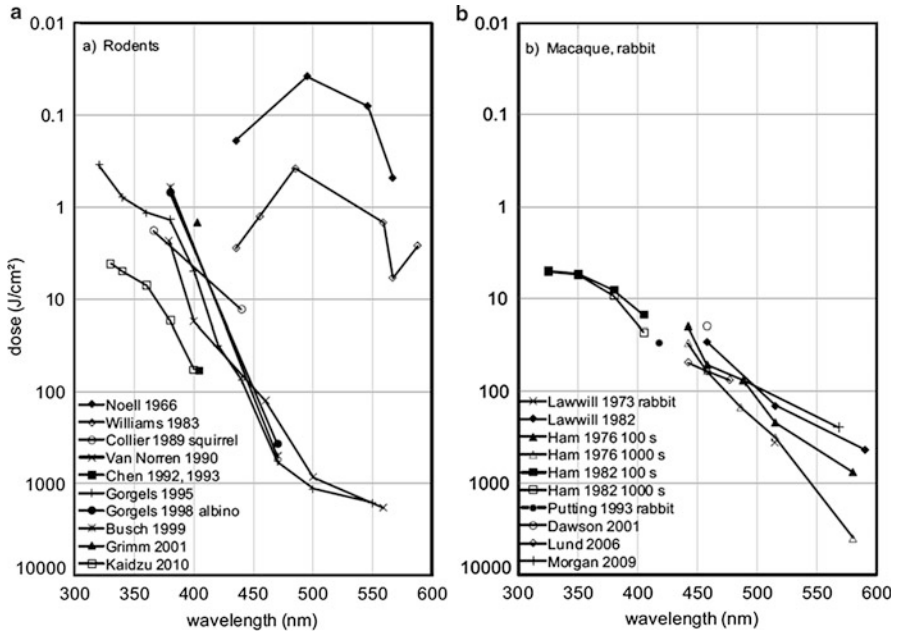


Fig. 5 Reproduced from Van Norren and Gorgels (2011). Dose for retinal damage as a function of wavelength. The literature source is indicated by first author and year of publication. (a) Data for rats, except when stated otherwise. (b) Data for macaque, except when stated otherwise

Type 2 damage was first recognized in humans in the 1960s as the major cause of photoretinitis for arc welders and people who observed a solar eclipse without eye protection (ICNIRP 2013).

Research is currently being carried out to investigate the dose and wavelength dependence of light-induced retinal damage (Shang et al. 2014; Boulenguez et al. 2014).

The Blue-Light Hazard

Unlike type 1 damage, type 2 damage is rather well established and serves as the basis of the ICNIRP guidelines concerning the blue-light hazard. For the general population, the action spectrum of the blue-light hazard is $B(\lambda)$, represented in the graph of Fig. 6. However, people born without crystalline lens (aphakic) or having received intraocular lens implants (pseudophakic) are exposed to a greater amount of retinal blue and UV light compared to phakic subjects exposed to the same light source. In these cases, the action spectrum defined by ICNIRP is $A(\lambda)$, also represented in Fig. 6. ICNIRP also recommends using the $A(\lambda)$ action spectrum when assessing the photobiological safety of infants under the age of two, due to the greater transparency of their crystalline lens (ICNIRP 2013).

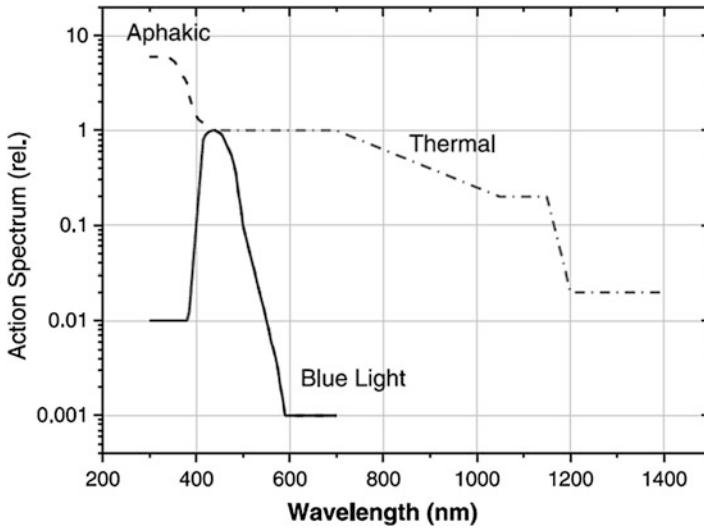


Fig. 6 Reproduced from ICNIRP (2013). Action spectra for blue-light photoretinopathy with crystalline lens (phakic) $B(\lambda)$ and without lens (aphakic) $A(\lambda)$ and for thermally induced photoretinopathy $R(\lambda)$

Table 1 Effective field of view angle (FOV) as a function of the exposure duration (ICNIRP 2013)

Exposure duration (second)	Acceptance averaging angle Y_{ph} (radian)
$t < 100$ s (about 1.7 min)	0.011
$100 \leq t < 10,000$ s (about 2.8 h)	$0.0011 \times t^{0.5}$
$t > 10,000$ s	0.110

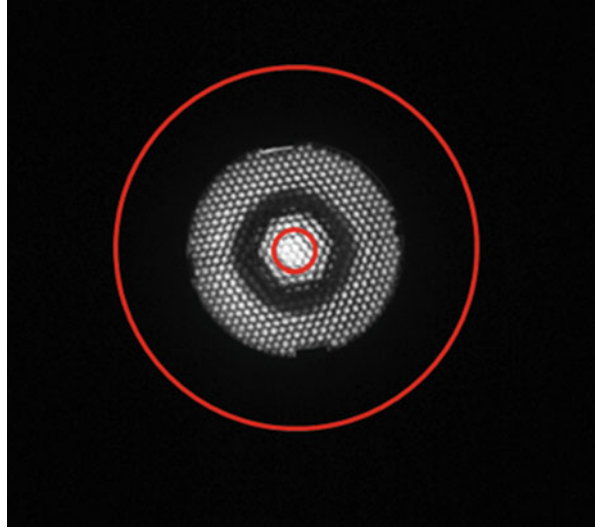
Note: t must be input in seconds to calculate Y_{ph} in radian

Retinal blue-light exposure can be estimated using the ICNIRP guidelines. A quantity called the blue-light weighted radiance L_B can be determined as a function of the spectral radiance L_λ of the light source and the action spectrum $B(\lambda)$, λ being the wavelength:

$$L_B = \sum_{380}^{1400} L_\lambda \times B(\lambda) \times \Delta\lambda \tag{1}$$

L_B is expressed in $W.m^{-2}.sr^{-1}$. As stated in the previous section, the natural movements of the eye tend to smear the retinal image of source over a wider effective area. The phenomenon is taken into account in the definition of the blue-light radiance L_B . This is the reason why the source radiance should be spatially averaged over an effective field of view (FOV) angle which varies as a function of the exposure duration t . The effective FOV angle is given in Table 1. The smallest and largest FOV angles defined in the ICNIRP guidelines are, respectively, 11 and 110 mrad. These values correspond to a retinal image of 190 μm and 1.9 mm,

Fig. 7 Image of an LED source observed at a distance of 200 mm. The *largest circle* shows a 110 mrad effective field of view corresponding to an exposure longer than 10, 000 s. The *smallest circle* shows an 11 mrad effective field of view corresponding to an exposure of 100 s or less



respectively. Figure 7 gives an example of the smallest and largest effective field of view including an LED light source. In the case of the largest field of view, the blue-light radiance is less than the true radiance of the light source since the field of view includes non-emitting areas (black zones).

ICNIRP defines the blue-light effective radiance dose ($\text{J.m}^{-2}.\text{sr}^{-1}$) as the time integral of the blue-light radiance over the duration of the exposure. For constant light sources, this dose is simply expressed by:

$$D_B = L_B \times t \quad (2)$$

The exposure limits (EL) set by ICNIRP are the following:

- For an exposure duration t greater than 0.25 s (aversion response) but less than 10,000 s (approximately 2.8 h), the exposure limit is expressed in term of radiance dose:

$$D_B^{EL} = 1 \times 10^6 \text{ J. m}^{-2}.\text{sr}^{-1} \quad (3)$$

This radiance dose exposure limit is equivalent to a blue-light radiance exposure limit:

$$L_B^{EL} = \frac{(1 \times 10^6)}{t} \text{ W.m}^{-2}.\text{sr}^{-1} \quad (4)$$

- For an exposure duration greater than 10,000 s, the exposure limit is expressed in term of blue-light radiance:

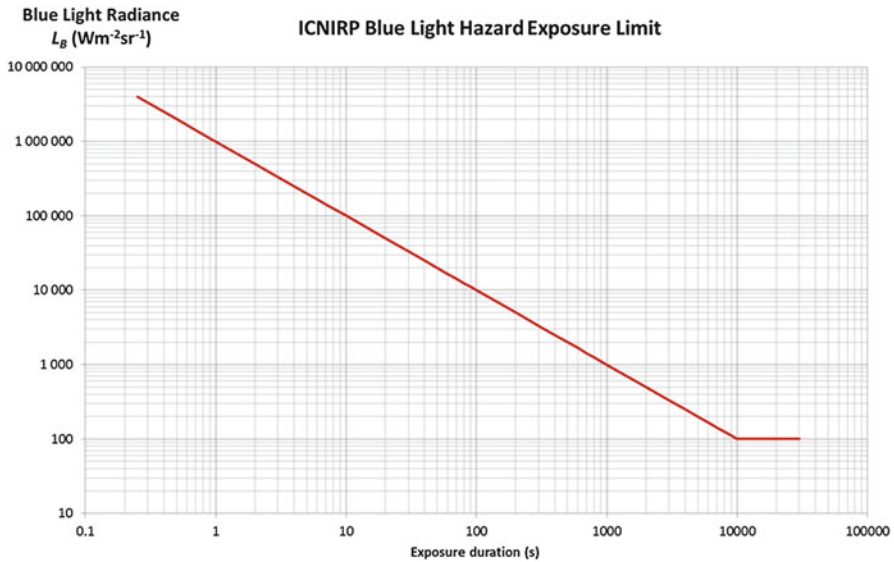


Fig. 8 Blue-light hazard exposure limit defined in ICNIRP (2013) in terms of blue-light weighted radiance

$$L_B^{EL} = 100 \text{ W.m}^{-2}.\text{sr}^{-1} \tag{5}$$

Figure 8 shows a graph of the ICNIRP blue-light exposure limit expressed in terms of blue-light radiance:

For small sources, corresponding to an angular subtense less than 11 mrad, it is possible to express the exposure limit in terms of blue-light hazard irradiance E_B . This case is called the “small source regime.” The blue-light weighted irradiance E_B can be determined as a function of the spectral irradiance E_λ of the light source and the action spectrum $B(\lambda)$:

$$E_B = \sum_{380}^{1400} E_\lambda \times B(\lambda) \times \Delta\lambda \tag{6}$$

Using simple photometric consideration, it can be showed that the small source irradiance is given by the ratio of the radiance to a factor of about 10^4 . Because of the eye movements involved in normal visual tasks, the maximum exposure duration that needs to be considered for small sources is 100 s. For this reason, the small source limit is constant for exposure durations longer than 100 s. Therefore, the blue-light exposure limit can be expressed in terms of irradiance as follows:

- For an exposure duration t greater than 0.25 s and less than 100 s:

$$E_B^{EL} = \frac{100}{t} \text{ Wm}^{-2} \quad (7)$$

- For an exposure duration greater than 100 s and less than 30,000 s:

$$E_B^{EL} = 1 \text{ Wm}^{-2} \quad (8)$$

Regulations on Personal Exposure to Optical Radiations Emitted by Artificial Light Sources

The exposure limit values (ELV) given in the ICNIRP guidelines are internationally accepted. In some regions such as the EU and the USA, they are transposed in regional and national regulatory documents. For example, in the USA, the American Conference of Governmental Industrial Hygienists (ACGIH) has set similar ELVs (ACGIH 2001). In the EU, the Directive 2006/25/EC on artificial optical radiations (EC 2006) requires limiting the personal exposure to artificial optical radiations at the workplace. In this regulation, the ELVs set by ICNIRP become mandatory limits that must not be exceeded for the workers. As far as personal exposure to LEDs is concerned, the blue-light hazard is included in the scope of the EU Directive. Therefore, employers should assess that workers are not exposed to levels in excess of the exposure limit values. Employers may be able to demonstrate this by using several means: generic assessments, theoretical assessments, or measurements. The directive itself does not specify a methodology. However, a number of standards were published to assist with verification of the compliance.

Assessment of Personal Exposure to Optical Radiations Emitted by Artificial Light Sources

The personal exposure to artificial optical radiation can be assessed by performing a comparison with the exposure limit values. The general methodology is to perform an assessment of the optical radiation emitted by all the artificial sources that may be incident on the human body in a given occupational scenario. In the EU, the EN 14255–2 standard (CEN 2005) describes the practical methodology used for visible and infrared radiation emitted by artificial sources in the workplace. This standard is applicable to any type of artificial sources. The methodology of EN 14255–2 relies on a work task analysis and the practical measurement of the exposure. This standard methodology is used by European health authorities to control the conformity of workplaces.

In the case of artificial lighting, this standard is applicable to the human exposure to a whole lighting installation, comprising all the lamps and luminaires emitting optical radiation toward the worker.

Photobiological Safety Standards for Lighting Products

The standard assessment of the photobiological safety of a whole installation as described in the previous section is not well adapted to the evaluation of the intrinsic safety of a single lighting product such as a lamp or a luminaire. Following the example of laser safety classes, the lighting industry has helped set up some standards in order to define the potential risks posed by lighting sources. These standards are useful because they provide a classification of a lighting source in different “risk groups.” However, it is important to mention that the notion of “risk group” is only applicable to a single product. The exposure to an installation comprising several lighting sources should be assessed using the guidelines of ICNIRP and the general methodology described in the previous section. In particular, the exposure to several lighting sources classified in a low-risk group does not guarantee that the total exposure is below the exposure limit. In the case of LEDs, which are concerned with the blue-light hazard, there are specific conditions to extend a low-risk classification valid for a single LED to typical installation situations. These conditions will be discussed in the following sections.

The photobiological safety of lamps and devices using lamps, such as luminaires and lighting modules, has been internationally addressed by the Commission Internationale de l’Eclairage (CIE), the Illuminating Engineering Society of North America (IESNA), and the International Electrotechnical Commission (IEC) through close collaborations and joint working groups. They led to the following standards describing the photobiological safety of lamps and lamp systems: joint publication CIE S009 (CIE 2006) and IEC 62471:2006 (IEC 2006) and IESNA/ANSI RP-27 series (IESNA 2000, 2005, 2007). These documents are not identical but similar in content.

The Photobiological Safety Joint Standard CIE S009/IEC 62471:2006

This standard, which concerns the photobiological safety of lamps and devices using lamps, provides a system of classification of the light source in several risk groups. The standard considers all the photobiological hazards listed by the ICNIRP that may affect the skin and the eye (thermal and photochemical hazards) from the ultraviolet to the infrared wavelengths. Guidance is provided to perform the physical measurements (radiance and irradiance) necessary to assess in a laboratory the exposure levels produced by a lighting product.

The standard introduces the notion of risk groups which depend on the duration of the maximum permissible exposure assessed for each type of photobiological hazard: hazards related to actinic UV, hazards related to UV-A, hazards related to blue light (retinal blue-light hazard), and thermal hazards related to visible and infrared radiations.

Four risk groups are defined:

- Risk Group 0 – Exempt group: no photobiological hazard under foreseeable conditions
- Risk Group 1 – Low-risk group: products safe for most use applications, except for very prolonged exposures where direct ocular exposures may be expected

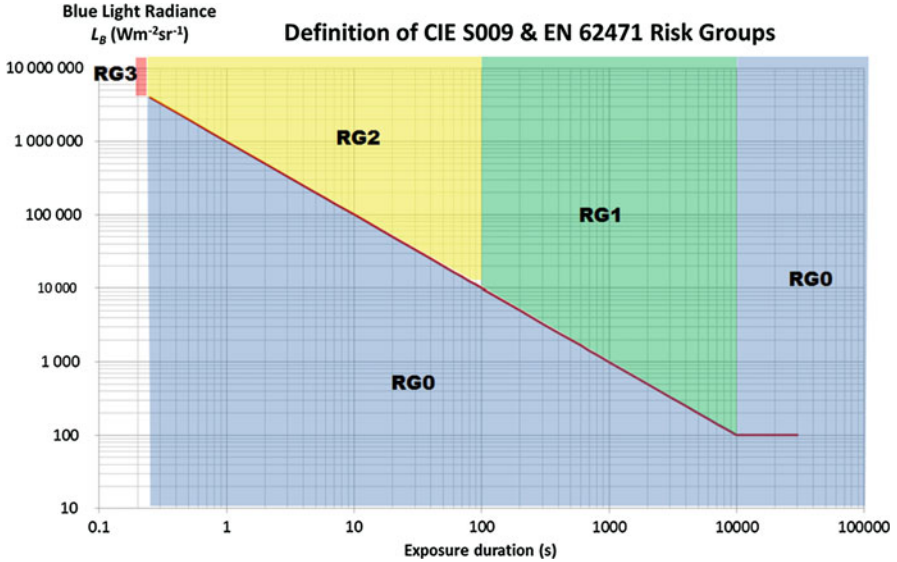


Fig. 9 Blue-light hazard radiance ranges for the defined Risk Groups with reference to the exposure limit defined in ICNIRP (2013) (red line)

- Risk Group 2 – Moderate-risk group: products generally do not pose a realistic optical hazard if the aversion response limits the exposure duration or when lengthy exposures are unrealistic
- Risk Group 3 – High-risk group: products pose a potential hazard even for momentary exposures

In the case of SSL, which is essentially based on LEDs and OLEDs, the current white-light sources only involve the blue-light hazard. In this case, the risk group classification depends on the duration of the maximum permissible exposure of the retina to blue light, as defined by ICNIRP and presented in Fig. 9:

- Risk Group 0: exposure limit is not exceeded within 10,000 s
- Risk Group 1: exposure limit is not exceeded within 100 s
- Risk Group 2: exposure limit is not exceeded within 0.25 s (aversion time)
- Risk Group 3: exposure limit is exceeded within less than 0.25 s

In the case of all types of artificial white-light sources, it is highly unlikely that RG3 is reached for blue-light hazard. This would occur when the blue-light level is a factor of 400 times higher than the RG2 lower limit (i.e., boundary between RG1 and RG2). RG3 for blue-light hazard at a color temperature of 6000 K is only reached when the luminance is above 4×10^9 cd/m² and when the illuminance at the eye is above 400,000 lx (see discussion in section 5.5.3). It should be noted that RG3 is

reached for hazards other than the blue-light hazard (IEC 2012) for non solid-state lighting sources.

IEC 62471 defines two different criteria to determine the viewing distance. Light sources used in general lighting should be assessed at the distance corresponding to an illuminance of 500 lx. Other types of light sources should be assessed at a fixed distance of 200 mm.

For LED components, there is no ambiguity in the distance since LED components are not used per se in general lighting. In this case, IEC 62471 requires using the distance of 200 mm.

However, the choice of the viewing distance in IEC 62471 is sometimes ambiguous and not realistic in the context of the real usage conditions. For instance, in stage lighting (theaters, concert halls, etc.), workers are exposed to an illuminance level higher than 500 lx. Sports participants in stadia lit for television coverage are also exposed to levels much greater than 500 lx for extended periods of time. Thus, applying the 500 lx criterion would underestimate the exposure, while the 200 mm criterion would greatly overestimate it. In a more common context, directional household lamps are supposed to be assessed using the 500 lx criterion, which corresponds to a typical viewing distance of a few meters. It is, however, quite common to have shorter viewing distances, as short as 200 or 500 mm at home. Another example is street lighting where the illuminance level is much lower than 500 lx, typically of a few lx. Assessing the exposure to blue light emitted by a street lighting luminaire at the distance giving an illuminance of 500 lx is clearly not appropriate.

IEC issued two technical reports to provide guidance to manufacturers of non-laser light sources when assessing and reporting the photobiological safety of their products. These documents are the IEC TR 62471–2 issued in 2009 (IEC 2009) and IEC TR 62778, issued in 2012 (IEC 2012).

The Technical Report IEC TR 62471–2

This technical report provides the basis for safety requirements dependent on the risk group classification of IEC 62471. A labeling scheme is provided according to the risk group of the light source, as shown in Table 2.

The technical report IEC TR 62471–2 recommends that products should be labeled, exhibiting the risk group of blue-light hazard when assessed to be RG2 or RG3. Furthermore, for all products in excess of the exempt group (RG0), the document recommends the manufacturer to provide the following user information:

- (a) Clear statement that the lamp or lamp system is in excess of the Exempt Group and that the viewer-related risk is dependent upon how the users install and use the product.
- (b) The most restrictive optical radiation hazard and other optical radiation hazards in excess of Exempt Group (see Table 2).
- (c) Exposure values and the hazard distances with optional graphical presentation of distance-dependent exposure values.

Table 2 Reproduced from IEC (2009). Hazard-related risk group labeling of lamp systems presented in IEC TR 62471–2

Hazard	Exempt risk group	Risk group 1	Risk group 2	Risk group 3
Ultraviolet hazard 200 nm to 400 nm	Not required	NOTICE UV emitted from this product	CAUTION UV emitted from this product	WARNING UV emitted from this product
Retinal blue-light hazard 300 nm to 400 nm	Not required	Not required	CAUTION Possibly hazardous optical radiation emitted from this product	WARNING Possibly hazardous optical radiation emitted from this product
Retinal blue-light or thermal hazard 400 nm to 780 nm	Not required	Not required	CAUTION Possibly hazardous optical radiation emitted from this product	WARNING Possibly hazardous optical radiation emitted from this product
Cornea/lens infrared hazard 780 nm to 3000 nm	Not required	NOTICE IR emitted from this product	CAUTION IR emitted from this product	WARNING IR emitted from this product
Retinal thermal hazard, weak visual stimulus 780 nm to 1400 nm	Not required	WARNING IR emitted from this product	WARNING IR emitted from this product	WARNING IR emitted from this product

- (d) Hazard distances for all relevant viewer-related risk groups below the assigned one.
- (e) Adequate instructions for proper assembly, installation, maintenance, and safe use, including clear warnings concerning precautions to avoid possible exposure to hazardous optical radiation.
- (f) Advice on safe operating procedures and warnings concerning reasonably foreseeable malpractices, malfunctions, and hazardous failure modes. Where maintenance procedures are detailed, they should, wherever possible, include explicit instructions on safe procedures to be followed.
- (g) Reproduction of the required labeling and an explanation of its meaning shown in Table 2.
- (h) Information on what type of user controls may be considered.

Table 3 is given in IEC TR 62417–2 to explain the labeling information and to provide guidance on control measure.

The technical report also suggests a procedure for the allocation of safety measures. This is done through the assessment of the light source at a distance of 200 mm and the applicable risk group exposure duration.

Table 3 Reproduced from IEC (2009). Explanation of labeling information and guidance on control measures

Hazard	Exempt risk group	Risk group 1	Risk group 2	Risk group 3
Ultraviolet hazard 200 nm to 400 nm	Not required	Minimize exposure to eyes or skin. Use appropriate shielding	Eye or skin irritation may result from exposure. Use appropriate shielding	Avoid eye and skin exposure to unshielded product
Retinal blue-light hazard 300 nm to 400 nm	Not required	Not required	Do not stare at operating lamp. May be harmful to the eyes	Do not look at operating lamp. Eye injury may result
Retinal blue-light or thermal hazard 400 nm to 780 nm	Not required	Not required	Do not stare at operating lamp. May be harmful to the eyes	Do not look at operating lamp. Eye injury may result
Cornea/lens infrared hazard 780 nm to 3000 nm	Not required	Use appropriate shielding or eye protection	Avoid eye exposure. Use appropriate shielding or eye protection	Avoid eye exposure. Use appropriate shielding or eye protection
Retinal thermal hazard, weak visual stimulus 780 nm to 1400 nm	Not required	Do not stare at operating lamp	Do not stare at operating lamp	Do not stare at operating lamp

However, when a lamp is integrated into another product, these assessment conditions may become nonrepresentative. In this case, the product may be assessed at the minimum distance and maximum exposure duration representative of the application-specific conditions of foreseeable access (viewer-related risk).

The applications can be divided into three groups, according to the likelihood of the viewing of the source:

- Unintentional short term : automotive, spot, flash, and projection
- Intermittent, occasional (or possible) short term: many toys, where the normal attention span of a child is short, laboratory equipment, home, and signaling
- Intentional (or likely) long term: displays and general lighting systems

When a product is assessed under application-specific conditions, this viewer-related risk group classification may differ from the risk group of the lamp incorporated into the product. IEC TR 62471–2 provides guidance on the maximum

Table 4 Reproduced from IEC (2009). Maximum acceptable risk group of products assessed for viewer-related risk under application-specific conditions

Risk group of the lamp system	Risk group assessed under application-specific conditions – viewer-related risk		
	Unintentional short term	Intentional short term	Intentional (or likely) long term
Exempt Risk Group	Exempt Risk Group	Exempt Risk Group	Exempt Risk Group
Risk Group 1	Risk Group 1	Risk Group 1	Exempt Risk Group – exposure limited by access distance or by controlled access
Risk Group 2	Risk Group 2	Risk Group 1 – exposure limited by access distance or/and exposure duration or product used in restricted location	Exempt Risk Group – exposure limited by access distance or by controlled access
Risk Group 3	Risk Group 2 – exposure limited by access distance or product used in restricted location	Risk Group 1 – exposure limited by access distance or/and exposure duration or product used in restricted location	Exempt Risk Group – exposure limited by access distance or by controlled access

permissible risk group of products accessible under application-specific conditions, as shown in Table 4.

For instance, if a Risk Group 2 lamp is incorporated into a lighting source (intentional long-term exposure), it is only acceptable if the viewer-related risk group of the lighting source is exempt (exposure limited by access distance or by controlled access).

If a Risk Group 3 lamp is incorporated into signaling equipment (intentional short-term exposure), it is only acceptable if the viewer-related Risk Group of the signaling equipment is maximum Risk Group 1 – foreseeable exposure is controlled by access distance and/or maximum exposure time.

The lighting industry has experienced some difficulties with the implementation of IEC TR 62471–2 when it is applied to the transfer of LEDs risk group to finished products. The procedure of Table 4 is based on worst-case conditions, not reflecting the real use of the LED. This procedure often results in RG2 classification, requiring the use of warning labels. The issue of transferring the risk group of LED components to finished products is more precisely addressed in the technical report IEC TR 62778.

The Technical Report IEC TR 62778

This technical report was published in order to clarify some ambiguities present in IEC 62471 when assessing the blue-light hazard of light sources and luminaires. It provides guidance on how to transfer the photobiological safety information of IEC

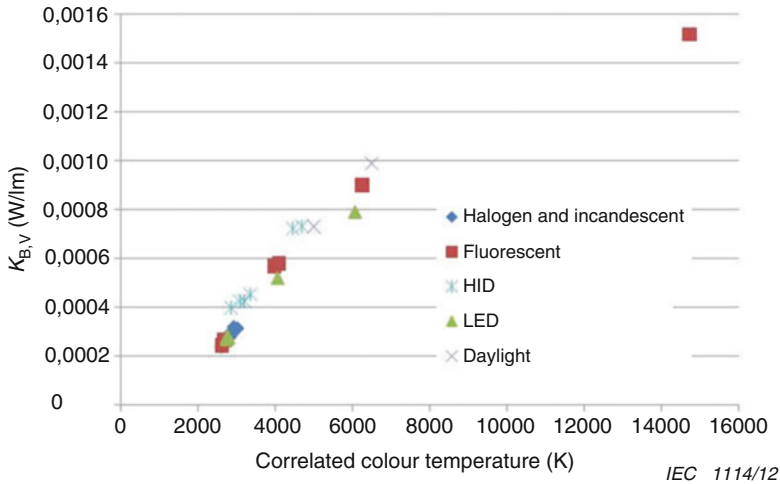


Fig. 10 Reproduced from IEC (2012). Blue-light hazard efficacy versus CCT for several white-light sources

62471 from components (for instance, LED package, LED module, LED lamp) to a higher-level lighting product (luminaire).

The technical report firstly demonstrates quite a strong correlation between the blue hazard efficacy and the correlated color temperature (CCT) of a white-light source. The blue-light hazard efficacy $K_{B,v}$ is defined as the ratio between the blue-light hazard radiance L_B and the photopic luminance (visual luminance) L , expressed in cd.m^{-2} . The blue-light hazard efficacy is expressed in W.lm^{-1} . Figure 10 shows a graph representing the blue-light hazard efficacy as a function of CCT.

This fact was also emphasized in the ANSES report (ANSES 2010). Figure 11 shows the data used in the ANSES report, obtained with a larger set of white LEDs.

This correlation is used in the IEC TR 62778 document to assess the blue-light weighted quantities (blue-light weighted radiance in $\text{W.m}^{-2}.\text{sr}^{-1}$ and irradiance in W.m^{-2}) as a function of photometric quantities (visual luminance in cd.m^{-2} and illuminance in lx).

The technical report presents an estimation of the light levels necessary to reach the exposure limits with an exposure of 100 s. The 100 s exposure corresponds to the boundary between RG1 and RG2. According to Eq. 3, the limiting blue-light radiance L_B is equal to $10,000 \text{ W.m}^{-2}.\text{sr}^{-1}$ in this case.

This estimation is performed in terms of luminance levels and illuminance levels in the case of small sources (see section 5.2.2). The result of this approach is illustrated in Figs. 12 and 13. These graphs show the luminance and illuminance corresponding to the blue-light exposure limits at 100 s, the boundary between RG1 and RG2.

As explained in the technical report, the small source regime presents a “worst case” in terms of source luminance. Knowing the E_B value at a certain illuminance level essentially gives the maximum exposure duration regardless of the luminance.

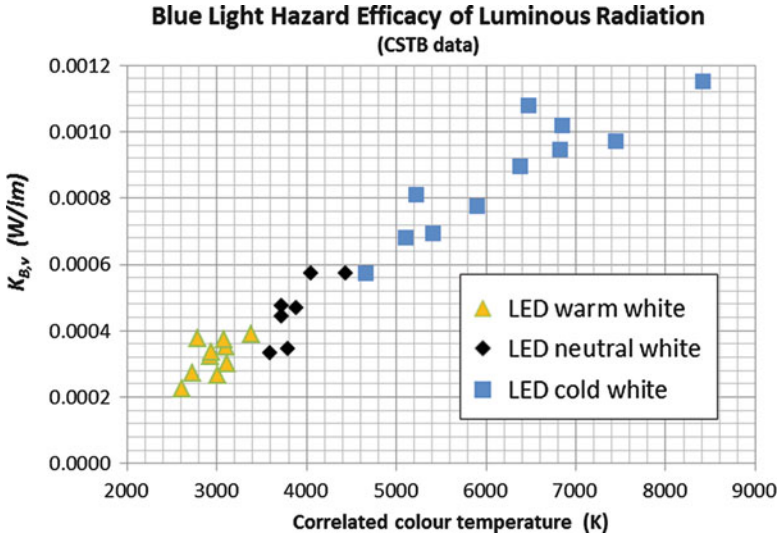
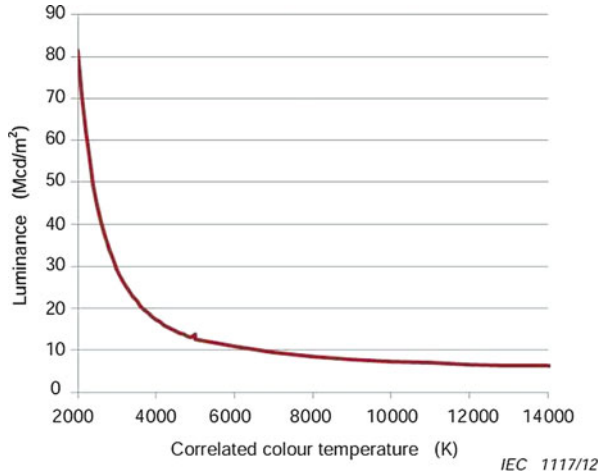


Fig. 11 Reproduced from ANSES (2010). Blue-light hazard efficacy versus CCT for a set of white LEDs classified as warm white, neutral white, and cold white

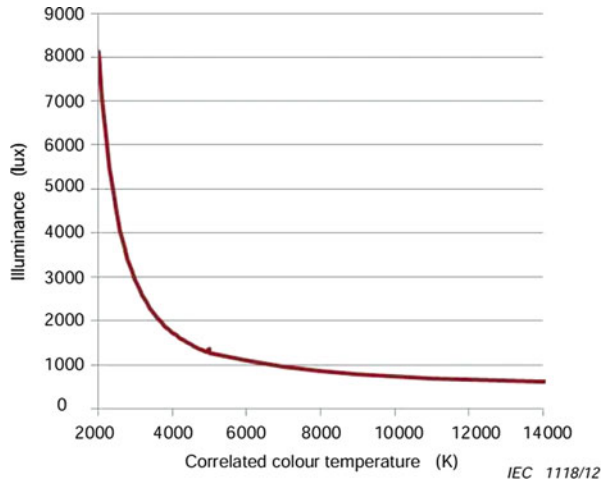
Fig. 12 Reproduced from IEC TR 62778 (IEC 2012). Estimate of the luminance level where $L_B = 10,000 \text{ W} \cdot \text{m}^{-2} \cdot \text{sr}^{-1}$, the boundary between RG1 and RG2, as a function of CCT



It means that if the illuminance level at the viewer’s eye position is well below the illuminance where $E_B = 1 \text{ Wm}^{-2}$ (the red line of Fig. 13), the maximum exposure duration cannot be below 100 s, regardless of the luminance of the light source.

Figure 12 also reveals that the 500 lx level is below the red line throughout the CCT range relevant for general lighting. In other words, *the 500 lx criterion can never generate a RG2 classification for white light.*

Fig. 13 Reproduced from IEC TR 62778 (IEC 2012). Estimate of the illuminance level where $E_B = 1 \text{ W.m}^{-2}$, the boundary between RG1 and RG2 for small sources, as a function of CCT



Another important conclusion drawn in IEC TR 62778 can be inferred from Fig. 12. The large source regime is valid at short distances, and radiance is a light source property independent of viewing distance. If a light source has a blue-light radiance L_B less than $10,000 \text{ W.m}^{-2}.\text{sr}^{-1}$, it will have a maximum permissible exposure duration greater than 100 s even with the shortest viewing distances. At longer distances, where it would pass from the large source to the small source regime, the maximum permissible exposure duration can only increase and never decrease. Therefore, if a light source has an L_B less than $10,000 \text{ W.m}^{-2}.\text{sr}^{-1}$ (i.e., its luminance lies below the red line in Fig. 12), it cannot be in RG2 no matter at what distance it is evaluated. It follows that whenever either of the two conditions is fulfilled, then a classification greater than RG1 is not possible. In order to give rise to a RG2 situation, both the luminance of the light source and the illuminance at the viewer’s eye have to be above a limiting value. In all other situations, the risk group is RG1 maximum, whatever the size and the viewing distance.

Based on Figs. 12 and 13, the technical report IEC TR 62778 proposes some recommendations to assist the consistent application of IEC 62471 for the assessment of blue-light hazard of light sources and luminaires. These recommendations are particularly relevant to LEDs and SSL products.

For large sources, having an angular subtense greater than 11 mrad, a measurement of spectral radiance needs to be performed at 200 mm in a FOV of 11 mrad in order to obtain a value of the blue-light radiance L_B which will be compared to the RG1 exposure limit of $10,000 \text{ W.m}^{-2}.\text{sr}^{-1}$. If the result is below this exposure limit, then the classification “RG1 unlimited” can be applied to the light source and higher products incorporating this light source. If the RG1 limit is exceeded, the light source is RG2 and there is a possibility that the final product will also be RG2. IEC TR 62778 then recommends determining the boundary between RG1 and RG2, expressed in terms of a threshold illuminance E_{thr} at which the boundary occurs.

The threshold illuminance should be included in the datasheet for transfer to the final product. In the case of a finished product, the threshold illuminance can be converted to a threshold distance d_{thr} corresponding to the boundary between RG1 and RG2. The recommended method to perform this conversion is to use goniophotometric data to identify the maximum luminous intensity. The inverse square law can be used to determine the minimum threshold distance. If goniophotometric data is not available, no guidance is provided but it seems possible to use an illuminance meter to experimentally find E_{thr} .

When the blue-light radiance of the large light source is below the RG0 exposure limit of $100 \text{ W}\cdot\text{m}^{-2}\cdot\text{sr}^{-1}$, then the light source is classified “RG0 unlimited.” The RG0 can thus be transferred to any type of luminaire using this light source.

For small sources, having an angular subtense less than 11 mrad, the measurement FOV can be reduced so that it underfills the light source. In this case, a blue-light radiance value L_B is obtained, and the assessment can be performed, yielding RG0, RG1, or RG2 classification.

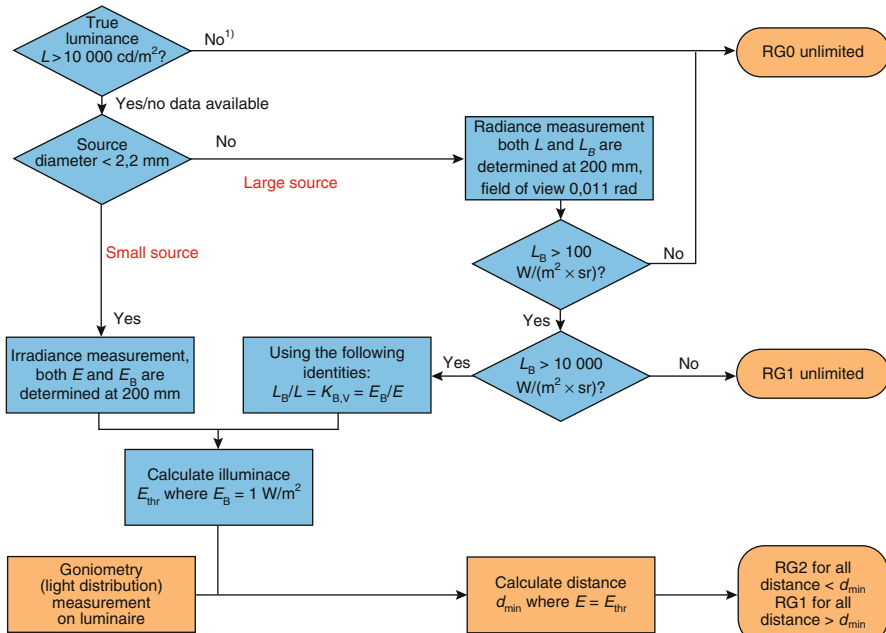
If the measurement is performed as an irradiance measurement, the resulting blue-light irradiance E_B should be compared with the RG1 exposure limit of $1 \text{ W}\cdot\text{m}^{-2}$. If E_B is above this limit, then the light source is RG2, and the boundary between RG1 and RG2 should be reported in terms of the threshold illuminance E_{thr} and the threshold distance d_{thr} . If E_B is below this limit, then the small light source can be classified RG1, but the risk group cannot be transferred to a finished product (this is due to the treatment of the light source as a point source where the inverse square law dictates that there will be a threshold distance where the exposure limit will be exceeded). In this case, the worst case is assumed (RG2) and the threshold illuminance should be reported to allow transfer to the final product.

The general methodology described by the IEC TR 62778 to transfer the blue-light hazard assessment from the primary light source to the luminaires including this light source is illustrated in Fig. 14.

Table 5 gives luminance values giving risk group not greater than RG0 and RG1, whatever the viewing distance and the source size. If the true luminance of the light source is greater than the values of Table 5, but the illuminance at the viewing distance complies with the values of Table 6 for the given correlated color temperatures (CCT), then its classification will not be greater than RG0 or RG1 at the considered viewing distance.

The methodology of IEC TR 62778 is more accurate than the one presented in IEC TR 62471–2 for the transfer of the risk group of LED components to a higher product which includes them. Tables 5 and 6 define situations of RG0 and RG1 classification not requiring radiance or irradiance measurement. This allows a luminaire manufacturer to make sure that a product incorporating such LED components will not be RG2 and therefore will not require the labeling and user information listed in IEC TR 62471–2.

In the case of RG2 LEDs, the specification of the threshold illuminance E_{thr} in the LED datasheet allows the luminaire manufacturer to define use conditions giving a RG0 or RG1 classification for the final product, without having to perform full



IEC 1120/12

1) The RG0 result following from the $\leq 10\,000\text{ cd/m}^2$ condition is only valid for white light sources.

Fig. 14 Reproduced from IEC TR 62778 (IEC 2012). Flow chart describing the flow of information from the primary light source (in blue) to the luminaire based on this light source (in amber)

Table 5 Luminance values giving risk group not greater than RG0 and RG1

Rated CCT	RG1 Luminance limits (cd m^{-2})	RG0 Luminance limits (cd m^{-2})
$\text{CCT} \leq 2350\text{ K}$	4×10^7	4×10^5
$2350\text{ K} < \text{CCT} \leq 2850\text{ K}$	1.85×10^7	1.85×10^5
$2850\text{ K} < \text{CCT} \leq 3250\text{ K}$	1.45×10^7	1.45×10^5
$3250\text{ K} < \text{CCT} \leq 3750\text{ K}$	1.1×10^7	1.1×10^5
$3750\text{ K} < \text{CCT} \leq 4500\text{ K}$	8.5×10^6	8.5×10^4
$4500\text{ K} < \text{CCT} \leq 5750\text{ K}$	6.5×10^6	6.5×10^4
$5750\text{ K} < \text{CCT} \leq 8000\text{ K}$	5×10^6	5×10^4

photobiological testing. The knowledge of the threshold distance is extremely useful for luminaires including RG2 LEDs as it can be compared with the minimum viewing distance that is expected in the use of the luminaire. Protection and control measures can therefore be implemented. This also assists with the determination of the threshold illuminance when multiple luminaires are used (including taking into consideration any variations in color temperatures).

Table 6 Illuminance values giving risk group not greater than RG0 and RG1

Rated CCT	RG1 Illuminance limits (lx)	RG0 Illuminance limits (lx)
$CCT \leq 2350 \text{ K}$	4000	40
$2350 \text{ K} < CCT \leq 2850 \text{ K}$	1850	18.5
$2850 \text{ K} < CCT \leq 3250 \text{ K}$	1450	14.5
$3250 \text{ K} < CCT \leq 3750 \text{ K}$	1100	11.0
$3750 \text{ K} < CCT \leq 4500 \text{ K}$	850	8.5
$4500 \text{ K} < CCT \leq 5750 \text{ K}$	650	6.5
$5750 \text{ K} < CCT \leq 8000 \text{ K}$	500	5.0

Limitations of the CIE/IEC Blue-Light Hazard Assessment

Potential Effects of Low-Level Chronic Exposures

The maximum exposure limits defined by the ICNIRP and used to define the risk groups in IEC 62471 are not appropriate for repeated exposures to blue light as they were calculated for a maximum exposure of one eight-hour day.

The effects of chronic and repeated low-dose exposures to visible light emitted by LEDs are currently being investigated by ophthalmologists, physicians, and photobiologists (I.Jaadane et al. 2015). Researchers are working on identifying the mechanisms of retinal damage caused by chronic exposures of rats to low-intensity LED lighting. The objective is to detect the death of retinal cells well before the retina is bleached, which is the visible signature of retinal damage, observed by fundoscopy.

The first published results show that retinal damage induced by chronic exposure to white LEDs can be detected at much lower levels than the ICNIRP exposure limits.

Potential Effects of Long-Term Exposures

The ICNIRP exposure limit values do not take into account the possibility of an exposure over an entire lifetime. Very little is known about the effects of life-long cumulated exposures to blue light emitted by LEDs. According to the Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR) of the European Commission (SCENIHR 2012), no evidence was found indicating that blue light from artificial lighting belonging to Risk Group 0 would have any impact on the retina graver than that of sunlight. The SCENIHR states that IEC 62471 gives limits that are protective against acute effects, while long-term effects are only marginally considered and estimated to be of negligible or small risk.

Sensitive Populations

IEC 62471 does not take into account the sensitivity of certain specific population groups, which can be characterized by an accrued sensitivity to visible light:

- People having preexisting eye or skin condition for which artificial lighting can trigger or aggravate pathological symptoms
- Aphakic and pseudophakic subjects who consequently either cannot or can only insufficiently filter short wavelengths (particularly blue light)
- Children, as their skin and visual system are not mature
- Elderly people as their skin and eyes, particularly the retina, are more sensitive to optical radiation

A general recommendation to these sensitive populations is to use lamps and luminaires emitting a small amount of short-wavelength light, which are characterized by a low CCT (warm-white light for instance).

The photobiological standards relative to lighting systems should be extended to cover children – especially infants less than 2 year old – and aphakic or pseudophakic individuals, taking into account the corresponding phototoxicity curve $A(\lambda)$ published by the ICNIRP in its guidelines.

Blue-Light Hazard Exposure Data Concerning LEDs and SSL Products

Since 2009, blue-light exposure data concerning LED have been provided by LED manufacturers and professional lighting associations (ELC and CELMA for instance) but also by independent laboratories and governmental agencies (ANSES and SCENHIR for instance). It was found that the blue-light weighted radiance L_B produced at a distance of 200 mm from the user by a significant number of blue and cold-white LEDs (bare LEDs and LEDs equipped with a focusing lens) exceeds the exposure limits set by ICNIRP for an exposure duration comprised between a few seconds for high-power blue LEDs to a few tens of seconds for high-power cold-white LEDs, making them classified as RG2.

For example, Fig. 15 extracted from ANSES (2010) shows the variations of the blue-light weighted radiance of cold-white LEDs as a function of the exposure duration. Table 7, also extracted from ANSES (2010), confirms that cold-white LEDs are the most critical white-light LED sources for the blue-light hazard. These values are consistent with the boundary values of IEC TR 62778 presented in Fig. 12.

The potential toxicity of some LED components viewed at short distances cannot be neglected. However, when the viewing distance is increased to one meter, the maximum permissible exposure duration rapidly increases to a few thousands of seconds, up to a few tens of thousands of seconds. These very long exposure durations provide a reasonable safety margin to assert that there is virtually no possible blue-light retinal damage from LEDs at longer viewing distances (statement valid for state-of-the-art LEDs at the time of writing this text).

Several classes of products and applications based on RG2 bare LEDs or LEDs covered by a focusing lens (collimator) can potentially create a high level of retinal

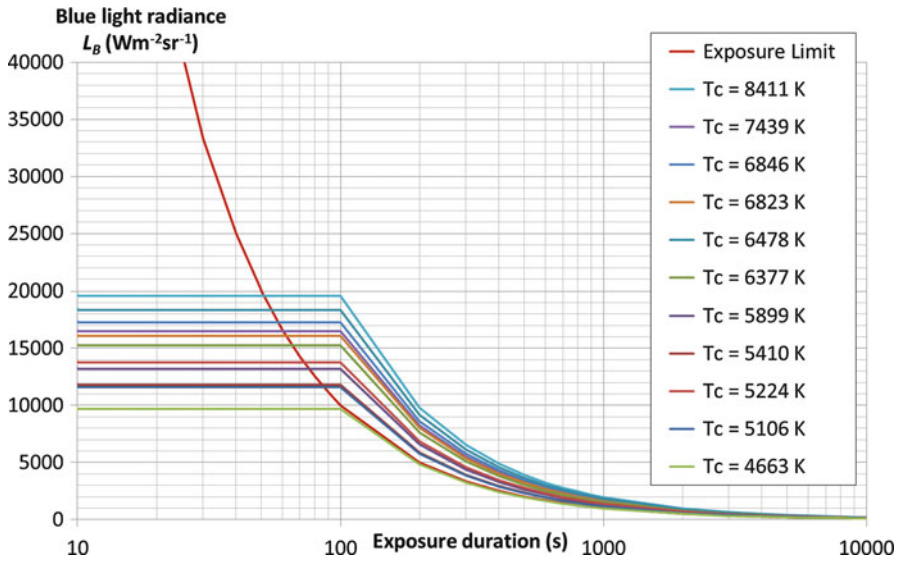


Fig. 15 Adapted from ANSES (2010). Blue-light radiance of single chip cold-white LEDs emitting a luminous flux of 200 lm, measured at a distance of 200 mm. The exposure limits are exceeded for exposure durations between 50 s and 100 s, corresponding to an RG2 classification

Table 7 Reproduced from ANSES (2010). Results of the IEC 62471 blue-light hazard assessment carried out on selected single chip high brightness white LEDs

	Luminous flux (lm)	Luminance (cd.m ⁻²)	Maximum permissible exposure duration	IEC 62471 Risk Group
Cold white	100	1.6×10^7	∞ (exposure limit is never exceeded)	RG0 (no risk)
	200	3.2×10^7	50 s to 100 s (according to correlated color temperature)	RG2 (moderate risk)
Neutral white	100	1.6×10^7	∞ (exposure limit is never exceeded)	RG0 (no risk)
	200	3.2×10^7	100 s to 10,000 s (according to correlated color temperature)	RG1 (low risk)
Warm white	100	1.1×10^7	∞ (exposure limit is never exceeded)	RG0 (no risk)
	200	2.2×10^7	(exposure limit is never exceeded)	RG0 (no risk)

blue-light exposure when short-viewing distances are possible. Examples are (but are not limited to):

- Testing and adjustments of high-power blue and cold-white LEDs by operators in lighting-manufacturing facilities or by lighting installers.
- Toys using LEDs. Children are a sensitive population to blue-light retinal exposure. Analysis similar to IEC TR 62778 using the aphakic action spectrum $A(\lambda)$, instead of the phakic action spectrum $B(\lambda)$, should be conducted.

- Automotive LED daytime running lights when activated near children and other sensitive people.
- LED lamps sold for home applications (consumer market) where situations may potentially occur that lamps are viewed at distances as short as 200 mm.

The conclusion drawn above cannot be extended to all SSL applications because, as explained in the previous sections, the photobiological safety of a final SSL product cannot always be assessed from its LED components. The L_B value of an SSL product can be different from the L_B value of the LED components that it uses. For instance, a higher L_B can be obtained with a luminaire using an array of low L_B small source LEDs. The case of LED arrays and clusters would necessitate a specific risk assessment method that is not yet included in the current version of IEC TR 62778. Also, a lower L_B can be obtained for a luminaire using a diffuser in front of a high L_B LED.

As explained in the IEC TR 62778 technical report, it is interesting to note that the strict application of IEC 62471 to LED lamps and luminaires used in general lighting (assessment at the distance corresponding to an illuminance level of 500 lx) always leads to an RG0 or RG1 classification, similar to traditional indoor light sources (fluorescent lamps, incandescent, and halogen lamps).

Nevertheless, when the 200 mm viewing distance was used, several measurement campaigns such as ANSES (2010) revealed that only a few indoor LED lamps and luminaires belonged to RG2, while traditional indoor light sources (fluorescent and incandescent) were always in RG0 or RG1. This result shows that LED technology potentially raises the blue-light risk in home applications where the viewing distance is not limited and light sources are accessible to children and other sensitive people. At the time of this writing, the general public is not aware of potential risks for the eye. It is expected that the application of the labeling system of IEC TR 62471–2 (warning in the case of RG2) will become mandatory in some economies. At the time of preparing this document, no mandatory labeling system was in place for RG2 lighting products.

For SSL products aimed at consumer applications (retrofit LED lamps for instance), the blue-light hazard risk group should be limited to RG1 at 200 mm, which can be considered as the minimum viewing distance encountered at home. The measurement campaigns carried out by several laboratories showed that the vast majority of indoor LED lamps and luminaires already comply with this requirement. This suggests that it is not a critical issue for the SSL industry at large.

It is worth noting that other widely used light sources, particularly high-intensity discharge lamps (metal-halide lamps for instance), are also in RG2 and even in RG3 for hazards other than retinal blue-light hazard. However, these lamps are intended for clearly identified uses and can only be installed by professionals who should be aware of the safety distance required to limit the exposure.

Conclusions

In a typical LED, the radiance and luminance levels may be extremely high, much higher than the values found in the case of common lamps used in general lighting, making them more susceptible to producing glare. Glare does not constitute a risk in itself but it is a source of discomfort and temporary visual disability. It can be the indirect cause of accidents.

In indoor lighting, glare is assessed with the index UGR, defined by the CIE. The UGR is not applicable to point sources such as visible LEDs incorporated in a luminaire. However, lighting manufacturers and designers usually perform UGR calculations on SSL luminaires having visible LED point sources but incorrectly considering the average luminance over the whole area of the luminaire. This approach is misleading as the resulting UGR is low and does not reflect the physiological perceived glare. Therefore, the use of UGR should be restricted to SSL products with large diffusers, without any point sources. In all cases, it is recommended to specify the maximum luminance of the SSL finished products. The luminance ratio between the light source and the background should be computed and adapted to each lighting installation according to visual ergonomics criteria.

The blue-light hazard is the only photobiological hazard currently required to be considered in the present SSL technologies, at the exception of LEDs using a UV-emitting semiconducting structure. The blue-light hazard is related to the photochemical damage caused by blue and violet light on the retina.

The blue-light hazard is associated with blue-light retinal irradiance. Due to the high radiance of LEDs, the retinal irradiance levels are potentially high and must be carefully considered. In general, the photochemical damage of the retina depends on the accumulated dose to which the person has been exposed, which can be the result of a high-intensity short exposure but can also appear after low-intensity exposures repeated over long periods. Blue light is recognized as being harmful to the retina, as a result of cellular oxidative stress. Blue light is also suspected to be a risk factor in the age-related macular degeneration.

Retinal blue-light exposure can be estimated using the ICNIRP guidelines. The retinal blue-light exposure levels produced at a distance of 200 mm by blue and cold-white LEDs often exceed the exposure limits after an exposure between a few seconds (blue LEDs) to a few tens of seconds (cold-white LEDs). As a consequence, the short-distance potential toxicity of these LED components cannot be neglected.

However, when the viewing distance is increased beyond one meter, the maximum exposure duration rapidly increases to a few thousands of seconds, even up to a few tens of thousands of seconds. These very long exposure durations provide a reasonable safety margin to assert that there is virtually no possible blue-light retinal damage from LEDs at longer viewing distances (statement valid for state-of-the-art LEDs at the time of writing).

Several usages and applications based on bare LEDs or LEDs associated with a focusing lens (collimator) are directly concerned by a potentially high level of retinal blue-light exposure. Examples are (but are not limited to):

- Testing and adjustments of high-power blue and cold-white LEDs by operators in lighting-manufacturing facilities and by lighting installers.
- Toys using LEDs. Children are a sensitive population to blue-light retinal exposure.
- Automotive LED daytime running lights when activated near children and other sensitive people (not in the scope of this annex).
- LED lamps sold for home applications (consumer market) in which case lamps can be viewed at distances as short as 200 mm.

For all SSL devices (LEDs, LED modules, LED lamps, LED luminaires, etc.), the blue-light hazard risk assessment must be carried out. The main tool is the IEC 62471 standard. It provides a system of classification of light sources in several risk groups, according to the maximum permissible exposure duration assessed at a given viewing distance: Risk Group 0 or Exempt Group (no risk), Risk Group 1 (low risk), Risk Group 2 (moderate risk), and Risk Group 3 (high risk).

IEC 62471 defines two different criteria to determine the viewing distance. Light sources used in general lighting should be assessed at the distance corresponding to an illuminance of 500 lx. Other types of light sources should be assessed at a fixed distance of 200 mm.

For LED components which will be integrated in a higher product, IEC 62471 requires using the distance of 200 mm. The application of the IEC 62471 measurement technique at 200 mm often lead to RG2 classification (moderate risk) for blue and cold-white LEDs.

The choice of the viewing distance in IEC 62471 is sometimes ambiguous and not realistic in the context of real usage conditions. The technical report IEC TR 62778 was published in 2012 to clarify and resolve this ambiguity of IEC 62471 when it is applied to the blue-light hazard assessment of LEDs and SSL devices.

Following the guidelines of IEC TR 62778, LED manufacturers should report the risk group of their component (RG0, RG1, or RG2). According to IEC TR 62778, it is sometimes possible to transfer the risk group of an LED to a higher product that incorporates it. In the case of RG2 devices, it is advised that the manufacturer provides the boundary between RG1 and RG2 with the threshold illuminance and the threshold distance, which can be viewed as a reasonable safety distance. RG2 products should be sold with clear information concerning the threshold distance. Otherwise, RG2 products should be labeled according to IEC TR 62471–2, in order to inform the user “not to stare” at the operating lamp as it may be harmful to the eyes.

For SSL products aimed at consumer applications (retrofit LED lamps for instance), the risk group should be limited to RG1 at 200 mm, which can be considered as the shortest viewing distance encountered at home.

IEC 62471 does not take into account the sensitivity of certain specific population groups, which can be characterized by an accrued sensitivity to visible light:

- People having preexisting eye or skin condition for which artificial lighting can trigger or aggravate pathological symptoms
- Aphakics (people with no crystalline lens) and pseudophakics (people with artificial crystalline lenses) who consequently either cannot or can only insufficiently filter short wavelengths (particularly blue light)
- Children, as their skin and visual system is not mature
- Elderly people as their skin and eyes are more sensitive to optical radiation

The photobiological standards relative to lighting systems should be extended to cover children and aphakic or pseudophakic individuals, taking into account the corresponding phototoxicity curve published by the ICNIRP in its guidelines.

Certain categories of workers are exposed to high doses of artificial light (long exposure times and/or high retinal irradiance levels) during their daily activities (e.g., lighting professionals, stage artists, etc.). Since the damage mechanisms are not yet fully understood, exposed workers should use appropriate individual means of protection as a precautionary measure (glasses filtering out blue light for instance). At the moment, there is no personal protective equipment available against the blue-light hazard resulting from exposure to artificial light sources. Some type of laser goggles designed to filter out blue and green laser lines may be used for this purpose, under the prescription of a qualified occupational hygienist.

New generations of LEDs emitting white light are currently being developed using violet and UV chips. This is the case of “GaN on GaN” LEDs which are now incorporated in several commercially available SSL products. Such devices are very interesting from a color rendering point of view as a more continuous and regular spectrum can be achieved with luminophores excited by shorter wavelength radiation than blue light. The photobiological safety of these LEDs and the products using them should be carefully assessed because of potential residual UV and deep blue radiation in the emission spectrum. The assessment should be conducted for the blue-light hazard and UV hazards as well. A careful examination of the aging of these products should also be conducted as the possible degradation of the luminophores may raise the level of short-wavelength light emission.

Acknowledgment This chapter is adapted from a review on health aspects of SSL products that was performed by the author in the context of the 4E-SSL Annex of the International Energy Agency between 2010 and 2014 (IEA 2014). The author expresses his gratitude to the experts involved in this annex for their contribution to reviewing this text.

References

- ACGIH (2001) TLVs and BEIs; threshold limit values for chemical substances and physical agents; biological exposure indices. ACGIH, Cincinnati
- AFNOR (2013) Ergonomics: ergonomic principles applicable to the lighting of workplaces for visual comfort. French standard NF X 35–103
- ANSES (2010) Health effects of lighting systems using light-emitting diodes (LEDs). Expertise report 2008-SA-0408, ANSES Editions

- ANSI/IESNA (American National Standard Institute/Illuminating Engineering Society of North America) RP27.2-00 (2000) Recommended practice for photobiological safety for lamps – measurement systems – measurement techniques. IESNA, New York
- ANSI/IESNA (American National Standard Institute/Illuminating Engineering Society of North America) RP27.1-05 (2005) Recommended practice for photobiological safety for lamps – general requirements. IESNA, New York
- ANSI/IESNA (American National Standard Institute/Illuminating Engineering Society of North America) RP27.3-07 (2007) Recommended practice for photobiological safety for lamps – risk group classification & labeling. IESNA, New York
- Behar-Cohen F, Martinsons C, Viénot F, Zissis G, Barlier-Salsi A, Cesarini JP, Enouf O, Garcia M, Picaud S, Attia D (2011) Light-emitting diodes (LED) for domestic lighting: any risks for the eye? *Prog Retin Eye Res* 30(4):239–257
- Boulenguez P, Martinsons C, Carré S, Torriglia A, Jaadane I, Chahory S (2014) RETINALED: Une étude in vivo du risque dû à la lumière bleue – vers une meilleure compréhension des pathologies rétinienne et une meilleure estimation du risque. In: Proceedings of the 9th conference of the French Radioprotection Society (SFRP). Limoges, France
- CEN (2005) Measurement and assessment of personal exposures to incoherent optical radiation. Visible and infrared radiation emitted by artificial sources in the workplace. EN 14255–2
- CIE (1995) Discomfort glare in interior lighting. CIE 117
- CIE (2006) Photobiological safety of lamps and lamp systems. CIE S009
- European Commission (2006) Exposure of workers to risks arising from physical agents (artificial optical radiation). Directive 2006/25/EC
- Ham WT Jr, Mueller HA, Sliney DH (1976) Retinal sensitivity to damage from short wavelength light. *Nature* 260:153–155
- ICNIRP (2004) ICNIRP guidelines on limits of exposure to ultraviolet radiation of wavelengths between 180 nm and 400 nm (incoherent optical radiation). *Health Phys* 87(2):171–186
- ICNIRP (2013) ICNIRP guidelines on limits of exposure to incoherent visible and infrared radiation. *Health Phys* 105(1):74–96
- IEA (2009–2014) Energy efficient end use equipment. Solid-State Lighting Annex <http://ssl.iea-4e.org/>
- IEC (2006) Photobiological safety of lamps and lamp systems. IEC 62471
- IEC (2009) Photobiological safety of lamps and lamp systems – part 2: guidance on manufacturing requirements relating to non-laser optical radiation safety. Technical report IEC TR 62471–2
- IEC (2012) Application of IEC 62471 for the assessment of blue light hazard to light sources and luminaires. Technical report IEC TR 62778
- IESNA (2010) IES lighting handbook, 10th edn. IESNA, New York
- I. Jaadane, P. Boulenguez, S. Chahory, S. Carré, M. Savoldelli, L. Jonet; F. Behar-Cohen, C. Martinsons, A. Torriglia, (2015) “Retinal damage induced by commercial Light Emitting Diodes (LED)”, *Free Radical Biology and Medicine* 84:373–384
- Kolb H (2011) Simple anatomy of the retina. The organization of the retina and visual system. website Webvision: The Organization of the Retina and Visual System, <http://webvision.med.utah.edu/book/part-i-foundations/simple-anatomy-of-the-retina>, as consulted on 30 June 2013
- Noell WK, Walker VS, Kang BS, Berman S (1966) Retinal damage by light in rats. *Invest Ophthalmol Vis Sci* 5:450–473
- SCENIHR (2012) Health effects of artificial light, Opinions of the Scientific Committee on Emerging and Newly Identified Health Risks, © European Union, ISSN 1831-4783, ISBN 978-92-79-26314-9
- Shang YM, Wang GS, Sliney D, Yang CH, Lee LL (2014) White light-emitting diodes (LEDs) at domestic lighting levels and retinal injury in a rat model. *Environ Health Perspect* 122:269–276
- Van Norren D, Gorgels T (2011) The action spectrum of photochemical damage to the retina: a review of monochromatic threshold data. *Photochem Photobiol* 87:747–753

Educational Lighting and Learning Performance

Thorbjörn Laike

Contents

State of the Art	899
Conclusions	904
References	905

Abstract

The education of our children is one of the most important assignments for society. The facilities where education is conducted have been seen as one important part of a good educational result. In order to create good facilities the lighting situation has for long been seen as one important factor both concerning daylight and electrical lighting. The chapter describes the development of research that initially only could relate to daylight, but with the technological development within lighting was extended to the study of electrical lighting, first the incandescent bulb, then the fluorescent tubes and today the light emitting diodes (LED). The research has for a long time dealt with the visual conditions, but since the late 70s also non-visual aspects has been taken into consideration, not least after the discovery of the ipRGC in the early 21st century. The research in this area has grown very rapidly since and the knowledge has developed. However, much is still to be done. The research is only in the beginning. Results from recent research shows that there is a potential to create better lighting situations with the LED by varying intensity and spectral distribution. Furthermore the distribution of the light should be taken into consideration, and of course the visual aspect must not be neglected. We should also recognize the importance

T. Laike (✉)

Faculty of Engineering, Department of Architecture and Built Environment, Lund University, Lund, Sweden

e-mail: thorbjorn.laike@arkitektur.lth.se

of daylight and how daylight and electrical light should work together for creating good lighting environments in the school.

Keywords

Ambient lighting • Daylight • Educational lighting • Nonvisual effects

The education of our children is one of the most important assignments for society. In a world where knowledge is more and more important, the educational levels of the inhabitants of a country become a competitive factor. The facilities where education is conducted have been seen as one important part of a good educational result. But what are the important features? The environment where knowledge and skills are taught has have been an area for research since the modern school system in the Western world was introduced. The communicative factor light has been one important environmental factor. Initially, the focus was on daylight, since it was the only main option. In the UK, a book on school architecture was published already in 1874. Daylight was seen as an important factor, and the advice was 20 % glazing area in relation to the floor area of the classroom. Furthermore, the daylight architecture also pleads for the indirect glare-free northern light (Robson 1877). However, during this period, the development of the electrical light was developing very fast, and in relation to this, the new environmental factor was considered. In this virgin area, the first attempts are benighted and few comments or advises have been found. Even when the development of the electrical lighting took off, the open-air school was the dominant idea during the early 1900s until the 1930s. Still the rule of thumb was the 1 to 5 ratio between floor area and window area. It was not until after the Second World War standards for classroom lighting were formed. The levels started at 100 lux in the 1950s, and in the 1970s, the lighting level on the working plane should be 200 lux in the UK. The lighting levels varied for different countries. For example, in the USA, the recommendations were set to 538 lux for regular class work and 1076 lux for instructions at the chalkboard during the 1980s.

Together with the development of the fluorescent tubes and the increase of their quality, there was a trend in classroom design toward windowless classrooms especially in the USA during the 1960s. From an environmental comfort point of view, the possibility to control the environment was essential. From a pedagogical point of view, some theorists even suggested that windows may distract students' attention. However, empirical studies showed that the environment was perceived as less positive than environments with windows among both students and teachers. The scholastic achievement could be kept, but the level of absenteeism was higher in classrooms without windows (Edwards and Torcellini 2002). Today, a revised view has finalized this discussion. For example, Tikkanen (1979) found that the incidence of eye problem was lower when having a normal side-view window in comparison with skylights. Today, we also know that the view of the outside world has an importance for the general well-being as shown by among others Ulrich, already in 1984. Finally, today, the impact of the nonvisual effects of light has been an integrated part of the research on light and lighting, something we will return to.

As shown above, the main focus has been on the visual aspects of light. However, knowledge about nonvisual effects of light has since the 1970s become an important part of the lighting design. Concerning the impact of daylight and artificial lighting, a study was conducted by my own mentor the late professor Rikard Küller who in an early study became an icebreaker (Küller and Lindsten 1992). The study showed that different types of light have a secondary impact on both children's behavior and physiology concerning circannual levels of cortisol. This study was the first to consider that not only short-term effects of light are important, but long-term nonvisual effects need to be taken into account. The study showed different results concerning the circannual cortisol levels depending on the accessibility to daylight, leading to the conclusion that side-view windows were the best solution since children sitting in the windowless environment show an annual pattern of cortisol that were lower during the whole year than their counterparts in classrooms with side-view windows. Furthermore, the children's behavior was also affected.

Since this early study, the knowledge has been much more elaborated, and today we have knowledge on the impact of qualitative aspects of light such as color temperature, lighting distribution, glare, and spectral distribution for the nonvisual effects of light. The details of this are presented elsewhere in this handbook. Also the knowledge of the visual effects of light in relation to education has been further elaborated. In the next section, the knowledge up to date will be presented.

To conclude, we must acknowledge the impact of lighting in the school environment. Light has an impact, as a part of the physical environment, but how big it is, is an often put question. Research within environmental psychology exemplified by a study of my own may shed some light on the issue. The impact of different physical environmental factors on children's well-being was assessed, together with other parameters such as the social environment, individual differences, and activities. One conclusion was that about 15–20 % of the total variance could be attributed to features in the physical environment. Some may say that this is a low figure, but the lighting conditions are a factor that gets the cup run over (Laike 1997). Furthermore, many studies show that young people are more sensitive to physical environmental factors than older people, and this is also of importance (Refs). I will argue that creating good and healthy school environment will gain the society as a whole. Today, we have the means, and the knowledge is growing.

State of the Art

As been described above, the daylight has an immense impact on the physical environment of schools. In the thorough review by Wu and Ng (2003), the growing understanding of the relation between man and daylight is clearly described. It could be concluded that the daylight is the source of light that functions best, even though it could also cause problems such as glare. The distribution of daylight in a room is of utmost importance.

When looking at the international literature in the field of education and lighting, there is an impressive amount of literature. However, the quality of the works varies

a lot, and rigorous studies are few. Furthermore, there is quite a big disagreement among researchers in the application of the presented research results, often due to imprecise descriptions of the design and methods. However, some facts are more clear and undisputed (Higgins et al. 2005).

Regarding visual perception, we know that with increasing illuminance, vision improves and this enhances the ability to perceive (van Bommel and Van den Beld 2004). Another important aspect is the impact of light on the developing visual apparatus. Myopia is on the rise in many countries, especially in developing countries, and studies show that increasing the illuminance level will reduce the incidence of myopia (Chen et al. 2007; Lee et al. 2013). In relation to this, it has been shown that light from LED is perceived brighter than the light from other light sources (Govén et al. 2014). From an application point of view, it could be argued that good LED light sources raise the possibility to arrange educational environments with sufficient amount of light to a low-cost of energy. Looking in a worldwide perspective, research in developing countries shows that access to artificial light not only in school but also in the home environment enhances the possibility for good school results. The implementation of LED together with solar-driven luminaries makes this possible. Interestingly, there is a parallel related to the lighting development in the Western countries for more than hundred years ago when the electrical lighting became more common. Some authors argue that the electrical lighting had an immense impact on the societal development toward a more equal and prosperous society because of the access to cheap electrical lighting (Gamert 1993).

Artificial light needs energy and we need to act sustainable in relation to energy use, and therefore, excessive use of artificial lighting is never a solution. It is necessary to look at the optimal lighting levels. The risk by using a figure for horizontal illuminance is that it will not cover the complete issue of good lighting (Boyce 2004; Rea 2012). Turning to alternatives, relations between the ambient lighting and the horizontal illuminance levels have been used to describe the lighting conditions in classrooms. One study (Govén et al. 2002) assessed the preferred levels of surrounding light when the horizontal illuminance level was fixed. The result showed that with an illuminance level of 500 lux in the horizontal work plane, the most preferred levels in the vertical plane were between 80 and 100 cd/m². This relation has been discussed with practitioners that often use this as a rule of thumb. When using LED light sources, it could be of importance that the luminaries have the possibility to give an indirect light toward the ceiling and the walls with the abovementioned ratio.

The impact of color temperature (CCT) has also been subject to research and shows some clear results indicating that at higher illuminance levels cooler color temperatures (17,000 K) have an arousing effect, and studies even show that the concentration ability goes up. It should be noticed that this research has mainly been carried out with fluorescent tubes which have a very different spectral distribution pattern and not all studies have been conducted within school environments (Viola et al. 2008; Mills et al. 2007). In the same way, lower illuminance levels together with lower CCT seem to have a calming effect (Baron et al. 1992; Knez and Enmarker 1998).

According to Barkman et al. (2012), the research on artificial lighting in schools may be divided into three groups relating to the type of light sources used in the studies. First, early studies using full-spectrum lamps, not used anymore, hypothesize that the complete color spectrum was needed in the school situation in order to reach positive effects. However, the results were never clear-cut, and according to McColl and Veitch (2001), the relation between visual, perceptual, and cognitive effects of the full-spectrum fluorescent tubes was very weak. Other researchers reached the same conclusion (Boyce 1994; Gifford 1994). One reason for this may be that the spectral distribution of those lamps was not comparable to the daylight they were supposed to simulate. Another type of studies was comparing modern fluorescent tubes with different color temperatures together with the impact of daylight (Küller and Lindsten 1992). The results indicated that the cool light (4000 K) resulted in higher ability to concentrate and the warmer (3000 K) light may enhance communication.

Colored filters have also been used to facilitate reading among pupils. Results indicate that using such filters reduces symptoms of visual stress and headache and speeds up the ability to read (Wilkins 2003).

The third type of studies has been assessing how dynamic or variable lighting affects school children in various ways. We will return to this question in a later part of this article. Before we go to this, let us look at the other side of the coin, namely, the discomfort of light in the school environment.

Studies have been conducted where different adverse effects of lighting have been investigated and their potential impact on children's school performance. However, one problem with these kinds of studies is that they vary quite a lot in design and sometimes the number of subjects has been small which make it difficult to generalize from the results. With this in mind, nevertheless, there have been results pointing in directions that certain factors may hinder. One such factor is discomfort glare. In a study by Winterbottom and Wilkins (2009), the luminance from whiteboards was so high that it may induce discomfort. The results fit well in with findings about students' complaints about visibility of data-projection screen (Hall and Higgins 2005; Smith et al. 2005). Another problem is the flicker from the light sources. Having lamps operated with AC supply (50 Hz in Europe, 60 Hz in the USA), there will be a modulation in light output twice the supply frequency. With incandescent bulbs, this modulation is very small due to the fact that the filament will not cool down, but with fluorescent tubes, there may be a modulation in illuminance between 9 % and 90 %. The 100 Hz modulation affects humans both perceptually and cognitive. Studies have shown that young people are more sensitive to these effects than older people (Küller and Laike 1998) and that children diagnosed with autism are extremely sensitive, for example, in one study, the repetitive behavior decreased and communicative behavior increased when using flicker-free lighting (Coleman et al. 1976). Concerning LED, there are indications that especially when dimming, there is a modulation of this kind, depending on the technique used for dimming and because of the ballasts used. However, studies using LED are few. A recent study by Govén et al. (2014) didn't find any adverse effects of this type in a school setting with high school students. On the contrary, the perception of the

lighting conditions was much more positive concerning the LED solution. However, more studies are needed and clear definitions of the concepts of flicker and light modulation must be developed.

In a study where the total physical environment of schools was investigated, it was found that concerning lighting, the schools differed quite a lot. More interesting was also that the behavior concerning the use of lighting was that normally it was switched on in the morning and then on during the whole day, despite changes in the visual conditions (De Giuli et al. 2012). The control of the lighting seems to be a factor that could be handled in different ways. Today's general knowledge about physical environments points that the physical environment should not be too static, if so the environment will not be stimulating, and a boredom will occur. On the other hand, it is important that the environment is not too chaotic or complex, which will lead to overstimulation. However, it seems that concerning lighting, the first mentioned situation seems to be the most common (Küller 1991). As mentioned above, the dynamic or variable light has been one recent topic for study. Several studies have been presented and are based on the assumption leading back to earlier research that preferences for certain light are dependent on situational factors and individual needs. Besides this psychological assumption, also the knowledge about different levels of light for handling human circadian rhythms has brought attention.

The factors that often have been varied are the illuminance levels and the color temperature (CCT). This may be done through certain preprogrammed schemes or preprogrammed scenarios that are chosen by the teacher. A German study compared two classrooms in two separate schools with variable lighting with two control classrooms. The variable lighting scenarios were set to seven different variants, standard, focus on board, board only, concentrate, activate, relax, and extreme relax. The teacher in the classroom was supposed to handle and chose the alternatives. The study went on for 9 months, and the results showed first of all that the students and the teachers were satisfied with the variable light. Furthermore, the result showed that under the variable light variant "Concentrate," the students displayed better attention than on the other solutions. The students also displayed a higher reading speed, and reading comprehension was also improved. On the other hand, the achievement motivation and the atmosphere of the classroom did not change with the variable lighting solution. The light source used in this study was fluorescent tubes (Barkmann et al. 2012).

In a Dutch study comprising of three different studies, Sleggers et al. (2013) compared different vertical illuminance levels (350–1000 lux) and different correlated color temperatures (3000 and 12,000 K). Two of the studies were quasi-experimental field studies using a dynamic lighting system, while the third study was conducted in a simulated school setting in a windowless laboratory setting. The aim of the two field studies was to investigate a lighting system with the possibility to address different needs in the classroom situation such as activation, attention, and calmness. The same aim was the target for the third study but under a controlled situation. The focus of the study laid on the cognitive performance as in the former presented German study. The results affirm that the lighting conditions have an impact on the children's performance. However, the results are not completely

clear-cut, since there are differences between students from grade four, where positive results were found, which not the case was for students from grade six. One explanation suggested by the authors is, as been mentioned earlier in this chapter, that younger children are more affected by environmental stimuli, than their older counterparts. The older children have learned to handle environmental factors in a better way. Furthermore, the controlled laboratory study did not confirm the results from the field studies. Concerning this result, the authors suggest that the reason may be the time of the year the studies were conducted. The two field studies were carried out during the winter season, and the laboratory study was conducted in spring. The seasonal effects should not be neglected. As Küller and Lindsten (1992) showed in their study, the reaction, both physiological and behavioral, varies over the year in countries far from the equator where big differences in the length of the day are at hand. The impact of daylight will be larger during the spring and summer season in comparison to the winter season. To summarize, the dynamic solution based on different settings aimed for different demands seems to have an effect, but to a relatively moderate amount.

Another way to look at the lighting system is to investigate the lighting distribution in the room. Research suggests that the surrounding light may be important, not least for the nonvisual effects of light. In a study where four classes were compared according to daylight and lighting distribution, pupils from grade three in a school in the UK were studied (Govén et al. 2011). Two classrooms had a large degree of surrounding light (75 % on walls and ceiling), while two control classrooms had (45 % on walls and ceiling, standard solution). In all classrooms, the light sources were changed in order to minimize the Hawthorne effect. Furthermore, in one experimental and one control room, the daylight factor was around 1.5 %, and in one experimental and one control classroom, the daylight factor was around 4.5 %. The pupils were followed during one school year at five different occasions with measurement of the children's subjective experience of sleep and emotional status together with measurements of the children's cortisol and melatonin levels at different occasions. The children's school performance was also rated. The results indicated that the two rooms with the highest amount of daylight described similar results, while the rooms on ground floor with lower daylight contribution, an effect during the darkest season was found, on both regarding arousal as measured by the morning cortisol and on the school performance, showing that the pupils in the experimental classroom displayed better results in school than their counterparts in the control room. The cortisol pattern revealed that the children in the experimental room displayed higher arousal during the dark season (November to February) indicating that the mere difference of the distribution of the light may have an impact on the children.

The knowledge from this study was brought to a second study where the idea was to compare LED light sources with fluorescent T5 tubes where the light distribution was as identical as possible. The study was conducted in southern Sweden with high school students. These students were followed for one school year, and the measurements conducted were the same as in the English study, with the exception that melatonin was not measured in this study. The student's activity during the measurement days was also checked.

The details of the study were as follows: in one conventional T5 fluorescent tubes with 4000 K and in the other a LED solution, also with 4000 K. The lighting level on the work plane in both situations was 500 lux (the standard in Sweden). In the study, both subjective experience of lighting conditions and the results of chronobiological marker (cortisol) were measured during three occasions during 1 day, at five consecutive times over the school year. The results concerning the experience were quite clear, that even though the conventional solution was good, the LED solution was experienced as brighter and the visibility displayed a higher rating. However, regarding the chronobiological marker, the results did show small differences between the two kinds of solutions; rather, daylight has an immense impact on the circadian rhythm. In Sweden, as other countries far from the equator, there is a large difference in the length of the day over the year, and this fact was clearly manifested in the results. In March, when the length of the day became longer, the results were clear. However, a small trend toward a more stable circadian rhythm was seen in November within the classrooms equipped with LED. Taken together, one could conclude that using LED is at least as good as using conventional fluorescent tube solutions, but it is important that the luminaire is taking care of all the possibilities for glare (Govén et al. 2014).

Conclusions

The overall results from different studies show that, with the new light source LED, there is a potential to produce better lighting solutions, both concerning the intensity, but also in relation to the spectral distribution. However, our knowledge is still limited and we need to know more about the impact of light from different wavelengths, especially in the long wavelength band. Based on today's knowledge, the intensity and the positive effects of short wavelength light in the morning hours are something that could be introduced relatively easy in the school environment. At the same time, it is important not to forget the impact of daylight, that the school environments need to use daylight in a proper way as much as possible. The artificial light should be seen as a complement to that should be used when the daylight could not fulfill the demands of the pupils. We must also remember the basic quality needs such as glare problem and enough light. Those two aspects are, as have been clearly shown, also very important when describing high-quality educational lighting.

What is the future for the lighting of schools? This question is difficult to answer since the development goes fast, but in environments where additional light is needed, the future may lay in a light that could mimic the daylight, a light that is free from glare, has the same properties as daylight, and are changing in line with the daylight. When working during evenings and nights, one must take this into account and have information on when the person should go to sleep and relate this to the total amount of radiation needed for a good solution that is both healthy and stimulating. Another thing may be to have lighting solutions that could be helpful for different activities in the school environment. The teacher may change the ambient lighting depending on the work conducted, if it is to work in group or to

work individually. The most important thing is that we create lighting environments that go together with the human needs. We are today in the beginning of the understanding of this, and there is still much to learn, but with the development of technology as well as the understanding of the impact of light on psychological responses, there is much to come.

References

- Barkmann C, Wessolowski N, Schulte-Markwort M (2012) Applicability and efficacy of variable light in schools. *Physiol Behav* 105:621–627
- Baron RA, Rea MS, Daniels SG (1992) Effects of indoor lighting (illuminance and spectral distribution) on the performance of cognitive tasks and interpersonal behaviours: the potential mediating role of positive affect. *Motiv Emot* 16:1–33
- Boyce P (1994) Is full-spectrum lighting special? In: Veitch JA (ed) Full-spectrum lighting effects on performance, mood and health (IRC Internal report no 659. National Research Council of Canada, Institute for Research in Construction, Ottawa, pp 30–36
- Boyce P (2004) Human Factors in Lighting. Boca Raton: CRC Press
- Chen R, Peng L, Yan Y, Lin Y (2007) Investigation on luminous environment of classrooms in the countryside. *China Illum Eng J* 2:3–9
- Coleman RS, Frankel F, Ritvo E, Freeman BJ (1976) The effects of fluorescent and incandescent illumination upon repetitive behaviors in autistic children. *J Autism Child Schizophr* 6:157–162
- De Giuli V, Da Pos O, De Carli M (2012) Indoor environmental quality and pupil perception in Italian primary schools. *Build Environ* 56:335–345
- Edwards L, Torcellini P (2002) A literature review of the effects of natural light on building occupants. Technical report NREL/TP-550-30769. National Renewable Energy Laboratory, Golden
- Garnert J (1993) *Anden i lampan (The Genie of the lamp)*. Carlssons, Stockholm
- Gifford R (1994) Scientific evidence for claims about full-spectrum lamps: past and future. In: Veitch JA (ed) Full-spectrum lighting effects on performance, mood and health (IRC Internal report no 659. National Research Council of Canada, Institute for Research in Construction, Ottawa, pp 37–46
- Govén T, Bångens L, Persson B (2002) Preferred luminance distribution in working areas. In: *Proceedings of right light 5, 5th European conference on energy-efficient lighting (nice)*. Borg & Co., Stockholm, pp 87–92
- Govén T, Laike T, Raynham P, Sansal E (2011) Influence of ambient light on the performance, mood edocrine systems and other factors of school children. *Proceedings 27th session of the CIE Sun City/South Africa 10–16 July 2011, volume 1, part 1, 112–121*. ISBN 978 3 901906 99 2
- Govén T, Gentile N, Laike T, Sjöberg K (2014) Energy efficient and study promoting lighting at high school: preliminary results, accepted for presentation at CIE 2014, Kuala Lumpur.
- Hall I, Higgins S (2005) Primary school students' perceptions of interactive whiteboards. *J Comput Assist Learn* 21:102–117
- Higgins S, Hall E, Wall K, Woolner P, McCaughey C (2005) *The impact of School Environments: A literature review*. Newcastle: The centre for learning and Teaching School of Education, Communication and Language Science, University of Newcastle
- Higgins S, Hall E, Wall K, Woolner P, McCaughey C (2005) *The impact of school environments: a literature review*. The Centre for Learning and Teaching School of Education, Communication and Language Science, University of Newcastle, Newcastle
- Knez I, Enmarker I (1998) Effects of office lighting on mood and cognitive performance and a gender effect in work-related judgement. *Environ Behav* 30:553–567

- Küller R (1991) Environmental assessment from a neuropsychological perspective. In: Gärling T, Evans GW (eds) *Environment, cognition and action: an integrated approach*. Oxford University Press, New York, pp 111–147
- Küller R, Laike T (1998) The impact of flicker from fluorescent lighting on well-being, performance and physiological arousal. *Ergonomics* 41:433–447
- Küller R, Lindsten C (1992) Health and behaviour of children in classrooms with and without windows. *J Environ Psychol* 12:305–317
- Laike T (1997) The impact of daycare environments on children's mood and behavior. *Scand J Psychol* 38(3):209–219
- Lee YY, Lo CT, Sheu SJ, Lin JL (2013) What factors are associated with myopia in young adults? A survey study in Taiwan military conscripts. *Invest Ophthalmol Vis Sci* 54:1026–1033
- McCull SL, Veitch JA (2001) Full-spectrum fluorescent lighting: a review of its effects on physiology and health. *Psychol Med* 31:949–964
- Mills PR, Tomkins SC, Schlangen LJM (2007) The effect of high correlated colour temperature office lighting on employee wellbeing and work performance. *J Circadian Rhythms* 5:2–10
- Rea M (2012) *Value metrics for better lighting*. SPIE Press, Bellingham US
- Robson ER (1877) *School architecture: being practical remarks on the planning, designing, building and furnishing of school-house*. John Murray, London
- Sleggers PJC, Moolenaar NM, Galetzka M, Pruyn A, Sarroukh BE, van der Zande B (2013) Lighting affects students' concentration positively: findings from three Dutch studies. *Light Res Technol* 45:159–175
- Smith HJ, Higgins S, Wall K, Miller J (2005) Interactive whiteboards: boon or bandwagon? A critical review of the literature. *J Comput Assist Learn* 21:91–101
- Tikkanen KT (1979) Spectral eye fatigue in a school environment. *Light Res Technol* 11:185–188
- Ulrich RS (1984) View through a window may influence recovery from surgery. *Science* 224:420–421
- Van Bommel WJM, Van den Beld GJ (2004) Lighting for work: a review of visual and biological effects. *Light Res Technol* 36:255–266
- Viola AU, James M, Schlangen LJM, Dijk DJ (2008) Blue-enriched white light in the work-place improves self-reported alertness, performance and sleep quality. *Scand J Work Environ Health* 34:297–306
- Wilkins A (2003) *Reading through colour*. Wiley, Chichester
- Winterbottom M, Wilkins A (2009) Lighting and discomfort in the classroom. *J Environ Psychol* 29:63–75
- Wu W, Ng E (2003) A review of the development of daylighting in schools. *Light Res Technol* 35:111–125

Ethnic and Social Aspects of Lighting

Shin Ukegawa

Contents

Introduction	908
Evaluation Index	908
Evaluation of the Preference Index of Skin Color (Index PS) to Assess the Preferred Appearance of Japanese Female’s Facial Skin Color	909
Deviation Duv from Blackbody Locus	909
Feeling of Contrast Index FCI	910
Clarification of Spectral Radiation Distribution	911
LED Spectral Radiation Distribution That Shows the Preferred Facial Skin Color	912
LED Light Source’s Spectral Radiation Distribution That Renders Visual Clarity of Fresh Food and Plants	913
Conclusion	916
References	917

Abstract

In the development of the LED light sources that will be introduced here, the adopted color rendering evaluation indexes include Duv (Methods for determining distribution temperature and color temperature or correlated color temperature of light sources, JIS Z (8725–1999)), PS (Light 82–11:895–901, 1998), and FCI (Color Res Appl 32–5:361–371) in addition to the General Color Rendering Index (Ra) (Method of specifying color rendering properties of light sources, JIS Z (8726–1990)). The individual target values are defined to clarify the spectral radiation distribution. This article focuses on the above characteristics: the technical details used in the development of a LED light source called “Favorable Color (Bikou-shoku),” which preferably renders facial skin color, and “Vivid

S. Ukegawa (✉)

Lighting Bussiness Group, Panasonic Corporation, Kadoma, Osaka, Japan

e-mail: ukegawa.shin@jp.panasonic.com

Color (Saikou-shoku),” which vividly renders the true colors of fresh food and plants.

Introduction

LED light sources, which have already spread to general households, are well known as long-life energy-efficient light sources. In comparison to conventional light sources, they are characterized by a spectral radiation distribution that is easier to control. This article focuses on the above characteristics: the technical details used in the development of a LED light source called “Favorable Color (Bikou-shoku),” which preferably renders facial skin color, and “Vivid Color (Saikou-shoku),” which vividly renders the true colors of fresh food and plants.

Evaluation Index

Figure 1 shows the LED’s general spectral radiation distribution. First, we must determine the type of distribution over a wide range (visible light), from 380 to 780 nm, because the spectral radiation distribution determines the light source’s color rendering performance. However, even in 5-nm increments, this task must find an 81st-dimensional solution and cannot be obtained easily by trial and error. To efficiently clarify the intended spectral radiation distribution, we must set the evaluation indexes to access the intended color rendering.

In the development of the LED light sources that will be introduced here, the adopted color rendering evaluation indexes include Duv (JIS Z (8725–1999)), PS (Yano and Hashimoto 1998), and FCI (Hashimoto and Yano 2007) in addition to the

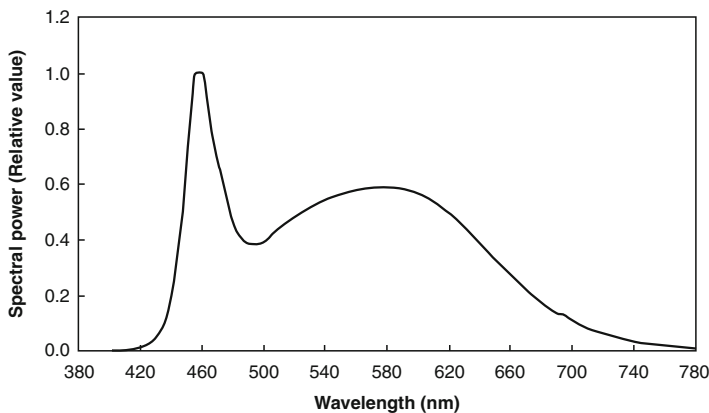


Fig. 1 Example of LED spectral radiation distribution

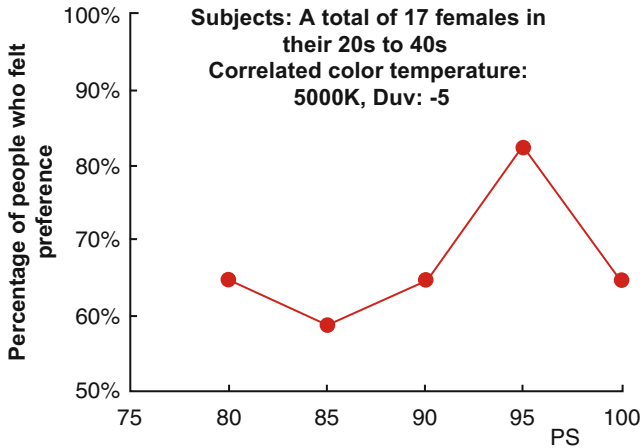


Fig. 2 Relationship between PS values and evaluations of facial skin color preferences

General Color Rendering Index (Ra) (JIS Z (8726–1990)). The individual target values are defined to clarify the spectral radiation distribution.

Evaluation of the Preference Index of Skin Color (Index PS) to Assess the Preferred Appearance of Japanese Female’s Facial Skin Color

Yano et al. identified the chromaticity point of Japanese female’s preferred facial skin color and proposed a (Yano and Hashimoto 1998) PS index that assesses the preferred appearance of Japanese female’s facial skin color based on the light source’s spectral radiation distribution. The closer the PS values are to 100, the higher the preference level of the facial skin color under illumination.

Index PS is quantified from the results of evaluation experiments where we used an illumination source consisting of three combined fluorescent lamps: red, green, and blue. Yamaguchi et al. carried out psychological estimation experiments with participants under LED light sources to verify the validity of their PS values.

Figure 2 shows the relationship between the PS values obtained from the above experiment and the level of the preferred appearances of the facial skin colors. Seventeen women in their 20s, 30s, and 40s participated in the experiment. The vertical axis in Fig. 2 shows the percentage who felt the preferred appearance of the facial skin color.

Deviation Duv from Blackbody Locus

As shown in Fig. 3, Duv indicates the amount of deviation expressed as a numerical value from the blackbody locus, which is a locus of the chromaticity coordinates of such natural light as sunlight. (JIS Z (8725–1999)) When the Duv value is positive (+), light is

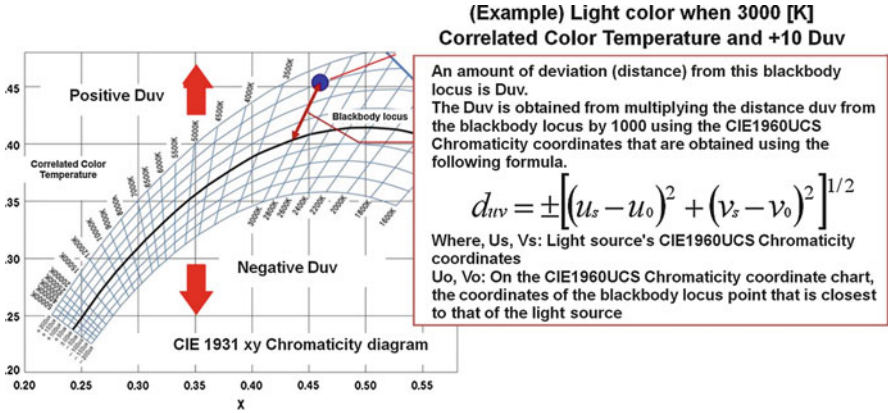
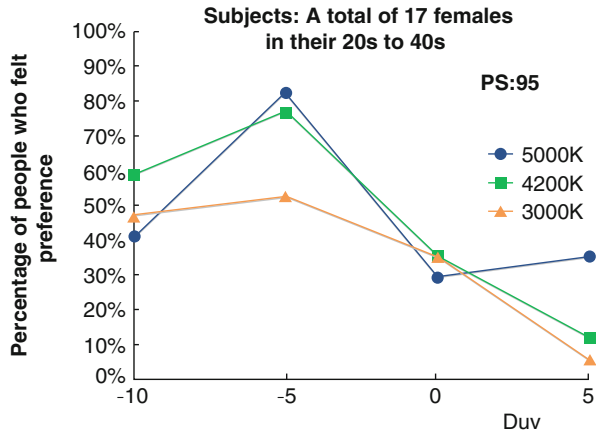


Fig. 3 Blackbody locus and Duv

Fig. 4 Relationship between Duv values and evaluations of facial skin color preferences



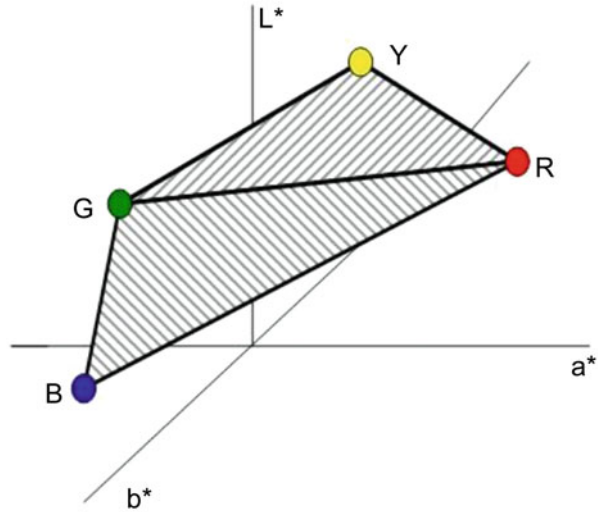
perceived as greenish. When it is negative (–), light is perceived as pinkish. Although the light sources have the same correlated color temperature, the perceived colors may be different. This is often caused by different Duvs.

Duv probably also affects facial skin color perception. Yamaguchi et al. carried out the above evaluation experiments (Yamaguchi and Saito 2012) in which both the PS values and the Duv values were changed. Figure 4 shows our experiment results. The relationship is shown between the Duv values and the evaluated preferred appearances of Japanese females’ facial skin colors.

Feeling of Contrast Index FCI

Under light sources with different color rendering properties, Hashimoto et al. experimentally related the visual clarity of colors from chromatic object groups

Fig. 5 Gamut area of four-color combination samples



(color). A four-color combination was formed that most effectively renders visual clarity (Hashimoto and Nayatani 1990). The four-color combinations are shown below in the Munsell color system: 5R4/12 (red), 5.5G5/8 (green), 4.5 PB3.2/6 (blue), 5Y2/10 (yellow). Using the gamut area (Fig. 5) enclosed by the chromaticity coordinates ($L^*a^*b^*$) of the four-color combination samples in the CIELAB color space (JIS Z (8729–2004)) under the light sources, Hashimoto et al. clarified that formula (1) obtains a feeling of contrast index FCI that allows visual clarity estimation.

$$FCI = [G_{LAB}(T)/G_{LAB}(D65)]^{1.5} \times 100 \tag{1}$$

where $G_{LAB}(T)$ is the gamut area of the four-color combination samples in the CIELAB color space under the evaluated test light sources and $G_{LAB}(D65)$ is the gamut area under base light source D65.

Under light sources with different FCI values, Tsukitani et al. experimentally evaluated the preferred visual clarity of green, red, and yellow plants placed inside a box (Tsukitani and Saito 2013). Figure 6 shows the relationship between the FCI values obtained from the above experiment and the level of the preferred visual clarity.

Clarification of Spectral Radiation Distribution

In developing LED light sources to achieve desired color rendering, we set the target values of the above four indexes to obtain the intended spectral radiation distribution.

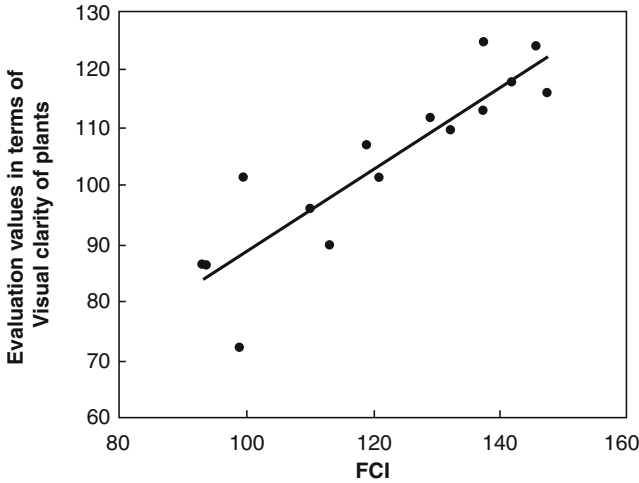


Fig. 6 Relationship between FCI values and evaluation of preferred visual clarity of plants

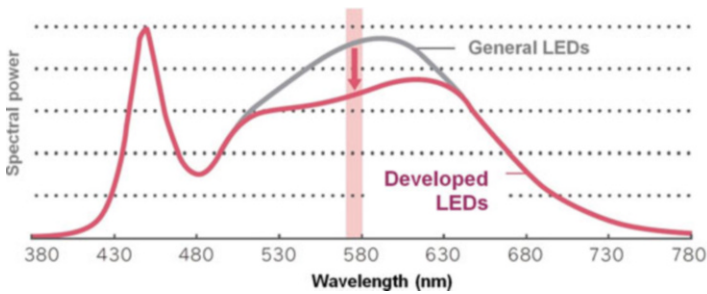


Fig. 7 Shows the spectral radiation distribution concept that renders the preferred appearances of facial skin colors

LED Spectral Radiation Distribution That Shows the Preferred Facial Skin Color

We set the following development target values of the LED light source’s spectral radiation distribution to obtain the preferred appearance of facial skin colors: PS 95 from the result of Fig. 2, Duv -5 from the result of Fig. 3, and a general color rendering index Ra of 90 or higher. Figure 7 shows the concept of the LED light source’s spectral radiation distribution that was developed to obtain the target values.

To clarify the spectral radiation distribution, we focused on wavelength light around 570–580 nm, which emphasizes meat’s delicious-looking color but is low in neodymium lamps. As a result, the light source developed this time has spectral radiation distribution that achieves PS 95, Duv -5 , and Ra 95.

Figure 8 compares the chromaticity coordinates (U^* , V^*) of the eight-color samples R_1 to R_8 in the CIE1964 uniform color space. The eight-color samples are used to obtain

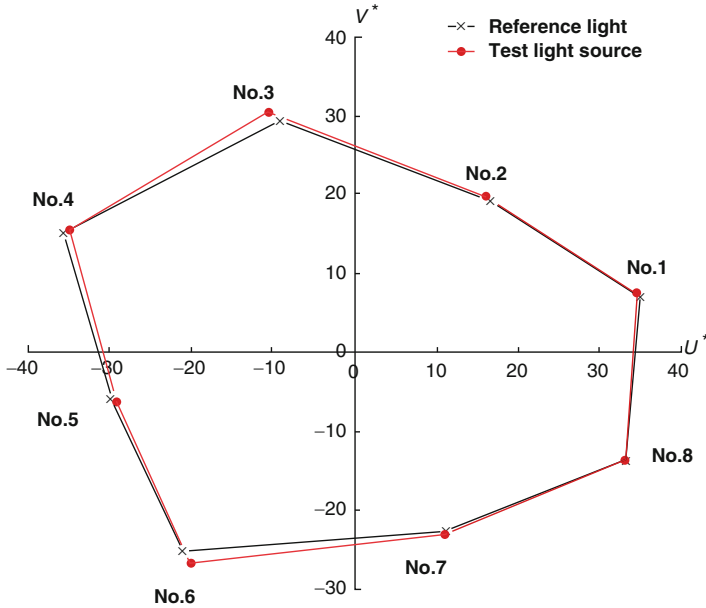


Fig. 8 Comparison of chromaticity coordinates of R₁ to R₈ (The reference light and the light that renders the preferred facial skin color are both 5000 K)

general color rendering index Ra. • indicates the chromaticity coordinates that are illuminated under the newly developed spectral radiation distribution. × indicates the chromaticity coordinates relative to the reference light with the same correlated color temperature. Since the chromaticity coordinates are closer to those of the eight-color samples under the reference light with the same correlated color temperature, general color rendering index Ra will be closer to 100. Since the chromaticity coordinates, which are illuminated under the newly developed spectral radiation distribution, has Ra 95, the value is close to that of the chromaticity coordinates under the reference light.

Figure 9 compares the chromaticity coordinates of samples R₉ (red), R₁₀ (yellow), R₁₁(green), and R₁₂ (blue) that are used to obtain the special color rendering index. (JIS Z (8726–1990)) As in the results of the eight-color samples shown in Fig. 8, their chromaticity coordinate values resemble each other.

Based on the above results, our newly developed spectral radiation distribution allows both a true color rendering under the base light and a preferred appearance rendering of the facial skin color.

LED Light Source’s Spectral Radiation Distribution That Renders Visual Clarity of Fresh Food and Plants

To avoid an uncomfortable light source under illumination, we set the development target values of the LED light source’s spectral radiation distribution to obtain the

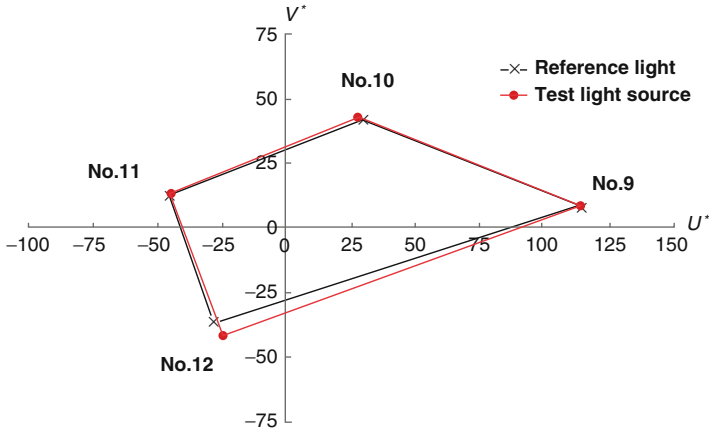


Fig. 9 Comparison of chromaticity coordinates of R₉ to R₁₂ (The reference light and the light that renders the preferred facial skin colors are both 5000 K)

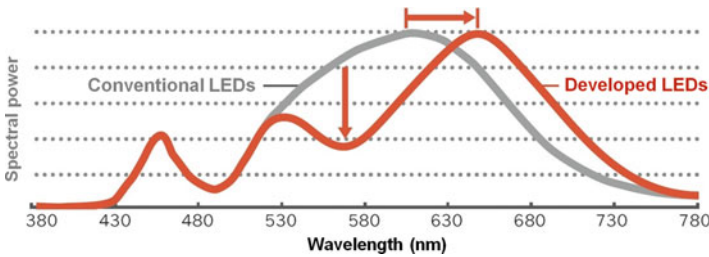


Fig. 10 Shows the spectral radiation distribution concept that renders the preferred object’s visual clarity

visual clarity of fresh food and plants as follows, based on the results of Fig. 6: FCI 135 or higher, Duv 0, and Ra 80 or higher. Figure 10 shows the concept of the LED light source’s spectral radiation distribution that was developed to obtain the target values. As in the spectral radiation distribution (Fig. 7) that renders preferred facial skin colors, the wavelength light around 570–580 nm was reduced, but the peak wavelength of the red component shifted toward the long wavelength side to improve red’s visual clarity. FCI 138, Duv 0, and Ra 84 were obtained.

Figures 11 and 12, as Figs. 8 and 9, compare the chromaticity coordinates that are obtained from the above spectral radiation distribution.

As the chromaticity coordinates (U^* , V^*) in the CIE1964 uniform color space move farther away from the original point, we obtain better visual clarity of the

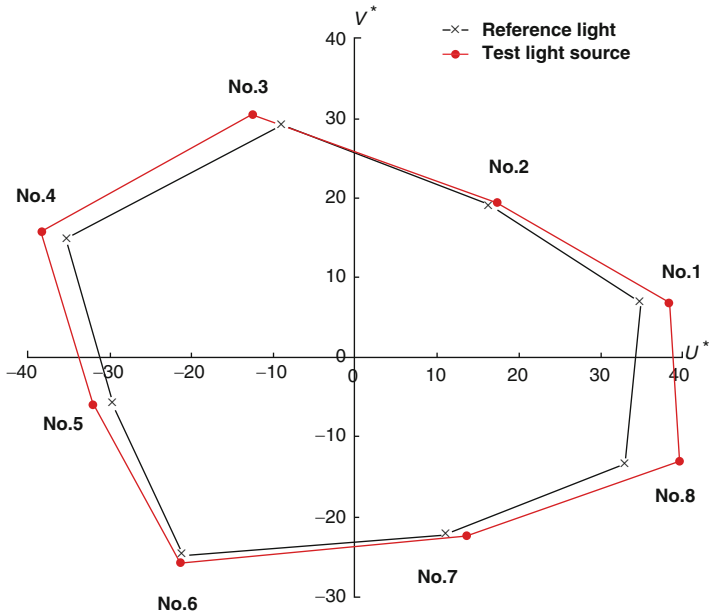


Fig. 11 Comparison of chromaticity coordinates of R_1 to R_8 (The reference light and the light that renders the object’s visual clarity are both 5000 K)

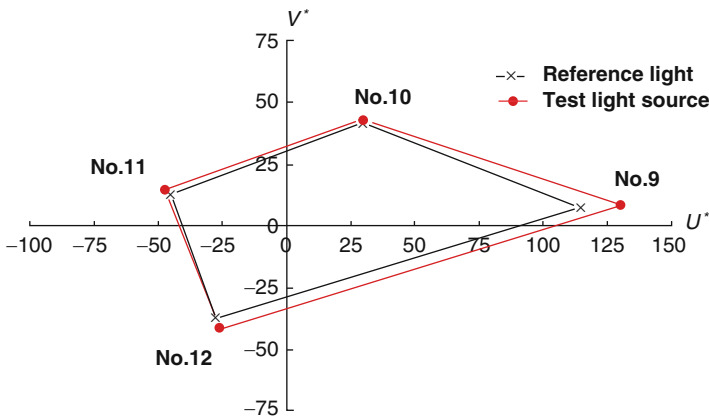


Fig. 12 Comparison of chromaticity coordinates of R_9 to R_{12} (The reference light and the light that renders the object’s visual clarity are both 5000 K)

objects. Figures 11 and 12 show that our newly developed spectral radiation distribution renders an object’s visual clarity better than the base light with the same correlated color temperature.



Fig. 13 Beauty salon with “Favorable Color” LED downlight: Musee Platinum Ikebukuro Higashiguchi

Conclusion

We successfully developed spectral radiation distribution (Fig. 7). We named our developed LED light source “Favorable Color (Bikou-shoku).” Fig. 13 shows an illuminated space of a beauty salon that is using “Favorable Color (Bikou-shoku)” as an LED light source. This LED light source not only renders preferred facial skin colors but also has general color rendering index Ra95 that enables true rendering of natural light colors without any yellow tint. Thus, this feature is utilized in spaces that want to emphasize cleanliness, such as department stores, boutiques, and bathrooms.

We also successfully developed spectral radiation distribution (Fig. 10). We named our developed LED light source “Vivid Color (Saikou-shoku).” Fig. 14 shows an illuminated space of a supermarket that is using the “Vivid Color (Saikou-shoku)” LED light source. Unlike conventional light sources for meat that increase reddish tints to improve its preferred appearance, the “Vivid Color” LED light source (Saikou-shoku), as indicated by its Duv 0, identically renders color as general white light sources. It can also be used for general lighting. Even though the use of a light source is mixed with general white light sources, the same level of comfort is maintained in the space. In addition, there is no need to use separate light sources for meat or vegetables. The “Vivid Color (Saikou-shoku)” LED light source itself alone is adaptable to any fresh food without an additional light source. As Fig. 6 shows, the light source is also suitable for illuminating plants.

We discussed the Spectrum Control Technology that allows spectral radiation distribution embodied in the “Favorable Color (Bikou-shoku)” and “Vivid Color



Fig. 14 Supermarket where “Vivid Color” LED spotlighting is being used: Ujie Super Misato

(Saikou-shoku)” LED light sources, which are distinctive LED characteristics. Further advancement of this technology is expected to achieve balanced goals of comfort and energy efficiency.

References

- Hashimoto K, Nayatani Y (1990) Evaluation and estimation of sense of brightness based on the feeling of four-color combination, The Illuminating Engineering Institute of Japan. *Journal of the Illuminating Engineering Institute of Japan* 74–10:96–101
- Hashimoto K, Yano T (2007) New method for specifying color-rendering properties of light sources based on feelings of contrast. *Color Res Appl* 32–5:361–371
- JIS Z (8726–1990) Method of specifying color rendering properties of light sources
- JIS Z (8725–1999) Methods for determining distribution temperature and color temperature or correlated color temperature of light sources
- JIS Z (8729–2004) Color specification $L^*a^*b^*$ and $L^*u^*v^*$ color spaces
- Tsukitani A, Saito T (2013) Light source to render the preferred color of plants – study of the color rendering optimization based on a visual clarity index, FCI. In: *The illuminating engineering society annual conference proceedings*. Nagoya, Japan, p. 39
- Yamaguchi S, Saito T (2012) Spectral characteristics of LED light sources for preferred appearance of facial skin color. *Panasonic Tech J* 58–2:62–66
- Yano T, Hashimoto K (1998) Preference index for Japanese complexion color under illumination, The Illuminating Engineering Institute of Japan. *Light J* 82–11:895–901

Part VII

Energy Efficiency

Energy Consumption and Environmental and Economic Impact of Lighting: The Current Situation

Georges Zissis

Contents

Introduction	922
Energetic Impact of Artificial Lighting	922
Environmental Impact of Artificial Lighting (Carbon Footprint)	926
Economical Impact of Lighting	927
References	932

Abstract

Artificial light sources play an indispensable role to daily life of any human being. Electrical light sources are responsible for an energy consumption of around 1/6 to 1/5 of the worldwide electricity production. Although classic lighting technologies are now mature, the luminous efficiency of the light sources together with their quality of light have not quite reached their limits: there is still room for innovation. Today, there are many opportunities for enhancing not only the efficiency and reliability of lighting systems but also improving the quality of light as seen by the end user. Furthermore, currently the next revolution in lighting is taking place: Solid State Lighting (SSL). In the long-term SSL, inorganic and organic light emitting diodes are now replacing massively legacy technologies. In fact, LEDs, with a continuous growth of their luminous efficiencies, establish themselves as breakthrough solutions.

G. Zissis (✉)

LAPLACE, UMR 5213 CNRS, INPT, UPS, Université de Toulouse, Toulouse, France

e-mail: georges.zissis@laplace.univ-tlse.fr

© Springer International Publishing Switzerland 2017

R. Karlicek et al. (eds.), *Handbook of Advanced Lighting Technology*,

DOI 10.1007/978-3-319-00176-0_40

921

Introduction

Man always desired to go on with his normal life after nightfall. For this reason, artificial light sources play an indispensable role to daily life of any Human being. Quality of life, health, and urban security related to traffic and crime prevention depend on light and on its quality. In fact, light is vital for life. Nowadays, the average artificial light quantity per capita is 600 times higher than that used by an average Englishman at the end of the nineteenth century (Fouquet and Pearson 2006). Today, the world of light sources has undergone a major revolution: incandescent lamps, after 150 years of service, are banned from the market while the Solid-State Lighting (SSL) systems come into play.

In just the last few years, LED performance has accelerated quickly and a wave of new commercial, industrial, and institutional LED fixtures has been introduced. LED technology is fulfilling its promise of offering the market the most efficient means of converting electrons into photons. LEDs have thus surpassed many conventional lighting technologies in terms of energy efficiency, lifetime, versatility, and color quality and due to their increasing cost competitiveness are beginning to successfully compete in a variety of lighting applications. Therefore, LED lighting is no longer “around the corner,” but it is here and has a solid market foothold. Performance is improving; production and purchase costs are coming down. In addition, Organic Emitting Diodes (OLEDs) are more and more mature and it is expected that they will penetrate massively lighting within the next 5–10 years. However, the OLED penetration is strongly dependent on the production costs that are still high and also on the lifespan of these components that cannot be considered as competitive until today.

Today, high-performance products offer added value beyond efficacy. However, as any new or emerging technology, SSL products should be proven to be at least as safe as the products they intend to replace. Also, in new lighting applications where older technologies could not be employed, the safety of SSL products should be assessed considering new or unusual conditions of usage.

Energetic Impact of Artificial Lighting

Electrically powered lighting is used daily throughout the world, in fact, approximately 30–40 billion electric lamps operate everyday. This corresponds roughly to 134.7 peta-lumen-hour (Plm.h) of electrical light. This, in 2005, corresponded to an average light energy of 27.6 Mlm.h per capita and per annum. Of course there are huge variations from country to country, thus an average North American consumer swallowed in a year 101 Mlm.h while an average Indian uses 3 Mlm.h. Figure 1 shows the average annual light use per capita for various world regions.

At the beginning of the twenty-first century and almost 150 years after the first incandescent bulb commercialization, electrical light sources are responsible for an energy consumption of around 3418 TWh per annum (Brown 2009). This quantity represented roughly 19 % of world’s total electricity consumption estimated at

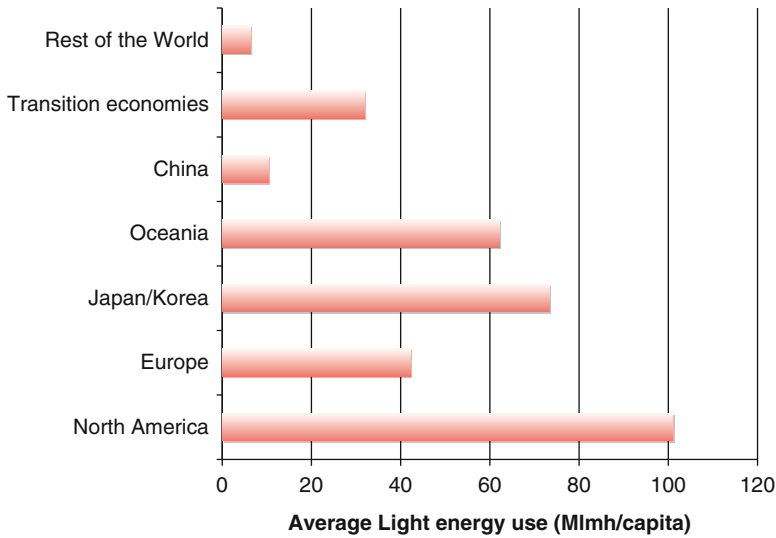


Fig. 1 Average Light use per capita and per annum in the early twenty-first century for different world regions (Data from Waide and Tanishima (2006))

17,982 TWh that corresponded to the full electricity production of 1265 large power plants. The global energy consumption has been recalculated in 2005 based on International Energy Agency (Waide and Tanishima 2006) estimation that omitted transmission and distribution losses; the IAE value has multiplied by a factor of 1.288 in order to correct that omission. Assuming that electricity represents 16 % of the global energy used worldwide (typically 20 % in western countries), lighting represents something like 2.4 % of the annual energetic resources of humanity.

Figure 2 shows how this enormous energy quantity splits among the major world regions. OECD countries are responsible for 65 % of the full consumption. North America (33 %), followed by Europe (15 %), are the major consumers today. China (9 %) and developing economies (10 %) will become the next large consumers.

Looking now closer to the energy consumption per economic sector, as shown in Fig. 3a, major contributors are tertiary buildings (43 %) and residential sector (31 %). The annual lighting electricity consumption per square meter of the tertiary building varies between 20 and 50 kWh/m². It should be noticed here that at the opposite of tertiary building lighting, the residential sector shows a very poor average luminous efficacy of just 21.5 lm/W as compared to approximately 50 lm/W for commercial buildings and 79 lm/W for industrial buildings.

There is a trend in the international community to reduce the electricity consumption of lighting with new technology to below 10 kWh/m² per year. The possible ways to reduce lighting energy consumption include: minimum possible power density, use of light sources with high luminous efficacy, use of lighting control systems, and utilization of daylight.

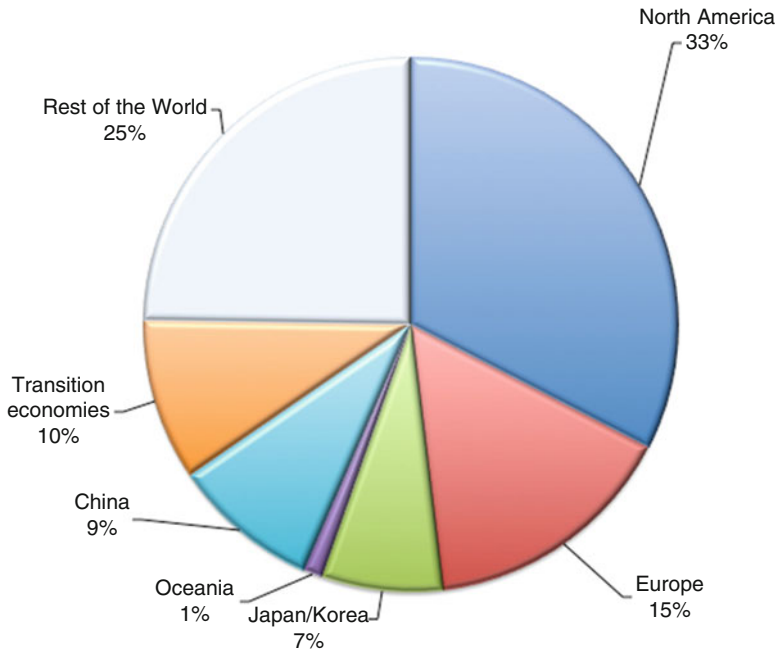


Fig. 2 Electricity consumption for lighting in the early twenty-first century for different world regions century (Data from [WAI-00])

The latest International Energy Agency estimates show the total savings potential in residential and services lighting at more than 2.4 EJ per year by 2030.

Furthermore, the electricity demand for lighting varies from country to country. There are some examples:

France consumes every year, approximately, 45 TWh of electricity for lighting. This is around 11 % of the annual electrical power needs of the country. The lighting of tertiary buildings counts for 45 % of the electricity used for lighting in the country. The street lighting consumes 10 % of the total, whereas the domestic lighting absorbs 30 % of the electrical energy. However, it has to be noticed that in a two decades period during the end of twentieth century, residential lighting has seen its consumption multiplied by three. (Domestic lighting energy consumption in France: 5 TWh in 1979, 14 TWh in 1999.) In fact, lighting in the residential sector accounts for 9 % of household electricity bill. Every French household has in average 28.3 lamps and consumes about 350 kWh per year for lighting.

In USA, the electric energy consumption for lighting in 2010 reached 690 TWh that represent 20 % of the electric energy generated in the country (in absolute value, this energy is equivalent to the combined annual electricity generation of France and Italy). Here again, the tertiary lighting prevails in this consumption as shown by Fig. 3. US Energy Information Administration (2013) estimates that in 2011 about 461 TWh of electricity were used for lighting by the residential and commercial sectors. This was equal to about 17 % of the total electricity consumed by both of

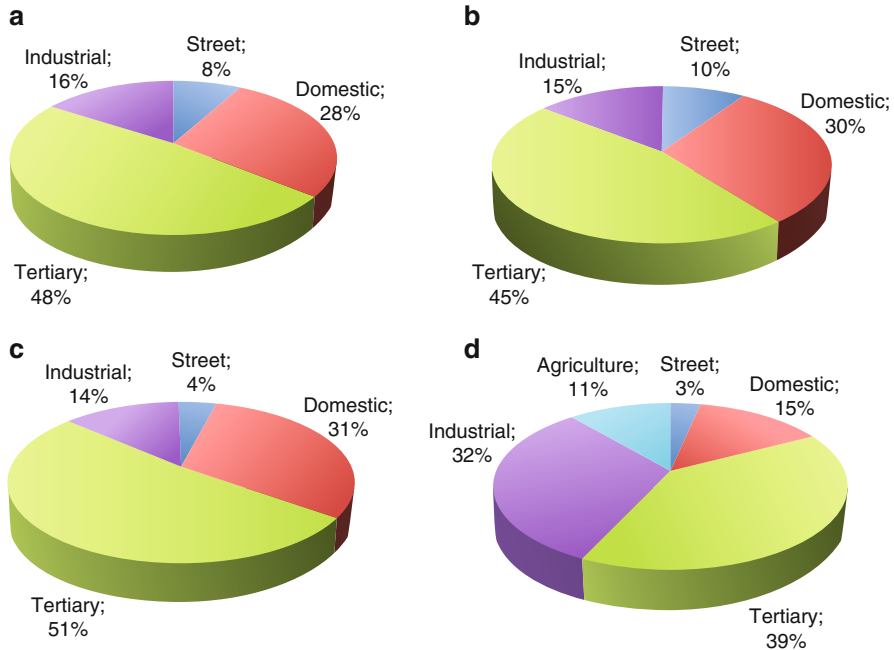


Fig. 3 Energy consumption comparison dedicated to lighting in early twenty-first century (a) worldwide 3418 TWh (2006 corrected values); (b) France 45 TWh (2006-values); (c) in the USA: 694 TWh (2010-values); (d) Russian Federation 137.5 TWh (2006-values)

these sectors and about 12 % of total US electricity consumption. Residential lighting consumption was about 186 TWh or 13 % of all residential electricity consumption. The commercial sector, which includes commercial and institutional buildings and outdoor street/highway lighting, consumed about 275 TWh for lighting or 21 % of commercial sector electricity at the same year.

In Russian Federation today, around 14 % of the country's overall electrical energy consumption is attributed to lighting; this corresponds to 137.5 TWh per annum, but tertiary and industrial sectors represent 70 % of the amount when private residential sector consumption is limited to 14 %, agriculture is still also visible with 11 % of the consumption.

In Japan, the annual electricity consumption for lighting, in 2011, was 150.6 TWh, accounting for 16 % of its total electricity consumption. This figure includes 38.2 TWh for the residential sector (13 % of residential consumption), 89.1 TWh for the commercial sector (33 % of commercial consumption), and 23.3 TWh for the industrial sector (6 % of industrial consumption).

As far as the "developing" countries are concerned, the situation is different. Lighting represents today the major part of their electric consumption: 30 % for Tunisia, about 40 % for Madagascar, and up to 86 % for Tanzania. The domestic lighting is here prevailing. For example, currently in India, lighting accounts for approximately 30 % of total residential electricity use.

This situation is easily understandable since light is an intimate need for Man, and, of course, as soon as electrification progresses, population takes profit from it by first installing lamps. On the other hand, in these countries, the cost of lamps is a major factor in the choice of the light sources type. Thus the incandescent lamps dominate the market because the “energy efficient lamps” are often inaccessible. Once again the energy efficiency is neglected to the benefit of other considerations. . .

Despite the dominance of electric lighting (99 % of the total energy used for light production), a significant amount of energy is also used in vehicle lighting and off-grid fuel-based lighting. Thus vehicle lighting accounted for 0.9 % of the total energy. It is responsible for 55 billion liters of gasoline use per annum; this is equivalent of 1.1 million barrels of oil per day. Fuel-based lighting, used roughly by 1.5 billion people not connected to electrical grids, accounted for only 0.1 % of the total energy used for light production. Even if people without access to electricity use only 50 klm.h per person per annum they are using for lighting 77 billion liters of kerosene per year, which is equivalent of 1.3 million barrels of oil per day.

In past century there had been remarkable increase in the amount of light used all over the world. The annual growth of artificial lighting demand in IEA countries was 1.8 % in last decade, which was lower than during the previous decades. This might be an indication of the start of demand saturation. However, the growth of lighting demand in the developing countries is increasing due to the rising average illuminance levels in those countries and also due to new construction of buildings. The consumption of light in developing countries is expected to increase more in the future due to increasing electrification rate in the regions with no access to electricity at the moment.

Environmental Impact of Artificial Lighting (Carbon Footprint)

Environmental impact of artificial lighting depends on the energy consumption, the material used to produce lighting equipment, the energy used for transportation, and the disposal of used equipment. Only a full Life Cycle Assessment following ISO 14040 methodology can give a realistic estimate of the full impact of lighting. A special chapter of this Handbook is dedicated to this analysis. In this paragraph we are interested only on the carbon footprint related to the energy use. In fact, as in the case of many appliances using electrical energy, the carbon footprint of lightening represents almost 85–90 % of its environmental impact (see in the chapter of this book “► [Life Cycle Assessment of Lighting Technologies](#)” written by L. Tähkämö for more details).

The electric energy generation to satisfy the lighting needs of humanity inevitably leads to environment pollution. Assuming that energy consumption during the lamp life-span represents approximately 90 % of its environmental impact whereas production and disposal/recycling phases correspond to 4 % and raw material use to 6 %, we can calculate the carbon footprint of lighting.

The total lighting-related greenhouse gas (GHG) emissions were estimated to be 1900 million tons (Mt) in 2005, which was about 7 % of the total global CO₂ emissions from the consumption and flaring of fossil fuels. This global amount is equivalent to 70 % of the emissions from light passenger vehicles. (Vehicle lighting is responsible for 161 Mt of GHG emissions.) Of course, increase of renewable energies proportion to the world's electricity production will proportionally reduce this percentage. Fuel-based lighting used in developing countries results in the release of 244 Mt of GHG to the atmosphere per annum.

However, the above calculations are world averages that cannot be extrapolated at national or regional levels. The lighting-related greenhouse gas (GHG) emissions depend strongly on the electricity generation method and the energy mix of each country. The energy mix can be described in a first approximation using the "total primary energy factor." This quantity is defined as the non-renewable and renewable primary energy divided by the delivered energy. In Europe, the total primary energy factor for electricity is 2.5. Furthermore, the CO₂ intensity in power generation differs in each European country; if an average emission factor for electricity of 527 g/kWh is used, the annual GHG emissions of Europe are in the order of 200 Mt.

Economical Impact of Lighting

In March 2011, the US Department of Energy (DoE) estimated that the global world market for lighting to be approximately \$110 billion (US Department of Energy 2011) (\$96 billion following another analyst (Archenhold 2010)). Figure 4 gives the global lighting market growth since 1997 and the trend till 2031.

The observed growth is mainly driven by (a) global population growth (b) developing nations demand for lighting, and (c) electrical grid development. Growth demand seen most in:

- China (19.6 %)
- Eastern Europe (8.1 %)
- Other Asia (6.3 %)
- Africa/Middle East (6.8 %)
- Latin America (9.5 %)
- Canada & Mexico (4.4 %)

The lighting industry sector usually includes lamps, luminaires, and lighting controls. As shown in Fig. 5, traditionally, the lighting market is divided into two major segments: Light sources and Ballast/fixtures.

If we admit that the dominating technologies at that moment (2009) were incandescent and fluorescent lamps, it is possible to split the \$110 billion as follows: 62.2 % for incandescent technology and 34.5 % for fluorescent segment. At the same year, revenue of Light Emitting Diodes for lighting was marginal.

In more details, the world lamp market is forecast to grow with 15 % Compound Annual Growth Rate (CAGR) through 2015. Figure 6a shows that from

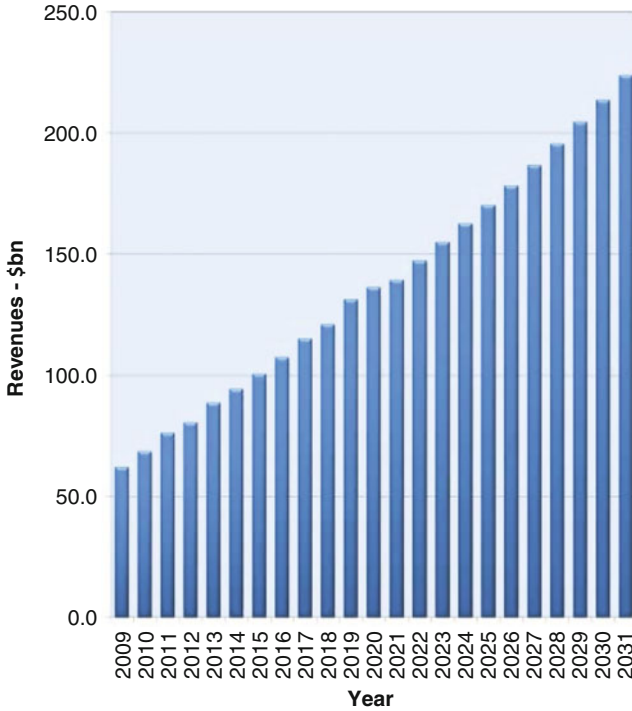
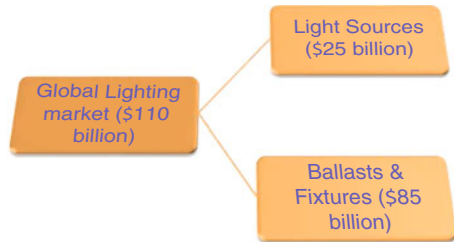


Fig. 4 World Lighting market (luminaires, lamps and lighting controls) forecast (Datapoint Press Release 2012)

Fig. 5 Global lighting market segmentation (2009 values)



approximately US\$ 23 billion in 2011 peaking at US\$ 41 billion in 2015 is projected to fall steadily to around US\$ 24 billion in 2020 (Tao 2012). Market saturation and decline are also reported by N. Bardsley based on a Navigant study (Bardsley 2013). Figure 6b shows this saturation and decline for different lamp technologies. It should be noticed here that even if between the two studies the tendencies are similar, the absolute values are rather different (there is a factor of 2 between the two studies). Values reported by Bardsley seem to be more reasonable.

World production of lighting fixtures (top 60 countries) is worth about \$64.4 billion for the year 2008, with 8.9 % increase in comparison with the previous year.

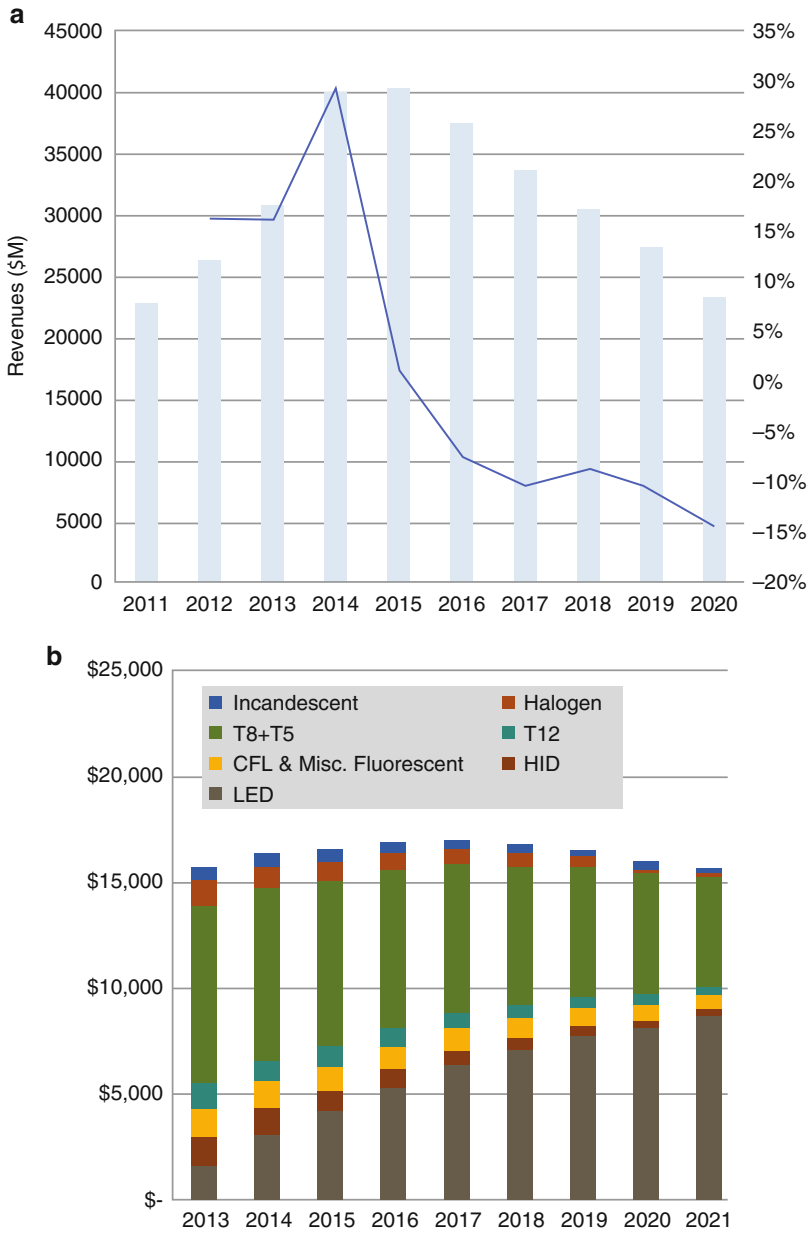


Fig. 6 (a) World lamp market revenue for lamps (all technologies) and CAGR (solid line) forecasts (Tao 2012) (b) lamp market size by technology (Bardsley 2013)

The ten major countries, in order of lighting fixtures production, are China, the United States, Japan, Germany, Italy, Mexico, the United Kingdom, India, Spain, and Brazil, which together produce 75 % in value of the world total production. The world trade of lighting fixtures can be estimated in \$26.5 billions of exports and \$27.3 billions of imports (the difference is mostly due to transport costs, included in the accounting from the import side). According to a recent study, 40 % of the world trade of lighting fixtures is linked to residential lighting and 60 % to technical lighting (Volpe 2010).

Concerning the legacy technology Light Sources market, it is dominated today by three majors: Philips (NL) with 20 % of the shares, Osram (D) with 18 %, and General Electric (US) with 10 % (Brunner 2006). These large lighting companies control the access to sales and distribution channels.

- Philips Lighting is the world number one of the lighting. It especially holds half of the market shares in Europe. The sales in 2005 rose up to 4.8 billion euros, an increase of 6.6 % with respect to the preceding year. The Company is employing 45,650 people.
- The German Osram, a subsidiary of the Siemens Group, is a company of 36,000 people, in 58 countries with a turnover of 4.2 billion euros. With a profit of 10.8 % of its turnover, Osram has reached the “highest level” since 5 years. Since its purchase of GTE-Sylvania (USA), in 1990, USA with 44 % of the sales is a priority target for Osram, followed by Europe (37 %) and by Asia-Pacific (16 %).
- General Electric Lighting (GEL, now part of Current by GE) is the American number one and represents 50 % of the market shares in the USA. Its turnover is about \$3 billions. The favorable target of GEL, since 10 years, is Europe and its market. To reach this goal, GEL has made many acquisitions: the Hungarian Thungsram, the British Thorn EMI, the Italian Sivi, and the German Linder Licht. The investments of the last 6 years exceed 750 million dollars.

These three major companies represent by themselves a turnover of 14 billion dollars. This type of structure is qualified of “oligopoly.” Other companies, like Havells-Sylvania and Matsushita, represent a turnover of only two billion dollars (this represents a turnover 10 times lower than the “big threes”). It remains that many national factories (of small or middle size) conserve a market and know-how, either in the source area or in the equipment area. On the other hand, the “lighting” market (first installation and maintenance) is always in the hands of “lighting designers” (engineers, architects, etc.) often linked to the principal parties such as the Government or the communities.

However, the landscape is rapidly changing since the arrival of Solid State Lighting revolution. Each of “big threes” has developed a strong presence in LED lighting through joint ventures with, and acquisitions of, specialty firms. Philips, for instance, acquired a large facility, Lumileds, in California thus becoming a major manufacturer of LED chips for use in the company’s own packaged LED lighting products; it also sells packaged chips to other firms. However, a couple of years from now Philips sold again Lumileds and all Philips Lighting is on the way to be spin-

offed from the mother company. OSRAM is a top manufacturer of LED components. Osram Opto-semiconductor branch has been established in 1999 and in September 2009 had a turnover of \$800 million and 4600 employees worldwide. It is established as market leader for Europe and a strong number 2 worldwide. General Electric is now in the LED business and remains on the lookout for partners. Once again, the domain is so rapidly moving that these lines have probably to be reviewed in a couple of years from now.

While these traditional lighting giants have so far played a leading role in the LED general lighting industry, they face competition from new LED lighting firms, especially in USA, Japan, Taiwan, Korea, and other Asian countries: Cree (US); Nichia and Toyoda Gosei (Japan); Samsung, LG Innotek, and Seoul Semi (S. Korea) are some of the top 10 LED lighting companies. A special mention has to be done for China, for which LED-Lighting industry growth becomes an important stake for the next years.

China lighting industry has maintained a rapid, continuous, and steady development trend for more than 10 years. Today, China is the largest lighting production country in the world. According to Chinese Association of Lighting Industry (CALI), there are over 10,000 lighting manufacturers in China. These manufactures are mainly distributed in China's south-eastern coastal areas, including Guangdong, Fujian, Jiangsu, Zhejiang, and Shanghai. Following Y. Yansheng (CALI), the sales volume of lighting products was \$55.6 billion in 2011. The export volume of lighting products from China was \$22.34 billion, which sets an historical record. The lighting products made in China have been exported to over 150 countries. Chinese export of luminaires represented in 2011 was 30 % of the global luminaires trade market (Yansheng and Chinese Association of Lighting Industry 2013).

In the past, Light Emitting Diodes penetrated gradually various market segments till saturation and then jumped from one saturated segment to another: For example, till 2004 the market was driven by nomad applications and the CAGR was maintained as high as 45 % per year. (Nomad applications: back lighting of mobile phones, laptops, torches, etc.) Then as this segment attained saturation, the global CAGR brutally reduced to 8 % per year for the next 4 years. Then, from 2008, the main market segments that LEDs try to conquer are backlighting for display applications (especially TV screens) and automotive. The projections show that this "jump" will guarantee a CAGR of around 11–12 % per year at least till 2016.

Based on these observations we can draw the following conclusion: LED market growth is "opportunistic" and "invasive." As has been seen in the following chapters of this report, LEDs represent today a significant market segment for the semiconductor and optoelectronics industry which is trying to counterpoise other classic market segments shrinking. Once LEDs reach a performance level that can be considered as satisfactory for a given application, the marketing effort is intensified and the market is invaded till saturation. The saturation is then followed by short lower-growth period till a new application is spotted and the cycle is starting again.

Looking at the historical data presented by various analysts (Fig. 7), we can see that the "marketing cycle" duration is in average less than 7–10 years: 3–7 years are necessary to saturate the market (following the global market size and the observed

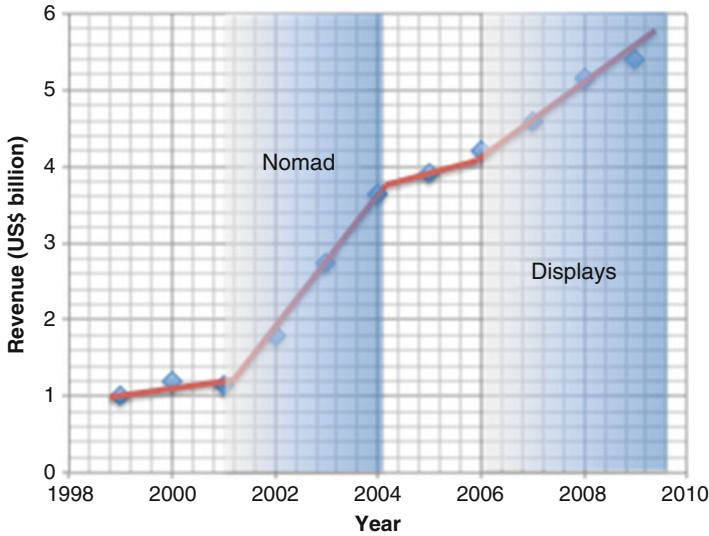


Fig. 7 Illustration of LED market growth mechanism based on historical revenues compiled from literature

OEM-Original Equipment Manufacturing life-span) and then 2–3 years are necessary to concentrate efforts to a new target. This cycle is rather short compared to observed values for OEMs (e.g., for lighting a product life cycle is estimated to be in the order to 25–30 years). This short cycle is however matching well with the performance evolution of the LED technology (R. Haitz’s law predicts that luminous flux will be multiplied by a factor of 2 every 18–24 months (Haitz et al. 2000)). Furthermore, semiconductor industry is producing “components” and not OEMs, thus the “product cannibalization” issue is rather marginal.

Assuming that this market growth mechanism is reproduced for each targeted application, we could speculate on what will happen for lighting during the next 5–10 years. Several forecasts predict a rapid growth period from 2009 to 2015 and that at saturation general lighting will represent 45–47 % of the global packaged LED industry revenue. Once again the estimated marketing cycle duration is less than 10 years. The ultimate question is then: “what will be the next LED target?”

References

- Archenhold G (2010) A new era for lighting, Light+Building. Integrated system technologies, 11-16.04.2010, Frankfurt (Germany)
- Bardsley JN (2013) Global impact of solid state lighting. CALI LED Forum, Shanghai
- Brown LR (2009) Plan B 4.0: mobilizing to save civilization. W.W. Norton, New York
- Brunner K (2006) LEDs for General Lighting Applications, EU Opera Project deliverable D4.3 presentation October 5th, 2006. Philips lighting

- Datapoint Press Release (2012) 'Perfect Storm' will drive the global lighting OEM market to \$108bn by 2016 & \$10.1bn in semiconductor revenues. <http://www.datapoint-research.com/index.php/market-research/by-application/lighting/item/304-perfect-storm-will-drive-the-global-lighting-oem-market-to-108bn-by-2016-10-1bn-in-semiconductor-revenues>
- Fouquet R, Pearson PJG (2006) *Seven centuries of energy services: the price and use of light in the United Kingdom (1300–2000)*. The Energy Journal, 27 (1). pp 139–177. ISSN 0195-6574
- Haitz R, Kish F, Tsao J, Nelson J (2000) Another semiconductor revolution: This time it's lighting! *Compd Semicond Mag* 6(2), 1–4
- Tao A (2012) LED lighting market overview. IMS Research, IHS Technology New York (USA)
- US Department of Energy (2011) Report on solid-state lighting research and development multi-year program plan
- US Energy Information Administration (2013) How much electricity is used for lighting in the United States? Washington DC (USA) <http://www.eia.gov/tools/faqs/faq.cfm?id=99&t=3>
- Volpe A (2010) Lighting fixture worldwide: a 50 billion Euro business, CSIL Special Report. Centre for Industrial Studies (CSIL), Milano, P. 3
- Waide P, Tanishima S (2006) Light's labour's lost: policies for energy-efficient lighting. IEA/OECD, Paris
- Yansheng Ch, Chinese Association of Lighting Industry (2013) Lighting industry in China. Global Lighting Association Conference, New Delhi, 9 Oct 2013

Life Cycle Assessment of Lighting Technologies

Leena Tähkämö and Heather Dillon

Contents

Introduction	936
Life Cycle Assessment	936
Methodology of LCA	937
Total Sustainability Assessment	940
Life Cycle Assessment of Light Sources	940
Environmental Impacts	944
Life Cycle Assessment of LED Lighting Products	946
Manufacturing	947
Use	950
End-of-Life	951
Challenges in LCAs of Lighting Technologies	952
Discussion and Conclusions	954
References	955

Abstract

Lighting is a major global energy consumer, and as such, it causes notable environmental impacts. The environmental impacts of lighting products are researched by life cycle assessment, a method that takes the whole life cycle of the product into account. It is important to study the product life cycle as whole so that the major environmental hot spots are identified and the environmental impacts are not shifted from one stage to another when choosing a different type of technology on the basis of environmental impacts.

L. Tähkämö (✉)

Lighting Unit, School of Electrical Engineering, Aalto University, Espoo, Finland
e-mail: leena.tahkamo@aalto.fi

H. Dillon

Donald P. Shiley School of Engineering, University of Portland, Portland, OR, USA
e-mail: dillon@up.edu

This chapter presents the basics of the life cycle assessment for evaluating the environmental impacts of light sources in particular. The typical results of the life cycle assessment of light sources in general are presented, but the chapter concentrates only on the lamps used in households. Household lighting is changing in several countries in the world from old, inefficient technologies (incandescent lamp) to more modern light sources of a higher luminous efficacy (CFLs, LED lamps). The change is often justified by environmental reasons. The environmental assessments of household lamps show clearly that the change from incandescent lamps to lamps of higher luminous efficacy is a beneficial decision from the environmental point of view.

Introduction

As a major global electricity consumer, lighting causes notable environmental impacts particularly due to the energy consumption in use. The electricity consumption in use accounts for approximately more than 90 % of the total life cycle environmental impacts of a light source. However, the energy consumption during use is not the only environmental impact of light sources but the total life cycle needs to be taken into account. The entire life cycle and its environmental impacts are evaluated in the life cycle assessment (LCA) method. The LCA enables the identification of the causes for environmental impacts over the life cycle of a product. The LCA may compare two or more products in order to verify which product is the most environmentally friendly. It is possible to concentrate on one product or even on one stage of the life cycle and to reduce the environmental impacts by changing the product design.

The basic method of the LCA is introduced in this chapter. The phases of the assessment are described after which the main results and special characteristics of the LCA of light sources are presented with the examples of LCA case studies. The LCA as a current methodology does not take the environmental impacts of light into account, but they are excluded. However, it must be noted that there are also other causes for environmental impacts than material or energy flows.

Life Cycle Assessment

Life cycle assessment is a tool to systematically evaluate the potential impacts of a product or a service over its life cycle. It collects the inputs and outputs for a system and can be used to quantify the environmental, energy, water, or cost impacts of a system. The formal LCA methodology was established in the 1990s. Numerous LCAs have been conducted and published on various products and services during the last two decades. The LCA provides a mechanism for evaluating the performance of the products for many purposes, such as the public procurement, enactment of the legislation, and the ecologically aware consumers to make a purchase decision.

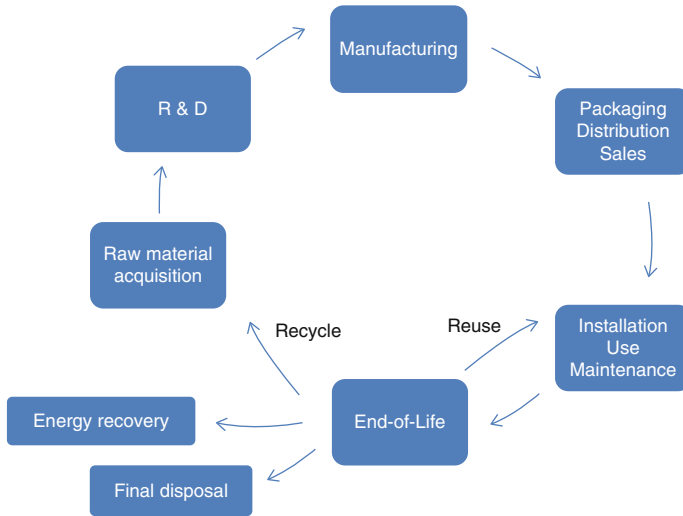


Fig. 1 Example of life cycle stages (Adapted from Tähkämö (2013))

The life cycle consists of several stages, e.g., raw material acquisition, manufacturing, distribution, use, maintenance, and end-of-life. An example of a product life cycle is presented in Fig. 1, where the life cycle starts from the raw material acquisition and ends in the end-of-life that is modeled to contain multiple alternatives including reuse, material recycle, incineration, and final disposal. It is possible to divide the processes in different ways; for example, the transport (distribution) may be tracked separately or in each process. The packaging, transport, and installation could also be combined as implementation. The LCA may be conducted concerning the whole life cycle or a part of the life cycle. The partial life cycle enables the comparison of certain stages of the life cycle in detail, while the LCA of a whole life cycle gives an overview of the total product impacts and is thus a holistic approach. The total LCA requires a large amount of data on each unit process in the system analyzed.

The stages considered in an LCA depend on the product system to be analyzed and the purpose of the assessment. No specific rules or recommendations exist for the unit processes in an LCA of light sources. However, a proxy may be used, such as the European Telecommunications Standards Institute (ETSI) 103 199 technical specification for LCA of information and communication technology (ICT) (ETSI 2011).

Methodology of LCA

There are three main types of the life cycle assessment: process LCA, economic input–output LCA (EIO LCA), and hybrid LCA. The process LCA is the conventional LCA method that evaluates the impacts as described in the following chapter. It concentrates on the examination of single processes in detail and is thus

a process-specific method. It enables product comparisons and identification of the improvement potential or the environmental “hot spots.” The EIOLCA is an input–output LCA that uses economic and environmental data to produce the LCA. It takes the entire industry sector into account and sets broad boundaries and scope of the product. It is a comprehensive technique. EIOLCA data is available for the US economies but less outside the USA, which restricts its use. Hybrid LCA combines the advantages of the two LCA methods as it may use EIOLCA for part of the processes and process LCA for the rest. In this way, the economy-wide effects are taken into account but also detailed data is used where possible.

The general LCA method is established in standards ISO 14040 (ISO 2006a) and 14044 (ISO 2006b). The standards introduce the procedure for conducting the LCA and define the basic terms, such as the functional unit. Yet, the LCA standards are sufficiently broad that they can be applied to any product or service. Due to the generic nature of the ISO LCA standards, there is often a need for more detailed rules for conducting an LCA of a certain product or service. These detailed rules are product category rules (PCR). No established rules exist for the lighting product parameters, e.g., the choice of functional unit and used energy sources for the LCA of light sources. For this reason, different authors may use different methods, but the reader should use caution when comparing the results of different LCA methods.

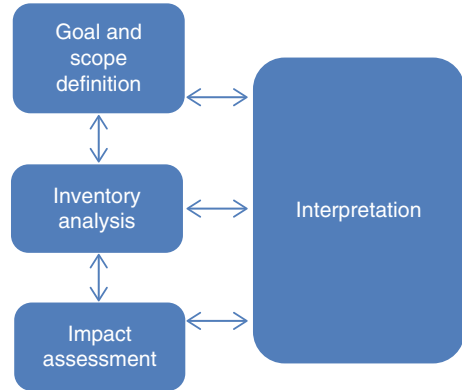
The LCA is framed on the use of *functional unit*. The functional unit is a unit to which the assessment is quantified and proportionated. It should be related to the function of the product. It enables the comparison of the environmental impacts of products in relation to their function. The functional unit is a key parameter in LCA, especially in comparative LCAs in which two or more products are compared to each other. There are no specific rules for choosing the functional unit, but ISO 14044 standard defines it to be “consistent with the goal and scope of the study,” “clearly defined,” and “measurable.” For example, in case of electricity production, the functional unit may be the production of 1 kWh of electrical energy. When it comes to light sources, an appropriate functional unit may be lumen-hours, since it takes both luminous flux and operating hours into account. The functional unit may be one piece of a lamp if the lamps are intended for the same application and possess comparable qualities, such as luminous flux, color characteristics, and luminous intensity distribution curve. To be more precise and to take the actual illumination into account, the functional unit of a light source may also consider the illumination on a surface, e.g., the illuminance on a 1 m² square surface at 1 m distance. However, in this case, the LCA should compare light sources of the same application.

The structure of a LCA process is described and defined in ISO 14040. The LCA process contains four phases as shown in Fig. 2.

Goal and Scope Definition

The goal and scope define the parameters of the assessment, such as the product system to be studied, the system boundaries, the functional unit, and assumptions used in the assessment. The system boundaries establish the inputs and outputs

Fig. 2 Phases of life cycle assessment (Adapted from ISO (2006a))



included in the LCA. Cut-off rules are also defined. The cut-off rules are needed, since it is often impossible to retrieve data on every input or output in the LCA. The inputs and outputs of the system may be cut off on the basis of their mass, energy, or environmental significance (ISO 2006b).

Inventory Analysis

The life cycle inventory (LCI) focuses on the data collection, data calculation, and allocation. The data is collected on the inputs, including energy, raw material, and ancillary inputs. The data is calculated relating it to the system by the functional unit. Allocation partitions the inputs and outputs between the product system in question and other product systems. Allocation is needed, since industrial processes that would yield a single output rarely exist. The inventory analysis is often performed using an existing commercial or freely available database of common unit processes and material inputs.

Impact Assessment

The life cycle impact assessment (LCIA) calculates the potential impacts. The impact assessment includes the selection of impact categories, category indicators, and characterization models. The LCIA assigns the LCI results into impact categories. There are numerous impact categories to choose from, but the most common for environmental LCA is the global warming potential. Other common categories for environment analysis include acidification potential, ozone depletion potential, and human toxicity potential (see Chapter [Environmental Impacts](#)). The LCIA should also include a data quality assessment that often takes the form of uncertainty and sensitivity analyses, and optional grouping and weighting of the results.

Interpretation

The interpretation phase combines the findings of the LCI analysis and the LCIA. It concludes the main findings in accordance with the goal and scope definition. The interpretation phase identifies the findings and presents them clearly and consistently.

Total Sustainability Assessment

Life cycle assessment may be considered as the umbrella term under which the economic and social impacts are, i.e., costs and social impacts, considered as environmental impact categories. On the other hand, the assessments may be defined and divided so that the umbrella term is total sustainability assessment in which environmental, economic, and social aspects form its three pillars. Sustainability, or sustainable development, aims at meeting the needs of the present generation without compromising the ability of future generations to meet their needs. The sustainability may refer to environmental, economic, and social needs. The life cycle sustainability assessment (LCSA) is defined as

$$\text{LCSA} = \text{LCA} + \text{LCC} + \text{SLCA} \quad (1)$$

in which LCA stands for the (environmental) life cycle assessment, LCC for life cycle costing, and SLCA for social LCA (Swarr et al. 2011; Klöpffer 2003, 2008).

The life cycle cost analysis has the longest history of the three pillars, as the monetary values have been of interest for decades. However, the life cycle costing as a sustainability measure may differ from the conventional costing. The life cycle costing may be similar with the conventional costing by including, for example, the time value of the money and calculating present value. The environmental approach may be included in the costs, e.g., in the costs of environmental protection.

The total sustainability assessment is a large, challenging entity to calculate over a product. Yet, it gives a very profound overview to the sustainability of a product system. The environmental and economic assessments are rather established techniques. They can be conducted from different perspectives, e.g., from the manufacturer's, consumer's, or municipality's point of view (Swarr et al. 2011). The social life cycle assessment suffers from difficulties in establishing the methodology and lack of data, and it is currently being developed, as the general interest in it is increasing in sustainability discussions (Klöpffer 2008). The social aspects include organization-specific aspects, and they may be classified according to the stakeholders, such as the workers, the society, and the customers, or to the impact categories, such as human rights, health and safety, and the cultural heritage (UNEP 2009). No international standards exist for SLCA.

Life Cycle Assessment of Light Sources

The environmental impacts of lighting may be studied from several points of view. The environmental impacts of the total life cycle of light sources have been studied in a few LCAs, which represent the product life cycle approach. Another approach is the comparison of use-stage performance, e.g., on the basis of energy consumption or related costs. This is a streamlined, simplified LCA. The environmental impacts of the light sources may be studied from the point of view of the light itself and its environmental impacts. Light causes multiple effects on human beings,

fauna, and flora. The effects depend on the time, location, and characteristics of the light. There is no method for quantifying the environmental impacts of light in an LCA. Neither the visual characteristics, such as correlated color temperature or color rendering index, are included in the previous LCAs, but they affect the application of the light source, and they should be at least discussed.

Light sources, i.e., lamps and luminaires, have been the subject in several life cycle assessments during the last two decades. A summary of the previous LCAs is presented in Table 1. The early studies have mainly compared the incandescent lamp and the compact fluorescent lamp (CFL) (e.g., Gydesen and Maimann 1991; Pfeifer 1996; Parsons 2006; BIOIS 2003), while the more recent assessments include also LED light sources (U.S. DOE 2012a, b; Osram 2009; Quirk 2009) or even a wide range of lighting technologies (DEFRA 2009; Dale et al. 2011).

A variety of functional units have been used in the LCAs of light sources. The functional unit is typically expressed in megalumen-hours, e.g., 1 Mlmh (Table 1). The lumen-hour seems to be an appropriate functional unit for light sources, as it considers both the burning hours and the luminous flux. The luminous flux of an incandescent lamp remains constant during its life. In contrast, the luminous flux of a lamp of fluorescent, high intensity discharge, or LED technology is not constant but depreciates over the operating time. None of the LCAs in Table 1 take the depreciation of the luminous flux into account in the calculations, yet some of the assessments acknowledge it (DEFRA 2009; Osram 2009; Slocum 2005). The lumen depreciation is stated to be too small to impact the results (Osram 2009). A lighting engineering approach for functional unit is presented by Yabumoto et al. by using two functional units: total luminous flux of 800 lm during 40,000 h and 100 lx floor illuminance at a distance of 1 m directly under the light source during 40,000 h (Yabumoto et al. 2010). In a methodology study of comparing nondirectional lamps (incandescent, CFL, LED lamp) (Tähkämö 2013; Tähkämö et al. 2012b), it was found that using Mlmh, hour, or illuminance as the functional unit did not significantly change the results of the comparison.

The data has been collected on the material contents of the light sources used in the LCAs (Tähkämö 2013). The materials of incandescent lamps were divided into glass (70–94 % of the total weight of the lamp) and metals (4–29 %). The weights of incandescent lamps varied between 15 and 38 g. No correlation between the weight and the wattage was found: the weight of the 60 W incandescent lamps ranged between 23 and 38 g. The weight of the CFLs was found to range between 46 and 120 g, and no correlation was found between the lamp weight and wattage. CFL consisted of glass (30–73 % of the weight of the lamp), metals (2–40 %), electronics (up to 31 %), and plastics (16–38 %). However, there are differences in the grouping of the materials in the references. For instance, electronic components may have been modeled as metals. The amount of mercury was between 3 and 5 mg per CFL. Only few references were found that provided the detailed material data for LED lamps. These references showed that LED lamps contained glass (0–13 % of the weight of the lamp), metals (45–78 %), electronics (3–21 %), and plastics (2–37 %).

Despite the differences found in the previous LCAs of light sources, the findings of the assessments were unanimous on two things: the use-stage energy consumption

Table 1 Summary of previous life cycle assessments of light sources (*IL* incandescent lamp, *HL* halogen lamp, (*CFL*) compact fluorescent lamp, *CFLi* CFL with integrated ballast, *CFLni* CFL with nonintegrated ballast, *CMH* ceramic metal halide lamp, *IND* induction lamp luminaire, *GWP* global warming potential, *AP* acidification potential, *EP* eutrophication potential, *POCP* photochemical ozone creation potential, *ODP* ozone depletion potential, *HTP* human toxicity potential, *ADP* abiotic (resource) depletion potential, a = future, b = hypothetical) (Adapted from Tähkämö (2013), based on Tähkämö et al. (2012a))

Light sources		Functional unit	Environmental impact categories	Reference
60 W IL	15 W CFL	10,000 h	ADP, AP, EP, GWP, ODP, POCP	Elijošūtė et al. (2012)
60 W IL 15 W CFL	12.5 W LED lamp 6.1 W LED lamp ^a	20 Mlmh	GWP; AP; POCP, ODP; HTP; freshwater aquatic, marine aquatic, and terrestrial ecotoxicities; EP; ecosystem damage; ADP; land use; hazardous, nonhazardous, and radioactive wastes	U.S. DOE (2012b)
60 W IL 35 W HL	14 W FL 11 W CFL	1 h of lighting	Cumulative energy demand, GWP, EcoIndicator'99	Welz et al. (2011)
150 W HPS 163 W MH	109 W IND 105 W LED	100,000 h of light	GWP, respiratory effects, ecotoxicity	Dale et al. (2011)
100 W IL 23 W CFL 2 × 28 W FL	20 W CMH 10 W LED lamp 16 W LED luminaire	1 Mlmh	GWP; AP; POCP, ODP; HTP; freshwater aquatic, marine aquatic, and terrestrial ecotoxicities; EP; ecosystem damage; ADP; land use; hazardous, nonhazardous, and radioactive wastes	DEFRA (2009)
40 W IL 8 W CFL	8 W LED	345–420 lm during 25,000 h	GWP, AP, POCP, HTP, EP, ADP, energy consumption	Osram (2009)
60 W IL 13 W CFL	6 W LED 6 W LED ^a	1 Mlmh	Primary energy consumption, GWP	Quirk (2009)
60 W IL	15 W CFL	1 kWh	Energy consumption	Landis et al. (2009)
60 W IL 13 W CFLi		500–900 lm during 10,000 h	Minerals, fossil energy sources, land use, GWP, EP, AP, ODP, POCP, ecotoxicity, respiratory effects, ionizing radiation, carcinogens	Michaud and Belley (2008)
100 W IL	23 W CFL	16 Mlmh	GWP, emissions of mercury, arsenic, and lead	Ramroth (2008)
100 W IL 18 W CFL		Equivalent luminous flux during 8,000 h	ADP; GWP; ODP; HTP; AP; EP; POCP; freshwater aquatic, marine aquatic, and terrestrial ecotoxicities; carcinogens; respiratory effects; minerals; fossil fuels	Parsons (2006)

(continued)

Table 1 (continued)

Light sources		Functional unit	Environmental impact categories	Reference
60 W IL 15 W CFL	7.5 W LED ^b	1 Mlmh	Energy consumption	Slocum (2005)
60 W IL 15 W CFLi	13 W CFLi 11 W CFLni	10 Mlmh	GWP, AP, primary energy, ADP, ODP, POCP, EP, HTP, ecotoxicity, costs of environmental impacts, metals, carcinogens	BIOIS (2003)
60 W IL 11 W CFLi	13 W CFLi 11 W CFLni	1 Mlmh	Primary energy consumption, Hg emissions, radioactive materials	Pfeifer (1996)
60 W IL	15 W CFL	1 Mlmh	GWP, SO ₂ , NO _x , CH ₄ , ashes, Hg, solid waste	Gydesen and Maimann (1991)

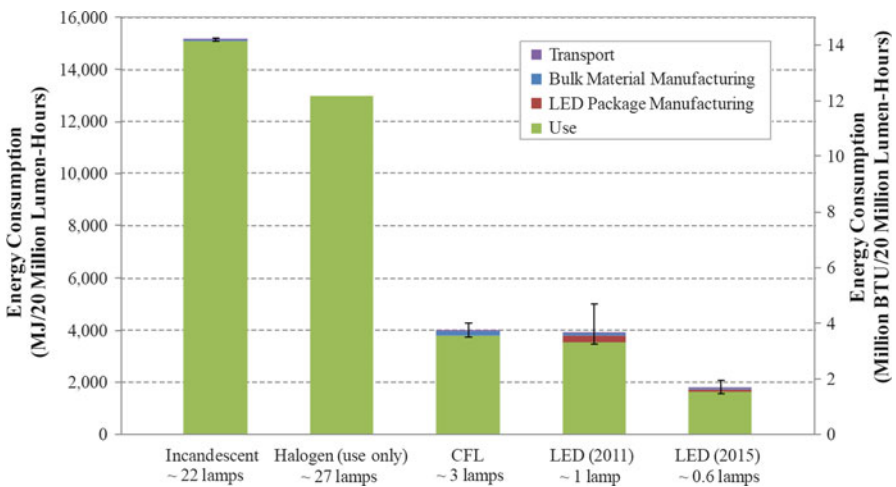


Fig. 3 Life cycle energy of incandescent lamps, CFLs, and LED lamps (DOE 2012a). The error bars indicate the variation between the 10 LCA reports summarized in the study

is the most important environmental aspect, and thus, the energy-efficient light sources, such as the CFLs and LED lamps and luminaires, are more environmentally friendly than their conventional counterparts from the life cycle point of view. This is also shown in Fig. 3, a summary of ten LCA results on the basis of energy compared by the US Department of Energy (U.S. DOE 2012a). The lamps are here compared on the basis of their primary energy consumption in manufacturing and use. Figure 3 illustrates that the primary energy consumption of incandescent and halogen lamps are clearly greater than the ones of CFL or LED lamps.

Generally, the LCAs of light sources include the raw material acquisition and manufacturing (often combined), use, and end-of-life. The relative impact of the use is expected to be reduced when the luminous efficacy of the light source increases. This increases the significance of other life cycle stages, notably the manufacturing and raw material acquisition, in terms of environmental impacts. In addition, the relative significance of the use-stage electricity consumption will be reduced when the electricity production is shifted towards low-emission electricity production, such as renewable energy sources or nuclear power.

Environmental Impacts

There are numerous environmental impacts to consider in an LCA. The LCIA methodologies introduce several impact categories from which to choose. The chosen environmental impact categories depend on the scope and purpose of the LCA. In general, it is recommended that authors include several impact categories in the LCA so that environmental impacts are taken into account in a wide range. In addition, it is possible to calculate the environmental impacts as single-scale indices, such as Ecoindicator'99, that weigh and factor several impact categories into one number.

The following subchapters describe briefly the most common environmental impact categories used in LCAs.

Acidification

Acidification is caused by the emissions of sulfur dioxide and nitrogen oxides. These oxides form acids in the atmosphere with water vapor and fall down as acid rain, acid snow, or dry acid depositions. It acidifies water and soil, corrodes buildings, and affects vegetation. It is measured in kilograms of SO₂ equivalents.

Eutrophication

Eutrophication is caused by nitrogen and phosphorus that originate from landfills, sewage, and fertilizers. Eutrophication causes excessive plant growth and oxygen depletion in water. It is measured in kilograms of PO₄ equivalents.

Global Warming

Global warming refers to the enhanced greenhouse effect. The atmosphere retains the heat due to the greenhouse gas emissions, such as carbon dioxide, methane, nitrous oxide, hydrofluorocarbons, perfluorocarbons, and sulfur hexafluoride. Global warming causes global impacts, such as melting of the polar ice, change in ocean and wind patterns, droughts, and floods. It is measured in kilograms of CO₂ equivalents.

Land Use

Land use refers to the occupation of the land and the change in land use. It relates to the loss of wildlife habitat and decrease in the land space. Land use affects the biodiversity. It is measured in square meter years (m²a).

Ozone Depletion

Ozone depletion refers to the thinning of the stratospheric ozone layer. It is caused by chlorinated and brominated substances, such as chlorofluorocarbons (CFCs) and halons. Ozone depletion increases the ultraviolet radiation on the surface of the earth. It is measured in kilograms of CFC-11 equivalents.

Photochemical Ozone Creation

Photochemical ozone creation (photochemical smog, summer smog) originates from the reaction of volatile organic compounds (VOCs) and nitrogen oxides with heat and sunlight in troposphere. It is formed generally in urban areas during summer. It decreases visibility, causes respiratory effects, and damages vegetation. It is measured in kilograms of ethylene equivalents or of formed ozone.

Resource Depletion

The depletion of natural resources describes the consumption of natural resources, such as fossil fuels and minerals. It may be divided into renewable and nonrenewable resources or to biotic and abiotic resources. Abiotic resource depletion is measured in kilograms of antimony equivalents.

Toxicities (Human Toxicity, Aquatic Ecotoxicity, Terrestrial Ecotoxicity)

Toxicities are caused by many substances, such as dioxins, heavy metals, and hydrochloric acid. The challenge in the toxicity categories is to know which quantity is harmful and in which time frame (long-/short-term impacts). The variety of toxicity categories enables the consideration of a specific toxicity target, e.g., marine or freshwater, aquatic or sediment, terrestrial or human. All the toxicity categories are measured in kilograms of 1,4-dichlorobenzene equivalents.

Waste

There are several waste categories, such as solid, radioactive, hazardous, and nonhazardous. They are usually measured in kilograms of waste.

Water Use

Water use reduces the availability of groundwater and surface water resources. It is measured in liters of water.

In addition to the environmental impact categories, it may be useful to calculate the LCIA results in primary energy consumption. Primary energy is the energy embodied in natural resources (raw materials, energy) that has not gone through any transformation. It is generally measured in joules (J). There are also other environmental impacts, such as noise and odor in air and water. In case of light sources, additional environmental impacts include the light pollution and the impacts of light on living organisms. However, they are difficult to calculate in a relative manner in an LCIA.

A recently completed LCA (U.S. DOE 2012b) compared the relative environmental impact of LED lighting products when compared to CFL and incandescent as shown in Fig. 4. Many of the environmental impact categories described above are

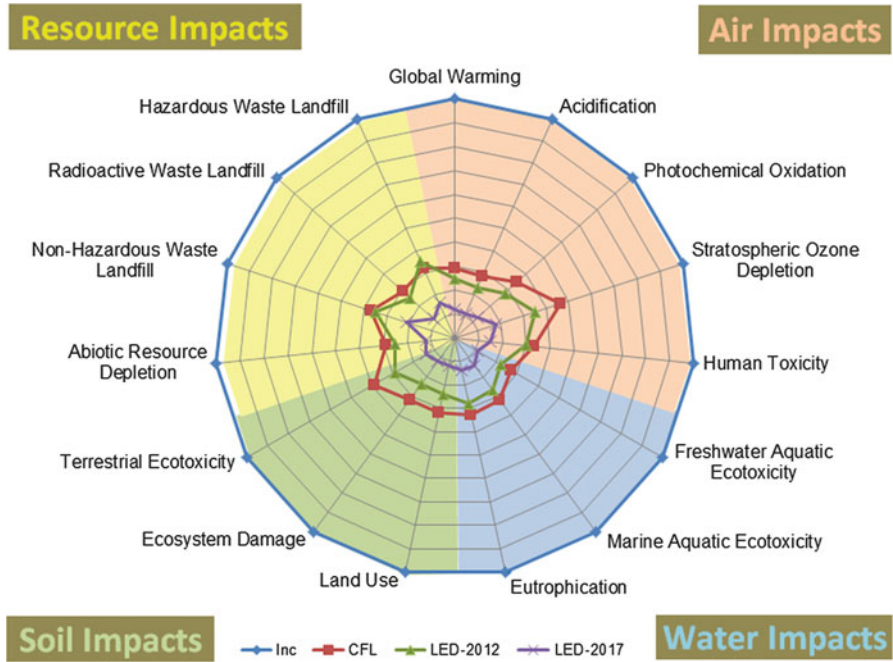


Fig. 4 Life cycle assessment impacts of the lamps analyzed relative to incandescent lamp (U.S. DOE 2012b). The data in the graph are normalized for the quantity of lighting service (20 Mlmh)

shown relative to the incandescent lamp on a relative scale. While it has substantially lower impacts than incandescent lamp, the CFL is slightly more harmful than the 2012 integrally ballasted LED lamp against all but one criterion – hazardous waste landfill – where the manufacturing of the large aluminum heat sink used in the LED lamp causes the impacts to be slightly greater for the LED lamp than for the CFL. The best performing light source is the projected LED lamp in 2017, which takes into account several prospective improvements in LED manufacturing, performance, and driver electronics (U.S. DOE 2012b).

Life Cycle Assessment of LED Lighting Products

The environmental impacts of LED lighting products have recently been studied in an LCA by the US Department of Energy (2012b). They compared three household lamp technologies: incandescent, CFL, and LED lamp. However, the especially valuable part of the study for the LCA community is the detailed description of the LED product manufacturing and its material and energy flows, since until this publication, there was no up-to-date data on LED manufacturing freely available. Figure 5 illustrates an example of the system boundaries of an LED lamp. The LED

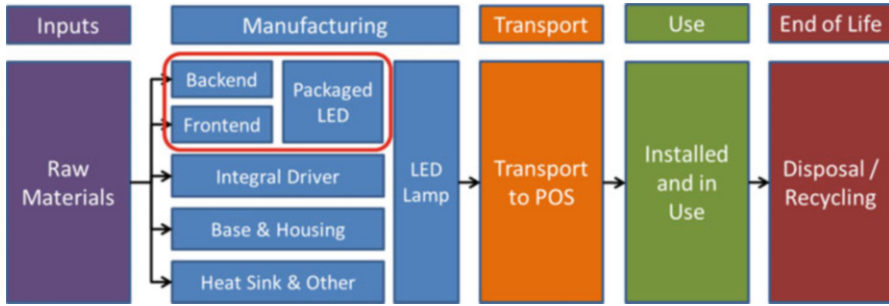


Fig. 5 Example of a system boundary of the life cycle assessment of LED lighting products (U.S. DOE 2012b)

lamp LCA and notably the manufacturing process are described in the following subchapters. The study had also two additional parts: the energy assessment of light source LCAs (U.S. DOE 2012a) and the report on the environmental testing of incandescent, CFL, and LED lamps (U.S. DOE 2013).

Manufacturing

The manufacturing of an LED package for lighting applications is the least defined part of an LCA of LED lighting products. The proprietary nature of most manufacturers' products has limited the typical methods for inventory analysis. The manufacturing process of LED lighting products was recently quantified in detail (U.S. DOE 2012b). In this work, the manufacturing was broken into three parts: (1) substrate production, (2) LED die fabrication, and (3) packaged LED assembly (Fig. 6).

Substrate Production

The substrate production focuses on preparing polished, cleaned sapphire wafers to use in a metal organic chemical vapor deposition (MOCVD) reactor for LED die fabrication (U.S. DOE 2012b).

The processing steps for sapphire wafers are described in more detail in the report (U.S. DOE 2012b). The energy and material summary for this unit process is shown in Table 2. This table provides both the quantity consumed per wafer both in terms of volume and in terms of mass.

LED Die Fabrication

The LED die fabrication process is divided into epitaxial growth and other front-end processes. In the epitaxial growth, the substrate is mounted in an MOCVD reactor, and it is heated, followed by the deposition of the nucleation layer, the n-type layer, the active layers (multi-quantum well), and finally the p-type layer. The result of this process is the LED epitaxial wafer (U.S. DOE 2012b).

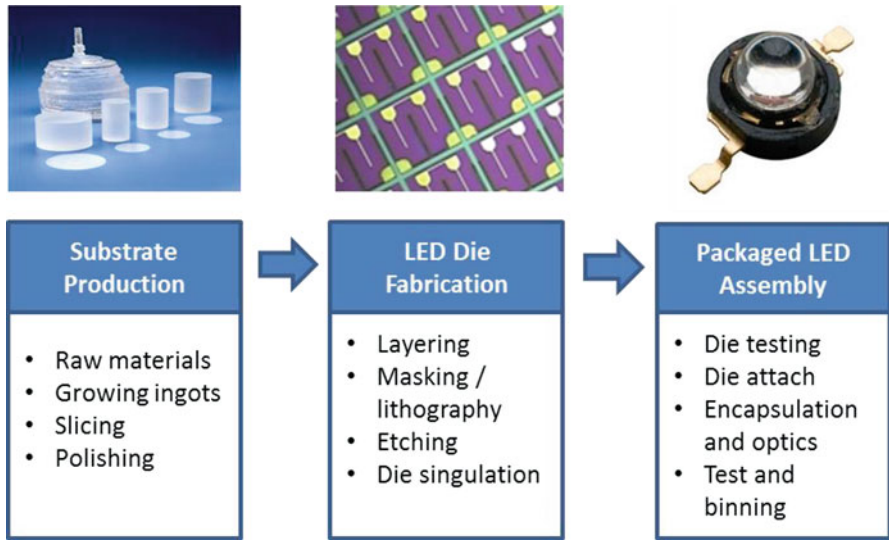


Fig. 6 Three main stages of packaged LED manufacturing and the major steps within (U.S. DOE 2012b)

Table 2 Energy and material consumption for three-inch sapphire wafer manufacturing (Adapted from U.S. DOE (2012b))

Stage	Material used	Amount	
		Volume per wafer	Mass per wafer
Material	Alumina (Al ₂ O ₃)	16.6 g/wafer	16.6 g/wafer
Material	Cleaning chemical (alkali detergent)	3.5 l/wafer	3.5 kg/wafer
Production	Energy consumption	18.3 kWh/wafer	18.3 kWh/wafer
Material	Diamond slurry	830.0 g/wafer	0.83 kg/wafer
Material	Water	105.3 l/wafer	105.3 kg/wafer

From the LED epitaxial wafer, several steps are needed to make the device and to prepare it for packaging. The wafer is inspected, subjected to masking and lithography, and etched, and then the contacts are attached (metallization) on the LED. These steps create the LED mesa-structure, and it results in visible LED dies on the wafer. Once these are developed, the substrate is separated from the LED dies. The dies are cut (die singulation) and tested and binned according to their performance. After these steps, the LED dies are ready to be packaged (U.S. DOE 2012b).

Table 3 summarizes the amounts of materials and energy consumed in the LED die fabrication. The table combines the material and energy consumption of both the epitaxy and p-n junction deposition stage and post-epitaxy steps associated with contacts, patterning, substrate removal, and preparing the finished LED die.

Table 3 Energy and material consumption for LED die fabrication (U.S. DOE 2012b)

Material	Quantity consumed	
	Volume/wafer	Mass/wafer
Acetone	0.59 l/wafer	467 g/wafer
AuSn solder	14.8 mm ³ /wafer	0.29 g/wafer
Developer	115 ml/wafer	115 g/wafer
Etchant Ag	30 ml/wafer	30 g/wafer
Etchant metal	60 ml/wafer	60 g/wafer
GaN etchant	0.192 l/wafer	192 g/wafer
H ₂	1.62 m ³ /wafer	136 g/wafer
N ₂	4.42 m ³ /wafer	5,527 g/wafer
NH ₃	0.447 kg/wafer	447 g/wafer
O ₂	2 l/wafer	2.3 kg/wafer
Photoresist	19 ml/wafer	19 g/wafer
Energy	42.57 kWh/wafer	42.57 kWh/wafer
SF ₆	0.1 l/wafer	13 g/wafer
SiH ₄	0.242 g/wafer	0.242 g/wafer
Slurry	2.3 l/wafer	2.3 kg/wafer
Target Ag	0.44 mm ³ /wafer	0.005 g/wafer
Target Al	1.27 mm ³ /wafer	0.003 g/wafer
Target Ni	0.417 mm ³ /wafer	0.004 g/wafer
Target Ti	0.467 mm ³ /wafer	0.002 g/wafer
Target W	3.089 mm ³ /wafer	0.06 g/wafer
TMAI	0.003 g/wafer	0.003 g/wafer
TMGa	1.47 g/wafer	1.47 g/wafer
TMIn	0.01 g/wafer	0.01 g/wafer
UPW	240 l/wafer	240 kg/wafer

Packaged LED Assembly

The third phase of LED manufacturing is referred to as the packaging of the device. A LED package is shown in Fig. 7. The packaging process includes the mounting of the LED die in housing, making electrical connections, and applying phosphor, encapsulant, and optics. In addition, the LED is tested and binned into the correctly classified product (U.S. DOE 2012b).

The substrate is cut into the individual packaged LEDs for use. Table 4 presents the aggregate consumption per LED produced including all the inputs for LED packaging and assembly.

Lamp Assembly

After the packaged LED, a self-ballasted LED lamp is created from several packaged LEDs. This self-ballasted LED lamp may be inserted into a mains voltage socket without auxiliaries.

An example of the LED lamp assembly was provided in US DOE (2012b) for the Philips EnduraLED lamp introduced in 2011. This particular LED lamp was commonly available in the US market in 2012. Table 5 presents the materials used in manufacturing of the LED lamp, the energy involved in the assembly and

Fig. 7 Example of the finished packaged LED, the Philips Luxeon Rebel (U.S. DOE 2012b)

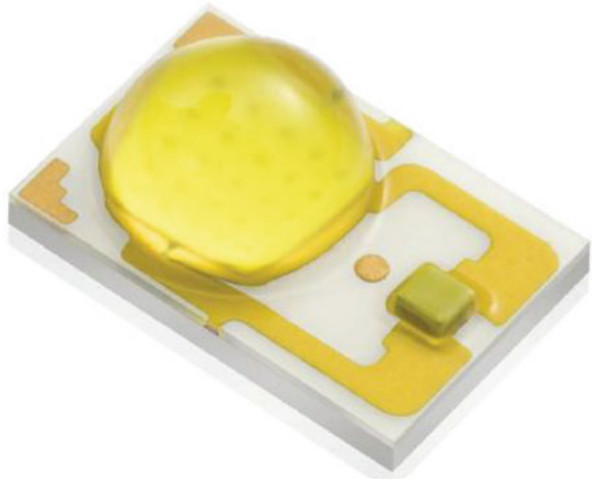


Table 4 Energy and material consumption for LED packaging assembly (Adapted from U.S. DOE (2012b))

Stage	Material used	Amount	
		Volume per LED	Mass per LED
Material	Ceramic substrate (2-layer alumina)	13.5 mm ² /LED	0.0135 g/LED
Production	Energy (kWh)	0.03 kWh/LED	0.03 kWh/LED
Material	ESD diode (Silicon)	0.22 mm ² /LED	0.055 g/LED
Material	Gold	0.004 mm ³ /LED	0.00006 g/LED
Material	Underfill	0.05 mm ³ /LED	0.0196 g/LED
Material	Silicone	8.4 mm ³ /LED	0.00006 g/LED

manufacturing, the estimated transportation of the lamp, the use-stage energy consumption during the lifetime of the lamp (12.5 W, 25,000 h), and the recycling rates of the lamp and packaging materials. The finished LED lamp weighs 178 g and the card-stock packaging 37 g.

There is a lack of information on the extent to which materials used in the manufacturing of LEDs are reused and recycled. If these materials are recovered, processed, and then reused, this would reduce the per unit production environmental impacts. To make the LCA conservative, relatively low rates of recycling or reuse of material is often assumed. To the extent that materials are recovered and recycled, the environmental impacts will be less than those reported in an LCA that uses conservative estimates for recycling.

Use

The use of the light sources – LED or other technologies – causes environmental impacts that depend strongly on the energy source. In case of a renewable,

Table 5 Life cycle inventory for an example 12.5 W LED lamp in 2012 (Adapted from U.S. DOE 2012b)

Stage	Material used	Amount	Stage	Material used	Amount
Material	LEDs (blue light)	12 units	Material	Resistor SMD	35 pcs
Material	Remote phosphor	1.0 g	Material	Resistor	3 pcs
Material	Plastic phosphor host	11.1 g	Material	Transistor	6 pcs
Material	Aluminum heat sink	68.2 g	Material	Resin glue	4.5 g
Material	Copper	5.0 g	Material	Solder paste	0.3 g
Material	Nickel	0.003 g	Production	Power	5.0 MJ
Material	Brass	1.65 g	Production	Manufacturing	178 g
Material	Cast iron	4.0 g	Material	Packaging	37 g
Material	Chromium	0.0002 g	Transport	Sea – 215g	10,000 km
Material	Inductor	5 pcs	Transport	Road – 215g	1,000 km
Material	IC chip	2.0 g	Use	Energy in use	312.5 kWh
Material	Capacitor SMD	8 pcs	End of life	Lamp, recycling	20 %
Material	Electrolytic capacitor	6 pcs	End of life	Lamp, landfill	80 %
Material	Diode	6 pcs	End of life	Package, recycling	30 %
Material	Printed wiring board	15.0 g	End of life	Package, landfill	70 %

low-emission energy sources, the environmental impacts are significantly lower compared to nonrenewable, high-emissions energy sources, such as coal. For example, in CFLs, the primary source of mercury on an LCA basis in the USA is driven by the upstream production of electricity from coal power plants that emit mercury rather than in the lamp.

It is important to note different electricity productions used in the LCA. The energy consumption in manufacturing is often modeled using an average electricity production for China (if product manufactured in China), while use is modeled as another electricity production, e.g., the one in the USA or Europe. It is important that the energy in use stage reflects the mix where the lamp is being actually used because the magnitude of the impact associated with the electricity consumed in use has been found to be very important.

End-of-Life

The end-of-life (EoL) includes several alternatives for the reuse or disposal of the product. The product or part of the product may be reused after repair or maintenance to prolong its lifetime. In many cases, the materials of the product could be recycled into “new” raw materials. The energy embodied in the certain materials, e.g., most plastics, may be utilized by incineration.

The EoL is a complex stage of the life cycle to model in an LCA. It may contain several possibilities (scenarios and recycling techniques), and the inclusion of by-products or recycled raw materials makes the calculation challenging. The EoL

stage shall include also the collection of the products from the users and the separation of the material fractions.

Reuse

After reaching the end of useful lifetime (usually 70 % of initial luminous flux), the LED luminaire is “scheduled” to be replaced in an optimal replacement scenario. At this point, the LED luminaire is still working, but the luminous flux has deteriorated so that it needs to be replaced. For this kind of situation, it is not likely for the LED luminaire to be repaired so that it would be used for a longer period of time because usually LED luminaires are not modular, and thus, there is no part to replace. Yet, the modularity may become more common thanks to global standardization collaboration regarding the LED products. In case of an LED lamp, the situation is similar: At the end of the useful lifetime, the whole product, i.e., the lamp, needs to be replaced.

Recycling

The material fractions of the LED product (lamp or luminaire) need to be separated, e.g., by dismantling and shredding. Screw fastening enables efficient separation of parts. The aluminum heat sink is a part of the LED product that is especially important to recycle due to the energy-intensiveness of the production of virgin aluminum. It has been found out that the CFL products outperform LED products in an environmental impact category primarily because of the aluminum heat sink (U.S. DOE 2012b).

The electronic components in the LED products should be recycled as waste electrical and electronic equipment (WEEE). There are no specific recycling processes for LED products, since the amount of LED products to be recycled is currently low. LED products are collected together with other light sources (fluorescent lamps, other discharge lamps) or as WEEE.

Landfill

Landfill is the worst option of the product to end up in from the raw material point of view, but it is useful to evaluate it in an LCA to determine a worst-case scenario for a product. There is often significant embodied energy in raw materials that would not be utilized in a landfill. Some countries and regions have adopted mandatory requirements that would prevent any WEEE from entering a landfill due to high concentrations of copper, aluminum, and other metals.

Challenges in LCAs of Lighting Technologies

Several challenges have been identified in the LCAs of light sources. First, the energy source affects the significance of the use-stage environmental impacts. It causes uncertainty in the results affecting the stage of the life cycle that has typically the greatest impacts – use. Using a specific energy source may distort the LCA results, while the use of average energy production of an area (country, state, continent) brings the average results. Second, there is a need for more detailed

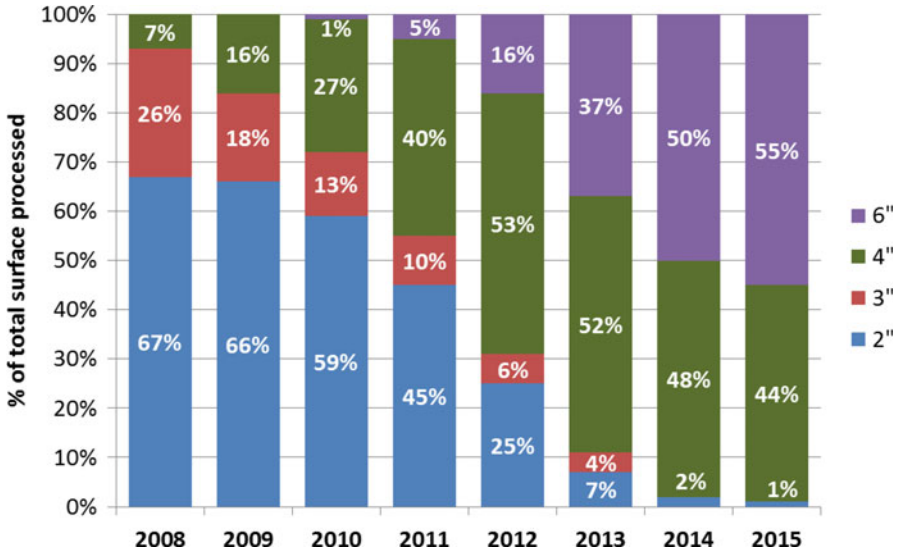


Fig. 8 Trends in diameter of sapphire substrates for LED manufacturing (U.S. DOE 2012b)

data in the environmental databases for manufacturing of light sources, especially the LED component. Up-to-date data is essential for reliable LCA results, and the industry is developing LED manufacturing methods at a rapid pace (e.g., sapphire substrate diameters in Fig. 8). Third, there is a wide variety of LED products on the lighting market: luminaires, lamps, modules, arrays, and components. This makes the comparison of products difficult. Yet, the comparison of different products is possible if the functional unit is carefully chosen. Finally, there are environmental impacts of light sources that cannot be taken into consideration in current LCA methodology: the environmental impacts of light itself.

There are uncertainties in the LCA of electronic products due to the complexity of the components. The electronic products are developed fast, and there is practically immeasurable number of different electronic components on the market. In addition, there is not always detailed, accurate data available on the exact component in a certain geographical location. Thus, it is not possible to analyze every component in detail, but proxies are necessary. For example, assuming a smaller wafer diameter will make the LCA conservative, while assuming larger diameter will make the LCA more accurate but may underestimate energy or environmental impacts. Some LCA authors have chosen to bind a study by including more modern estimates as a separate case. As detailed the life cycle inventory of a self-ballasted LED lamp presented in this chapter is, it is likely to need updating and possibly elaborating to cover other LED manufacturing technologies in the future.

There are only two unit process estimates currently available for manufacturing of an LED. Prior to 2012, all LCA studies had been based on manufacturing estimates for an indicator LED based on LED manufacturing technology from 2007.

The indicator LED was found to have a luminous flux of 4 lm, while the high-brightness LED was found to have a luminous flux of 100 lm (Radio-Electronics 2012; Philips 2012). A study showed that the environmental impacts were reduced by 94.5 % on average in a per-lumen comparison of the 2007 indicator LED and the 2012 LED (U.S. DOE 2012b). It was concluded that the high-brightness LEDs manufactured in 2012 are significantly less harmful for the environment than the 5 mm indicator LEDs produced in 2007.

Discussion and Conclusions

The LCAs of light sources typically conclude that the use (energy consumption) of the light source causes the greatest environmental impacts. This is, however, sensitive to the used energy source, and the environmental impacts differ if low-emission or high-emission energy source is used. Other stages of the life cycle tend to cause only small environmental impacts in the scope of total life cycle. Yet, more detailed modeling is recommended especially regarding manufacturing and end-of-life of the product. In addition, the shift towards renewable energy sources that is happening globally will change the dynamics of the LCA of light sources and generally all energy-using products. Thus, it is likely that manufacturing and end-of-life become more important in the LCA in the future.

Lighting design requires numerous factors to be taken into account. They depend on the lighting application, since the requirements differ in different applications, such as indoor, outdoor, road, area, general and local lighting. On the basis of the LCAs of light sources, it can be concluded that the main parameter in the design of environmentally friendly lighting is the luminous efficacy of the light source. The higher the luminous efficacy is, the lower the life cycle environmental impacts the light source has on the average. In design of a light source, other environmental parameters may be considered to reduce the environmental impact, such as reducing the weight of the light source, designing for easy dismantling (for recycling), avoiding the use of hazardous materials, and ensuring the long operating life by design of the electronic components and the heat transfer.

As light sources and other lighting-related products have been the subject of several LCAs, new topics may be introduced in the research of environmental performance of lighting. A method for including the environmental impacts of light may be created, similarly with the noise or odor impact calculation methods. The uncertainties could be analyzed and ways to reduce them may be developed. More accurate data for the life cycle inventory of light sources should be made available. The LCA of new products may be conducted, including especially the evaluation of OLED products. The discussion of the functional unit remains valid: it may be directed towards an application-specific functional unit or towards a functional unit that could be used for *all* light sources.

References

- BIOIS (2003) Study on external environmental effects related to the life cycle of products and services. Appendix 2: Case studies. Bio Intelligence Service, European Commission. The report can be found online at: http://ec.europa.eu/environment/ipp/pdf/ext_effects_appendix2.pdf
- Dale AT, Bilec MM, Marriott J, Hartley D, Jurgens C, Zatcoff E (2011) Preliminary comparative life-cycle impacts of streetlight technology. *J Infrastruct Syst* 17(4):193–199
- DEFRA (2009) Life cycle assessment of ultra-efficient lamps. Department for Environment, Food and Rural Affairs, London
- Elijošiūtė E, Balciukevičiūtė J, Denafas G (2012) Life cycle assessment of compact fluorescent and incandescent lamps: comparative analysis. *Environ Res Eng Manage* 61(3):65–72
- ETSI (2011) Environmental Engineering (EE); Life Cycle Assessment (LCA) of ICT equipment, networks and services; General methodology and common requirements. ETSI, Sophia Antipolis
- Gydesen A, Maimann D (1991) Life cycle analyses of integral compact fluorescent lamps versus incandescent lamp. In: *Proceedings Right Light 1*, Stockholm, Sweden, pp. 411–417
- ISO (2006a) 14040: Environmental management – life cycle assessment – principles and framework. International Organization for Standardization
- ISO (2006b) 14044: Environmental management – life cycle assessment – requirements and guidelines. International Organization for Standardization
- Klöpffer W (2003) Life-cycle based methods for sustainable product development. *Int J Life Cycle Assess* 8(2):157–159
- Klöpffer W (2008) Life cycle sustainability assessment of products. *Int J Life Cycle Assess* 13(2):89–95
- Landis AE, Bilec MM, Rajagopalan N (2009) Life cycle assessment for evaluating green products and materials. In: *Proceedings of US-Japan workshop on LCA of sustainable infrastructure materials*, Sapporo
- Michaud R, Belley C (2008) Analyse du cycle de vie comparative d'ampoules électriques: incandescentes et fluorescentes compactes. Centre interuniversitaire de recherche sur le cycle de vie des produits, procédés et services (CIRAIG), Montréal
- Osram (2009) Life cycle assessment of illuminants – a comparison of light bulbs, compact fluorescent lamps and LED lamps. Osram Opto Semiconductors GmbH, Regensburg
- Parsons D (2006) The environmental impact of compact fluorescent lamps and incandescent lamps for Australian conditions. *Environ Eng* 7(2):8–14
- Pfeifer RP (1996) Comparison between filament lamps and compact fluorescent lamps. *Int J Life Cycle Assess* 1:8–16
- Philips (2012) Luxeon rebel datasheets. Philips Lumileds Lighting Company. <http://www.philipslumileds.com/products/luxeon-rebel/luxeon-rebel-white>. Accessed May 2012
- Quirk I (2009) Life-cycle assessment and policy implications of energy efficient lighting technologies. University of California, Berkeley
- Radio-Electronics (2012) High brightness LED, HBLED tutorial. Adrio Communications, Surrey. <http://www.radio-electronics.com/info/data/semicond/leds-light-emitting-diodes/high-brightness-hbled-basics-tutorial.php>. Accessed May 2012
- Ramroth L (2008) Comparison of life-cycle analyses of compact fluorescent and incandescent lamps based on rated life of compact fluorescent lamp. Rocky Mountain Institute
- Slocum A (2005) A technology assessment of Light Emitting Diode (LED) solid state lighting for general illumination. National Center for Environmental Economics (NCEE), Washington
- Swarr TE, Hunkeler D, Klöpffer W, Pesonen H-L, Ciroth A, Brent AC, Pagan R (2011) Environmental life cycle costing: a code of practice. Society of Environmental Toxicology and Chemistry (SETAC), Pensacola
- Tähkämö L (2013) Life cycle assessment of light sources – case studies and review of the analyses. Doctoral dissertation, Aalto University. Unigrafia Oy, Helsinki

- Tähkämö L, Puolakka M, Halonen L, Zissis G (2012a) Comparison of life cycle assessments of LED light sources. *J Light Vis Environ* 36:44–53
- Tähkämö L, Zissis G, Martinsons C (2012b) Methodology study of life cycle assessment of light sources. In: *Proceedings of the 13th international symposium on the science and technology of lighting*, Troy
- U.S. DOE (2012a) Life-cycle assessment of energy and environmental impacts of LED lighting products, part 1: review of the life-cycle energy consumption of incandescent, compact fluorescent, and LED lamps (updated August 2012). United States Department of Energy
- U.S. DOE (2012b) Life-cycle assessment of energy and environmental impacts of LED lighting products, part 2: LED manufacturing and performance. United States Department of Energy
- U.S. DOE (2013) Life-cycle assessment of energy and environmental impacts of LED lighting products, part 3: LED environmental testing. United States Department of Energy
- UNEP (2009) Guidelines for social life cycle assessment of products. United Nations Environment Programme, Druk in de weer, Belgium
- Welz T, Hischer R, Hilty LM (2011) Environmental impacts of lighting technologies – life cycle assessment and sensitivity analysis. *Environ Impact Assess Rev* 31:334–343
- Yabumoto N, Hatta A, Jinno M, Hattori H (2010) Life Cycle Assessment of LED lamps and CFLs as the ideal light source for “sustainable development”. In: *Proceedings of the 12th international symposium on the science and technology of light sources and 3rd white LED conference*, Eindhoven, The Netherlands

Further Reading

- Baumann H, Tillman A-M (2004) *The hitch hiker’s guide to LCA*. Studentlitteratur AB, Lund
- Boyd SB (2012) *Life-cycle assessment of semiconductors*. Springer, New York
- Tähkämö L, Martinsons C, Ravel P, Grannec F, Zissis G (2014) Solid state lighting Annex: Life cycle assessment of solid state lighting, Final report. Energy Efficient End-Use Equipment (4E), International Energy Agency. Available at: http://ssl.iea-4e.org/files/otherfiles/0000/0068/IEA_4E_SSL_Report_on_LCA.pdf

Impact of Lighting on Flora and Fauna

Sibylle Schroer and Franz Hölker

Contents

Introduction	958
Light Perception and Signaling Outside Human Perception	959
Light Intensity	959
Color Spectra	959
Direction of the Light	961
Receptors for Light	962
Responses to Light	963
Responses to Light in Plants	964
Responses to Light in Arthropods	967
Responses to Light in Fish and Amphibians	971
Responses to Light in Birds and Reptiles	973
Responses to Light in Mammals	977
Conclusion	979
Future Research	980
References	981

Abstract

Technology, especially artificial light at night (ALAN), often has unexpected impacts on the environment. This chapter addresses both the perception of light by various organisms and the impact of ALAN on flora and fauna. The responses to ALAN are subdivided into the effects of light intensity, color spectra, and duration and timing of illumination. The ways organisms perceive light can be as variable as the habitats they live in. ALAN often interferes with natural light information. It is rarely neutral and has significant impacts beyond human perception. For example, UV light reflection of generative plant parts or the direction of light is used by many organisms as information for foraging, finding

S. Schroer (✉) • F. Hölker

Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany

e-mail: schroer@igb-berlin.de; hoelker@igb-berlin.de

spawning sites, or communication. Contemporary outdoor lighting often lacks sustainable planning, even though the protection of species, habitat, and human well-being could be improved by adopting simple technical measures. The increasing use of ALAN with high intensities in the blue part of the spectrum, e.g., fluorescent light and LEDs, is discussed as a critical trend. Blue light is a major circadian signal in higher vertebrates and can substantially impact the orientation of organisms such as numerous insect species. A better understanding of how various types and sources of artificial light, and how organisms perceive ALAN, will be an important step towards more sustainable lighting. Such knowledge is the basis for sustainable lighting planning and the development of solutions to protect biodiversity from the effects of outdoor lighting. Maps that describe the rapid changes in ALAN are urgently needed. In addition, measures are required to reduce the increasing use and intensity of ALAN in more remote areas as signaling thresholds in flora and fauna at night are often close to moonlight intensity and far below streetlight levels.

List of Abbreviations

ALAN	Artificial light at night
BAT	Brown adipose tissue lux
UV	Ultraviolet
WAT	White adipose tissue

Introduction

Light allows us to perceive distances, spaces, movements, and rhythms. It is pure energy – light is life. Humans differ from animals in being able to overcome the rhythms of natural light. Artificial light at night (ALAN) became an everyday tool whose absence can no longer be contemplated. It has become an asset with enormous high technology development. Today light is not only used for perceiving visual information but also to deliver data around the globe. New technological developments in the use of light will shape the human economies and production activities significantly in the future. But how much do we understand about this signaling tool we use everyday? Most humans cannot imagine living without ALAN, but at the same time, they do not realize to what extent natural nightscapes are illuminated. Our awareness rises when black-outs accidentally switch off ALAN in urban areas and the natural nightscape becomes visible. The introduction of ALAN has caused an unprecedented disruption to the transformation of nightscapes over large areas of the Earth. There have been no natural analogues, at any timescale, to the nature, extent, distribution, timing, or rate of spread of ALAN (Gaston et al. 2015). The use of ALAN increased by about 6 % annually during the last decades (Hölker et al. 2010a) and ALAN continues to extend further in space, time, and intensity. Its extended use can be perceived far from outside the world, but we are only starting to understand its impact on the biosphere.

This chapter addresses both the perception of light by various organisms and the impact of ALAN on flora and fauna. All life on Earth has evolved to live in cycles of light and dark. For most organisms, their temporally differentiated niche has been promoted by highly developed senses, often including specially adapted eyesight (Hölker et al. 2010a). Light receptors are highly tuned on organisms' habitat and ecological function. A high diversity of solutions evolved to cope with the challenges posed by the different light environments, in order to exploit it most efficiently. The responses to ALAN are subdivided into the effects of light intensity, color spectra, and duration and timing of illumination within this chapter.

Light Perception and Signaling Outside Human Perception

Many organisms see the world in different light than humans do. Light, which is visible to human eyes, represents only a small part of the full spectrum of biological relevant radiation. In addition, most animals have developed distinct sensory mechanisms that perfectly cope with their temporally differentiated niche and activity pattern. For example, human vision has evolved accordingly to surviving needs for diurnal activity and is thus specialized on daylight vision but limited in perceiving low-light visual cues which are important, e.g., for nocturnal and crepuscular organisms or those inhabiting dark caves.

Light Intensity

A moonless night is about 100 million times darker than a day with bright sunshine. Nocturnal organisms, approximately 30 % of all vertebrates and more than 60 % of all invertebrates, evolved highly developed senses to be active during low-light conditions (Hölker et al. 2010b). One of the greatest challenges for low-light vision is to absorb sufficient photons to reliably discriminate color. The eyes of animals living in the world's dimmest habitats are usually adapted to capture and absorb as many photons as possible having, e.g., large apertures and short focal lengths (Warrant 2004) (Fig. 1). The *tapetum lucidum*, a layer of tissue immediately behind the retina in the eye of many crepuscular or nocturnal vertebrates, contributes to the superior night vision (Fig. 2). It reflects light back through the retina, increasing the photon absorbance (Ollivier et al. 2004). The enhanced sensitivity gained from a reflective *tapetum* was an early step in vertebrate evolution to enhance sensitivity during crepuscular periods of changes in light intensity and thereby from cone-based to rod-based vision (Collin et al. 2009).

Color Spectra

In vertebrates, the photoreceptive cells directly sensitive to light are classified as rods and cones. Long rods function mainly in dim light <0.001 cd and provide

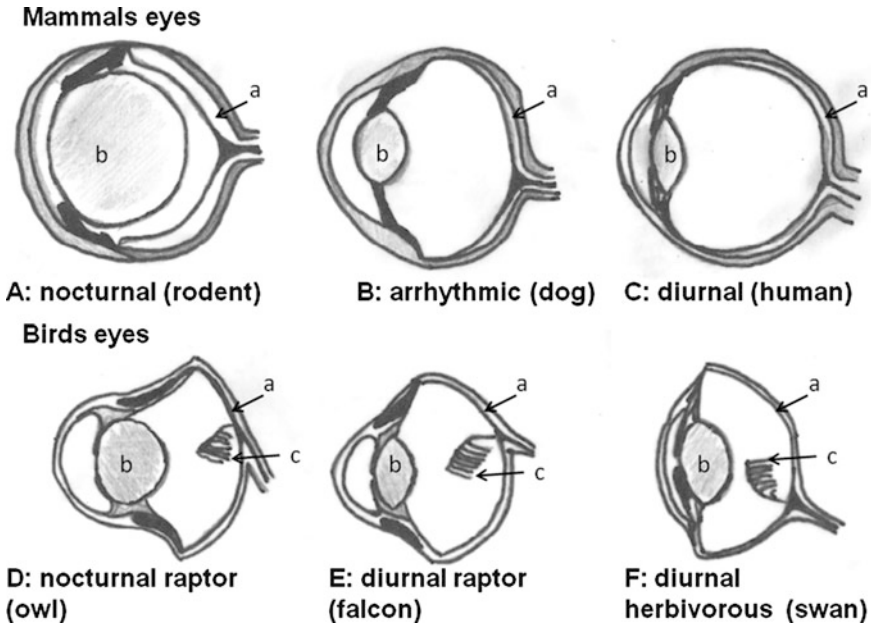


Fig. 1 Differences in eye forms of mammals and birds with different activity patterns. Letters indicate (a) the retina, (b) the lens, and (c) the pecten (Image illustrated by Sibylle Schroer)

color-blind vision in most vertebrates. The shorter cones support vision in light intensities that are stronger than 3 cd and allow color perception. This perception results from the varying spectral sensitivities of different cone cell types to various parts of the spectrum, according to their structure and length (Kelber and Roth 2006). High light intensities at wavelengths outside cone-type sensitivities may cause the same physiological response in the photoreceptor as a dim light at the peak sensitivity level. At brightness levels from 0.001 to 3 cd, both cell types are activated for mesopic vision. Retinal ganglion cells receive the signals from cones and rods and transmit the information via long axons into the brain (Berson et al. 2002). A small percentage of retinal ganglion cells are photosensitive. Some of the retinal neurons form a circadian network, a clock to reconfigure retinal circuits for enhancing light-adapted visual function during the day- and dark-adapted rod-mediated visual function at night (McMahon et al. 2014).

The wavelengths corresponding to visible light for human perception range from about 390 nm to 700 nm. Therefore, ultraviolet (UV) light, which has shorter wavelengths, cannot be perceived by most humans. UV is, for example, necessary for the synthesis of vitamin D, but excessive exposure can cause sunburn and skin cancer. Infrared radiation (IR) can only be perceived as warmth. Conversely to human vision is the perception of the UV spectrum rather important for most birds, insects, amphibians, crustaceans, and some fish and bat species (e.g., Bennett and Cuthill 1994; Mazza et al. 2002; Winter et al. 2003). This sense can be useful for orientation, foraging, and sexual attraction; it plays a critical role for pollinators to

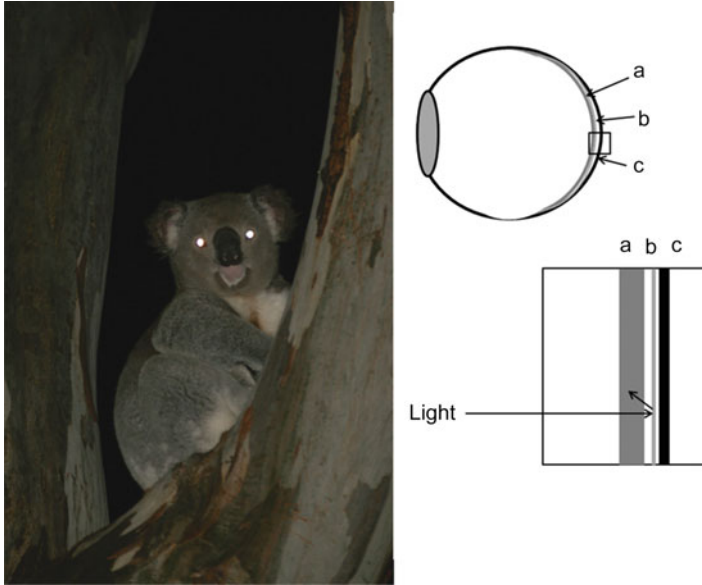


Fig. 2 *Tapetum lucidum* of nocturnal and crepuscular vertebrates. The light gets reflected by the layer behind the retina and thus increases photon absorbance. Small letters indicate (a) the retina, (b) the *tapetum lucidum*, and (c) the choroid (Image illustrated by Schroer. Photo courtesy of Annette Krop-Benesch)

find flowers (Johnson and Andersson 2002; Winter et al. 2003) and for birds finding fruits (Altshuler 2001; Bennett and Cuthill 1994). Interestingly, UV sensitivity is present in both diurnal and nocturnal species, e.g., in nocturnal caudata (Mège et al. 2016) and insects (Eisenbeis 2006). Not only direct UV light but also the reflection of UV light in fruits and its perception by frugivorous animals is a fine-tuned ecological interaction that supports nighttime pollination and fruit dispersal (Altshuler 2001). Hence, the ban of UV emitting outdoor lighting is a necessary and timely step toward species protection.

Direction of the Light

The information of the direction of light is used by many organisms as a visual cue in addition to discriminate brightness and color. Light polarization is invisible to most humans; it refers to the state of the light describing the proportion of waves in a beam that rotate in parallel planes to each other, or at the same rotation direction, resulting in fully, partially, or unpolarized light signals (Horváth 2014; Sabbah et al. 2005).

In nature light tends to be partially linearly polarized and thus delivers information for orientation and navigation. Dung beetles (*Scarabaeus zambesianus*) orientate on polarized moonlight to escape in a straight line from the dung pile avoiding competition (Dacke et al. 2003). Linear polarization of clear skylight is always perpendicular to the

sun–antisun direction, and so it is used as a sun compass for navigation, as revealed for the desert ant (*Cataglyphis*) (Wehner and Müller 2006). This sense is also used by underwater organisms, at clear water conditions and especially at lower sun elevation, e.g., during their daily vertical migration at sunset and at sunrise (Lerner et al. 2011). Crustaceans use polarization for navigation and orientation toward food. Many of their prey species are either reflective or transparent, but they change the polarization of the light (Kleinlogel and White 2008). Cuttlefish use polarization vision to improve detection of predators in turbid conditions (Cartron et al. 2013). Water-emerging insects find their habitat mainly on horizontally polarized light reflected from the water surface (Horváth and Csabai 2014). Nonbiting midges (Chironomidae) orient on reflected polarization of the water surface to oviposit at sites with suitable food availability for their larvae (Lerner et al. 2008).

Artificial polarization is a risk for populations that depend on this signal. For example, aquatic insects attracted to highly horizontally polarizing light, such as asphalt under streetlighting, may mistakenly land or oviposit on the dry surfaces causing the fatal consequence of desiccation (Horváth et al. 2009). The lunar skylight polarization signal can be polluted by urban lighting, which indicates that nocturnal animal navigation systems, which depend on perceiving polarized moonlight, likely fail to operate properly in highly light-polluted areas (Kyba et al. 2011). Consequently, future light management should take polarization into account for species protection.

Receptors for Light

Opsins are responsible to mediating the conversion of a photon of light into an electrochemical signal. They are essential joints of communication between the internal and external environment of photoreceptor cells in prokaryotes, some algae, and animals. In plants, fungi and Placozoa, however, no opsins have yet been recorded. The great variety of opsins is classified according to the function and spectral absorption maximum. Photopsin is a superior term for photoreceptor proteins found in the cone cells of the retina that are the basis of color vision. Rhodopsin is the primary pigment found in rod photoreceptors. It is enabling vision in very low-light conditions without color discrimination (Litman and Mitchell 1996).

Cryptochromes are blue light-sensing photoreceptors found throughout the kingdom of life, in bacteria as well as in plants, animals, and humans. These flavoproteins are involved in the circadian rhythm and development of plants and animals and in some species in the sensing of magnetic fields (Lohman 2010). They absorb light associated with short, blue light wavelength spectra, not exceeding 500 nm (Ahmad and Cashmore 1996), and activate photocatalytic processes. These photobiological processes are considered as an evolutionary composition, which evolved approximately 3.5 billion years ago (Kritsky 1984).

The best-known photoreceptors of plants, fungi, and bacteria are phytochromes, a family of chromoproteins involved in circadian rhythms, growth, and development (Auldridge and Forest 2011; Chen et al. 2004; Giraud and Verméglio 2008;

Rodriguez-Romero et al. 2010). Phytochromes are mainly sensitive to red and far red light (660 and 730 nm, respectively). In microbial systems, phytochrome sensitivities extend in yellow, green, blue, and violet portions of the spectrum (Rockwell and Lagarias 2010).

Carotenoid pigments are responsible for the reception of photoperiodic responses in invertebrates, such as induction of the seasonal metabolism break or migration (Veerman and Veenendaal 2003; Veerman 2001).

Another recently detected opsin, the melanopsin is involved in circadian rhythm signaling processes and pupillary reflex in retinal ganglion cells in the eyes of humans and other vertebrates (Berson et al. 2002; Hankins et al. 2008; Hattar et al. 2002; Provencio et al. 2000; Zaidi et al. 2007). The circadian rhythm signaling is mediated over the suprachiasmatic nucleus, which regulates body functions associated with the 24-h rhythm. It triggers the pineal gland, also called the “third eye,” a small endocrine gland in the epithalamus of the vertebrate brain. It is shaped like a pine cone and lies between the two halves of the thalamus. It produces melatonin, a serotonin-derived hormone, which affects the modulation of recreational patterns in both seasonal and circadian rhythms (Pévet et al. 2006). While in mammals, the pineal gland purely serves as a neuroendocrine organ, it is photoreceptive in nonmammalian vertebrates such as fish and amphibians (Ekström and Meissl 2003).

Image-forming eyes are a marvel of evolution to identify suitable prey and detect potential predators (Collin et al. 2009). The predominant eye type in vertebrates is classed as “camera” eye, because light is penetrating through a single opening and projected onto the retina (Fig. 3), a layer of tissue, lining the inner surface of the eye. Camera eyes typically adjust to differences in light intensity by reducing the size of the pupil when exposed to bright light reducing the retinal irradiance. This pupil dilation reflex allows higher visual acuity at daytime and adaptation of light penetration onto the retina and protects the retinal photoreceptors from damaging by extreme light intensities (Gerkema et al. 2013).

Predominant photoreceptors in insects are compound eyes, multiple lenses, up to tens of thousands, each focusing light onto a small number of retinula cells. Next to the compound eyes for visual responses, most insects have further photoreceptors, referred to the ocelli (Fig. 4). The dorsal ocelli are found in most insect species with various number, forms, and functions. The lateral ocelli, or stemmata, are found in holometabolous larvae and certain adults of several insect orders (Matthews and Matthews 2009).

Responses to Light

Light is not neutral; it is the main source of energy for most primary producers and a major factor in controlling many physiological and behavioral processes. It is an important signal for growth, spatial movement, orientation, and communication, triggering community structure and energy flow through the food web. Responses to light are multifaceted and highly species depended. ALAN used inappropriately threatens biodiversity by impacting trophic, social, and competitive interactions,

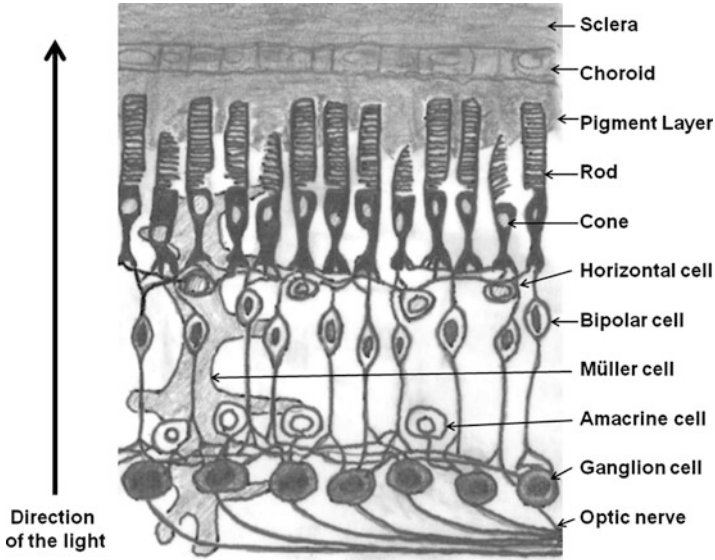


Fig. 3 The retina is a layer of light-sensitive nerves and receptors, coating the inner eye of vertebrates (Image illustrated by Sibylle Schroer)

Fig. 4 Compound eyes and ocelli of a hornet (*Vespa crabro*). Photo courtesy of Tim Haye. Letters indicate *a*: the ocelli and *b*: the compound eye



masking seasonal and daily rhythms and thus altering community structure, ecosystem processes, and properties (Longcore and Rich 2004; Hölker et al. 2010b; Gaston et al. 2013, 2015; Kurvers and Hölker 2014).

Responses to Light in Plants

Plants react to their light environment with stomata opening and chloroplast movement and alter their growth in response (Briggs and Christie 2002; Keller et al. 2011).

Signals perceived by the phytochrome system adjust growth according to the daily and seasonal needs, e.g., photoperiodic induction of flowering, leaf senescence, and abscission. It also regulates other responses including the germination of seeds; the elongation of seedlings; the size, shape, and number of leaves; or the synthesis of chlorophyll. Plants use the phytochrome system to grow away from shade and toward light and to compare the length of dark periods (Keller et al. 2011). Physiological responses are triggered by light intensity, wavelength spectra, and duration.

Light Intensity

Low-light conditions inhibit plant growth and development by affecting gas exchange, whereas excess light intensity has detrimental effects on the photosynthetic apparatus. As a result, plants have developed sophisticated mechanisms to adapt their structure and physiology to the prevailing light environment.

Already in the 1930s, Matzke (1936) discovered that light levels of only 10 lx, or less, supplied by a bulb in 13 m distant, may delay leaf fall by a month beyond the normal season. To induce photoperiodic reactions in plants, light intensities can be as low as 0.1 lx, when the light encounters the spectral sensitivity maximum. Reaction saturation in photoperiodic time measurement is reached in most species at light level differences of 10 or maximal 100 lx (Bünning and Moser 1969). With intensities beyond saturation, photoperiodic reactions no longer depend on the light intensity. The signal of light for leaf abscission is overruled by temperature at a threshold of 12 °C (Cathey and Campbell 1975). The effect of ALAN on trees is highly species specific. Sycamores (Platanaceae) and elms (Ulmaceae: *Ulmus* sp.) are, for example, very sensitive to ALAN and keep continuous growth for a longer period in the fall when treated with 10 lx light intensity during their first years (Cathey and Campbell 1975) (Fig. 5). Streetlight was also found to alter the vegetative growth of crops. Sinnadurai (1981) found significant higher size of maize plants, higher number of leaves, and number of cobs in plants close to high-pressure sodium lights than those 60 m away from the light source. Bennie et al. (2015) described an impact of ALAN on flower head density in a leguminous food plant at approximately 10 lx at the surface.

For the optimal development of, e.g., young tomato plants, a flux density of about 300 $\mu\text{mol m}^{-2} \text{s}^{-1}$ photosynthetic photon, comparable to a sunny day, is recommended (Fan et al. 2013). At diffused light conditions, the need in photosynthetic photons for optimal development is decreased: This is probably due to light penetration into the lower layers of the leaf canopy, resulting in an increased CO₂ fixation rate by the whole canopy (Tani et al. 2014). Diffused light condition can be achieved in greenhouses by foil and glass to save energy costs. Diffused ALAN by the atmospheric scattering, the so-called skyglow, might consequently also have a higher impact on plant responses, than direct radiation. Skyglow can lead to nighttime brightness up to thousandfold above that experienced by organisms during their evolutionary history (Kyba et al. 2015a). Although the intensity of skyglow is small compared to direct streetlight, it extends over vastly larger areas (Kyba and Hölker 2013). Surprisingly, the impact of skyglow on plants is today neither considered nor studied.

Fig. 5 Defoliated *Betula pendula* except in the light cone of high-pressure streetlight (Image taken in Berlin, Germany, December 2015, by Sibylle Schroer)



Color Spectra

Responses in plants to wavelength are numerous. Certain spectra regulate photosynthesis, morphology, phototropism, volatile organic compound emission, and synthesis of secondary metabolites, leaf thickness, or quantity of cuticle wax (Vänninen et al. 2010).

The UV-B (280–320 nm) and the UV-A light spectrum (320–390 nm) are especially critical for plants. Low doses induce specific photomorphogenetic and developmental responses, whereas high doses result in stress signal transduction, triggering protection from damage to the photosynthetic apparatus in excess of light (Johanson et al. 1995).

The responses to the UV-A light spectrum contribute to maximizing photosynthetic potential and activated respiration in acceptable light intensities (Tsuboi and Wada 2011). The blue wavelength spectrum from 390 to 500 nm is catalytic for several metabolic and growth processes in plants, e.g., phototropism, chloroplast relocation, stem elongation, photoperiod-dependent flowering induction, resetting of the circadian oscillator, and control of stomatal opening. Strong blue light activates the incorporation of carbon in amino acids, leading to a lower amount of starch formation in leaf chloroplasts, and increases the biosynthesis of proteins (Vänninen et al. 2010). Blue light upregulates genes that encode key enzymes in the Calvin cycle, whereas green to red light spectra (500–600 nm) downregulate these genes. The phytochrome responds to red light with initiation of cell growth.

Especially sodium streetlighting radiating a big ratio at 589 nm, the so-called sodium line, can stimulate phytochrome responses and plant growth (Cathey and Campbell 1975).

Not only single narrow bandwidths of spectra stimulate plant responses but rather the ratio of different spectra. The relative amount of blue to red, e.g., triggers photosynthetic activity, or the relation of red to far red is critical for seed germination, seedling establishment, shade-avoidance response, and floral induction. The green spectrum of the light (500–580 nm) tends to temper or negate the effects of blue and red light in plants (Vänninen et al. 2010).

Timing and Duration of Illumination

Light duration can induce flowering, nutrient uptake, volatile organic compound emission, and synthesis of secondary metabolites (Vänninen et al. 2010). The circadian rhythm is triggered by light and temperature. It is intensely studied with *Arabidopsis* mutant plants, which changed the rhythm accordingly to altered associated genes. Interestingly, the photoreceptor expression (phytochrome and cryptochrome) is itself rhythmic, indicating that the clock gates its sensitivity to light (McClung 2006). *Arabidopsis* clock mutants with longer periods (28 h) accumulated lower biomass than those with short periods (20 h) when grown under short cycles (10 h light/10 h dark), indicating impaired physiological function, including lower rates of chlorophyll production and carbon fixation (McClung 2006).

Plants with light-dependent flower induction have critical associated daylengths. Long-day plants induce flower buds when the days are longer than their critical daylength. In northern latitudes, these plants flower in summer. Short-day plants induce flowering when the days are shorter than their critical daylength. They flower in spring or fall in northern hemisphere. Day-neutral plants form flowers independent of the daylength. The perception of daylength in long-day *Arabidopsis* adjust to the phase angle of circadian rhythms relative to the light–dark cycle, rather than measure the absolute duration of light and darkness (Roden et al. 2002).

The photoperiod is a critical determinant of the oxidative stress response. Queval et al. (2007) have shown links between daylength and the rate of oxidative cell death. Defense genes and oxidative stress-responsive transcripts are induced to a greater extent in short days than in long days. Dim nocturnal light can in some species inhibit recovery from leaf damage caused by atmospheric ozone, e.g., in subterranean clover (*Trifolium subterraneum*) (Vollsnes et al. 2009). Since the patterns of anthropogenic light pollution and ozone pollution are spatially correlated on a global scale (Cinzano et al. 2001; Ashmore 2005), Gaston et al. (2013) demand a closer look on the extent to which low-intensity nighttime light could affect repair and recovery from ozone damage.

Responses to Light in Arthropods

“There is more mechanistic evidence for caterpillar-booms than for baby booms following power outage” (van Geffen 2015).

Moths are not only attracted to ALAN sources in great number; female moths of the orders Geometridae (inchworms) are also less active and emit less sex pheromones under ALAN conditions, resulting in reduced mating success (van Geffen et al. 2015).

Arthropods use visual cues to orient, navigate, and avoid predators or locate host plants, prey, and mates (Prokopy and Owens 1983). As many insects are attracted to light, ALAN functions like a vacuum cleaner. It is able to suck them out of their natural habitat (Eisenbeis 2006). Several theories try to explain the diversity of behaviors of insects around artificial sources of light. One reason is their navigation behavior. The optomotor system stabilizes the course by retinal images of the sky. The retinal image does not change as long as the animal moves along a straight line, when rotating the signal changes. The disturbance resulting in involuntary rotation can be corrected by compensatory body movements. The animal will rotate until it has reestablished its former retinal image (Wehner 1984; Frank 2006). Furthermore, an insect flying from artificial light into darkness or from darkness into light may be functionally blind until eye pigments have returned to their dark-adapted positions. Similar mechanisms might also be responsible for affecting both abundance and composition in aquatic arthropods such as amphipods (Navarro-Barranco and Hughes 2015).

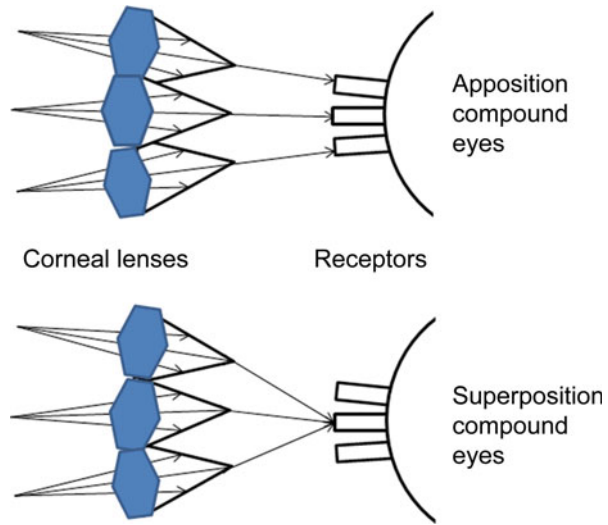
The massive attraction of disoriented prey to ALAN sources results in predator accumulation. The bridge spider (*Larinioides sclopetarius*), for example, rapidly increases its reproductive activity by the periodically excessive supply of emerging water insects (Kleinteich and Schneider 2011) and is attracted to artificial lights in manmade habitats (Heiling 1999). However, the ability to benefit from prey attraction to artificial light differs between nocturnal spiders. Most orb spider species are rather sensitive to light, and some even require absolute darkness for web building, e. g., the Walnut orb-weaver spider (*Nuctenea umbratica*) or the silver-sided sector spider (*Zygiella x-notata*) (Zschokke and Herberstein 2005).

Light Intensity

Habitats of arthropods differ greatly and thus their visual senses. Nocturnal insects have highly sensitive visual systems. Some display superposition compound eyes (Fig. 6) that are 100–1000 times more sensitive compared to the ones of diurnal insects of the same size (Hagen et al. 2015). In some nocturnal species, like nocturnal hawk moths (Sphingidae), color perception is possible at very low light intensities, as low as $0.0001 \cdot \text{cd m}^{-2}$, enabling nighttime pollination with color vision comparable to those for diurnal pollinators (Kelber and Roth 2006).

Already at light levels around moonlight (0.3 lx at maximum), many arthropods are affected. For some the threshold lies even beyond. Such light levels are already reached by urban skyglow (Kyba et al. 2015). The occurrence of species communicating with bioluminescence, e.g., fireflies (Lampyridae), rapidly decreases when a threshold of 0.5 lx is exceeded (Hagen et al. 2015). Aquatic insects in stream systems drift at nocturnal light levels below 0.001 lx to avoid predation by fish (Bishop 1969). And for water fleas (*Daphnia* sp.), a keystone herbivore in lakes, it has been described that even skyglow can suppress both the amplitude and the magnitude of

Fig. 6 Simplified image of apposition – in contrast to superposition eyes. Corneal lenses of apposition eyes each form a small inverted image, whereas in superposition eyes the lenses form a single erect image on the receptors of the retina



the diel vertical migration. The reduced nocturnal upward movement into upper water layers may release phytoplankton from grazing pressure and thus may influence water quality in urban waters (Moore et al. 2000).

Both skyglow and direct light might play a role for the onset of long-distance flights of moths. For example, Riley et al. (1983) describe that the onset of “dusk flights” occurred when the irradiance level fell on average to $2.7 \times 10^{-5} \text{ Wm}^{-2} \text{ nm}^{-1}$ in the 450–800 nm range (about 2 lx).

Color Spectra

The spectral sensitivity in arthropods has as many facets as their habitats and ecological functions. Most insects have three different receptor types, one being sensitive to UV light with a peak sensitivity around 350 nm, another to blue light at 440 nm, and a third to green light around 540 nm (Kelber et al. 2003). In contrast, cockroaches and ants possess two photoreceptors, water fleas four, and flies and some butterflies five, or stomatopod crustaceans have evolved even twelve spectral photoreceptor types (Kelber and Roth 2006). Arthropods’ responses to color are influenced by the dominant wavelength and by the composition of wavelengths.

The broader the spectrum of the ALAN light source, the wider is the range of arthropod species that will respond to the light (Davies et al. 2013). The streetlight type attracting the most insects is the white and UV light-emitting mercury vapor light (Eisenbeis 2006). It draws seven times as many insects as comparable LED light (van Grunsven et al. 2014). The attraction of moth species to fluorescent lamps with various filters for wavelengths and radiance in the UV spectrum correlates negatively with the wavelength and positively with morphological characteristics of moth species, especially with their eye size (van Langevelde et al. 2011). LEDs are considered by some authors as relatively unattractive for insects compared to long wavelength emitting low-pressure sodium lamps (e.g., Eisenbeis and Eick 2011).

In contrast, a study by Pawson and Bader (2014) demonstrates greater attraction to LEDs than to high-pressure sodium lamps irrespective of the color temperature of the LEDs. The diverse attraction to LED light is today not understood. It is most possibly related to the different radiation angle and species affinity to short wavelengths, which are emitted to a great part by LEDs.

UV light was observed to be attractive to winged aphid specimen, triggering the rising from the leaf to the sky and hence migratory behavior; instead green light attraction is related to vegetative behavior for host selection mechanism (Döring 2014). The attraction to green light was found to be altered in plant virus-infected whiteflies potentially leading to advantages in spreading the virus to other plants (Jahan et al. 2014).

The shorter wavelengths of the UV-B spectrum have adverse effects on the survival, development, egg hatchability, and fecundity. This part of the spectrum seems to be crucial for the flight activity, takeoff, and initiation of dispersal (Johansen et al. 2011). The tolerance of UV radiation is related to species and their associated habitat, and for some species living in dark habitat, e.g., fruit flies (*Drosophila melanogaster*), longer wavelength of the visible blue light spectrum is rather detrimental than the UV light. Hori et al. (2014) therefore consider the potential of high-intensity monochromatic blue LED radiation for species-specific economic pest control.

In future knowledge about various effects of the spectral power distribution of artificial light might be more exploited to trigger favorable processes for plant production and pest reduction in greenhouses. For example, the illumination with low-intensity violet-blue light can prevent diapausing in a beneficial insect predator (*Orius insidiosus*), while allowing optimal flowering in the short photoperiod host plant (*Dendranthema grandiflora*) of its prey (Stack and Drummond 1997). The use of LED lighting in greenhouses is of high interest to plant growers as the narrow spectrum light emission from LEDs can be matched to plant responses without wasting energy on nonproductive wavelengths (Massa et al. 2008; Stutte 2009). However, the physiological interpretation or evolutionary patterns of wavelength sensitivity in arthropods today still bear many open questions. It is, for example, not yet understood, why two phylogenetically closely related orders Trichoptera and Lepidoptera show different sensitivity patterns (van Grunsven et al. 2014).

Timing and Duration of Illumination

In arthropods like in plants, the photoperiod often acts in concert with other physical or biological background information, like temperature, moisture, humidity and cues from food, and conspecific density (Danks 2005). The responses are often triggered by hormones such as melatonin, which was found to have different circadian concentrations in various organs, possibly related to differences in melatonin function. In the hemolymph of Australian black field crickets, *Teleogryllus commodus*, constant illumination was linked to a low circulating melatonin concentration and a negative impact on immune parameters (Durrant et al. 2015). Female crickets (*Gryllus bimaculatus*) held at a 12/12 light–dark cycle expose significantly higher melatonin levels in the compound eye, brain, and palp during the dark period.

Conversely, melatonin levels were significantly higher during the light period in the cercus, ovipositor, antenna, hind leg, and ovary (Itoh et al. 1995). Daily rhythms are found for many activities related to reproduction, such as the onset and length of the calling time by females, the response of males to the call, pheromone production, and release and mating activity (Saunders 2002).

The critical photoperiod for the induction or termination of diapauses or eliciting migration is species specific and often dependent on its geographical origin (Veerman 2001). Animals living in the far north are thought to have a longer critical photoperiod than animals at the equator. Van Geffen et al. (2014) found the pupation duration of the noctuid moth, *Mamestra brassicae*, reduced by artificial light, indicating that diapause can be inhibited by ALAN, when temperature is clement.

Responses to Light in Fish and Amphibians

Aquatic systems and their adjacent terrestrial ecosystems are very likely to be affected by changed natural light due to ALAN, because humans tend to live close to waters (Kummu et al. 2011; Perkin et al. 2011). Especially the illumination of running waters in urban environments is sometimes disproportionately higher than other natural areas such as forests, lakes, meadows, and pastures (e.g., Kuechly et al. 2012). Given the expected prevalence of light along aquatic systems, it may affect typical aquatic vertebrates such as fish and amphibians. Most amphibian species are nocturnal and many are endangered by anthropogenic stress (Hölker et al. 2010b; Stuart et al. 2004). Baker and Richardson (2006) demonstrate that green frogs (*Rana clamitans melanota*) produce fewer advertisement calls and move more frequently under streetlighting than under ambient light conditions. The study clearly shows that male frog behavior is affected by the presence of ALAN in a manner that has the potential to affect population dynamics.

Also fish species are affected by ALAN. Artificial light is used worldwide in fisheries to attract fish, squid, and other marine food (Davies et al. 2014). Several fish species behavior including the hatching process is light dependent (McAlary and McFarland 1993). Thus, ALAN has a great potential to interfere with the circadian behavior of aquatic organisms (Perkin et al. 2011; Davies et al. 2014).

Light Intensity

The physiological and behavioral responses of many fish can be triggered at very low light levels, below 1 lx, which may make even the light levels produced by skyglow an ecologically relevant light source. Such low light levels have been shown to be an important cue for both predator avoidance and feeding, for example, in salmoniformes. ALAN at 1 lx can delay the dispersal timing of Atlantic salmon (*Salmo salar*) fry by up to 2 days and extend the dispersal period into daylight hours (Riley et al. 2015). The period between fry dispersal and the establishment of feeding territories is of critical importance in the dynamics of salmonid populations, and any disruption may significantly increase predation and reduce fitness. In caged Atlantic salmon (*Salmo salar*), the swimming depth could be adjusted by artificial lighting

(Juell and Fosseidengen 2004). Salmon groups swam deeper and at lower density, both day and night, in cages illuminated by lamps at different depths. When underwater lamps were lowered from 1 to 15 m and subsequently raised again during a period of 48 h, swimming depth correlated with lamp depth.

Brüning et al. (2015) showed nocturnal suppression of melatonin and suppressed gene expression of gonadotropins (Brüning et al. 2016) in European perch (*Perca fluviatilis*) at 1 lx. However, cortisol levels were not affected by the tested light intensities, indicating rather hormonal response than a stressful disturbance by the light (Brüning et al. 2015).

Streetlight and especially illumination of bridges might potentially interfere with migratory behavior of certain fish species, resulting in excessive energy loss and spatial impediments to migration, which in turn can result in reduced migratory success (Hölker et al. 2010b). For example, flume experiments demonstrate a strong avoidance reaction of silver eels (*Anguilla anguilla*) to ALAN (Hadderingh et al. 1999), and streetlighting delays and disrupts the dispersal of Atlantic salmon (*Salmo salar*) fry (Riley et al. 2015).

Anuran activity in natural habitats under nocturnal conditions has been found down to 0.00001 lx (Buchanan 2006). Slight increases in illumination caused by nearby lights or even by skyglow may alter foraging behavior or antipredator response of frogs (Buchanan 2006; Baker and Richardson 2006). Red-backed salamanders (*Plethodon cinereus*) orient toward prey sooner at higher ambient illuminations (0.001 lx) compared to lower light levels indicating improved visually based foraging ability at higher light levels (Perry et al. 2008). However, fewer salamanders were active in lighted transects (0.01 lx) compared to unlighted transects.

Color Spectra

Frogs typically exhibit phototaxis and move toward blue light (less than 500 nm) at intensities higher than ambient illumination (Buchanan 2006). However, strong interaction effects between intensity and wavelength preference are observed, making it difficult to predict the behavior of frogs. Thus, more field studies of intensity-dependent and spectrally dependent phototaxis are needed. Most frogs studied have trichromatic color vision and possibly tetrachromatic color vision with sensitivity in the ultraviolet wavelengths (Buchanan 2006).

The natural light spectra in water depend on the composition of the water. In seawater only mid-wavelengths such as blue-green (450–550 nm) are reaching deeper parts, whereas in shallow waters a broader spectrum with short (UV) or long (red) wavelength can be present (Myrberg and Fuiman 2002). In most lakes, however, yellow light penetrates the water the deepest (Lythgoe 1988). Shallow water organisms have broad spectral sensitivities. With depth the sensitivity of water organisms alters toward middle wavelengths (Cronin et al. 2010). Animals that live in different water depths may perceive color throughout almost the entire spectral range of available light. In species, which change their habitat depending on their size, the vision perception can change with their developmental stage (Boeuf and Le Bail 1999). Findings in fishes of the upper part of the water column indicate great sensitivity of the pineal or other nonvisual photoreceptors to blue light (e.g., Vera et al. 2010). Brüning et al. (2016)

found that all colors (blue, green, and red) suppressed melatonin production in perch. They describe blue light as less effective for melatonin suppression, which corresponds to the different light conditions in perch habitats.

UV and red fluorescent light serves species recognition and short distance communication in coral reef inhabiting fish with the benefit that many predatory fish are unable to detect the wavelengths (Gerlach et al. 2014). Red fluorescence is particularly well suited for short-range visual interactions in reef fish as its information content is rapidly lost at greater distances and thus not detectable for most predators. For example, the fluorescent color pattern of *Cirrhilabrus solorensis* peaks at around 660 nm, a visual range for which most reef fish inhabiting depths below 10 m have poor or no sensitivity.

Timing and Duration of Illumination

Fish species are either more active in light, less active in darkness, or vice versa. The photoperiod is important in synchronizing locomotor activity rhythms and food intake (Boeuf and Le Bail 1999). Photoperiod affects reproduction in fish. It is considered the most important environmental cue for breeding seasonality and maturation. In aquaculture, the manipulation of photoperiod is an efficient tool to induce reproductive events. It is used to induce, e.g., reproductive maturation and egg ovulation outside of natural spawning periods (Kolkovski and Dabrowski 1998; MacQuarrie et al. 1979). In natural environments, ALAN might cause similar effects or problems such as disturbance of synchronous hatching and swim bladder inflation of fish larvae, reducing survival chances (Riley et al. 2015; Brüning et al. 2011).

For amphibians the daylength is one of the most important factors to predict seasonal changes in their environment. The trigger is required to anticipate the future changes in their thermal environment for bio-thermal adaptation (Sanabria and Quiroga 2011) and behavioral synchronization (Canavero and Arim 2009). For many fish and amphibians, rapid changes in light levels at night (e.g., by vehicle headlights) can lead to a massive loss of visual information. The periods of light and dark adaptation in juvenile pacific salmon and several nocturnal frogs can be more than 1 h (Buchanan 2006; Nightingale et al. 2006). During these transition periods, organisms are temporarily “blind” for their visual environment.

Responses to Light in Birds and Reptiles

Among birds, there are many species record holders in visual perception. The highest sensitivity for light in birds is measured in the oilbird (*Steatornis caripensis*). The cave-inhabiting bird breeds in depth, where daylight often is excluded and the bird forages for fruit only at nighttime (Fig. 7). With 1 million rods per mm², the oilbird exceeds any counted rod density in vertebrates. The brown falcon (*Falco berigora*), a diurnal fast flying raptor, on the other hand, holds the record of cone density in vertebrate eyes with 380.000 cones per mm² (Martin et al. 2004). The pectens in bird eyes reduce the number of blood vessels in the retina, leading to sharpened vision (Fig.1).

Fig. 7 The eye of an oilbird (*Steatornis caripensis*), which is nesting in caves and a nocturnal feeder on fruits. For foraging the oilbird has specially adapted eyesight and uses echolocation for navigation (Photo courtesy of the Asa Wright Nature Centre)



Birds and reptiles perceive light even in the absence of input from the eyes or associated neurotransmitters. Recently, Fulgione et al. (2014) found opsins at the skin of the belly of the moorish gecko (*Tarentola mauritanica*), which perceive ambient color and initiate camouflage color changing of the skin without the input via the eye.

Magnetoreception is another necessary tool for birds and reptiles to navigate. The signaling is a complex system, involving interactions between magnetoreceptors and visual cues. Artificial light might interfere with these highly species-specific developed senses which are not yet fully understood (Wiltschko et al. 2009).

Light Intensity

The extension of the individual daylength in birds due to ALAN is today well known. Male birds probably start singing at dawn when stimulation by increasing light intensities has reached a certain threshold level. This threshold varies among species, leading to species-specific timing of dawn song. Along an urban gradient ranging from an urban forest to the city center, the onset of blackbird (*Turdus merula*) dawn song differs up to 5 h (Nordt and Klenke 2013). The period of onset of singing before dawn increases with light intensity by 1.5–2 min per lux (Da Silva et al. 2014). Streetlighting affects the timing of dawn song, the strongest in species that under natural conditions start singing early, e.g., the blackbird or robin (*Erithacus rubecula*), and is neglectable in species that naturally initiate singing late, e.g., the chaffinch (*Fringilla coelebs*) (Kempnaers et al. 2010). The rooster crowing can also be induced with light stimulation using intensities of 1 lx for 30 min at dawn, and the number of crows can be increased with lighting intensity (Shimmura and Yoshimura 2013). In an experiment, artificial lighting at 1.6 lx caused great tits (*Parus major*) to wake up earlier, sleep less (–5 %), and spend less time in the nest box as they left their nest box earlier in the morning. Females spent a greater proportion of the night awake (Raap et al. 2015). De Jong et al. (2016) found that the increased activity is not limited to a certain threshold but occurs even when nocturnal light levels are slightly increased.

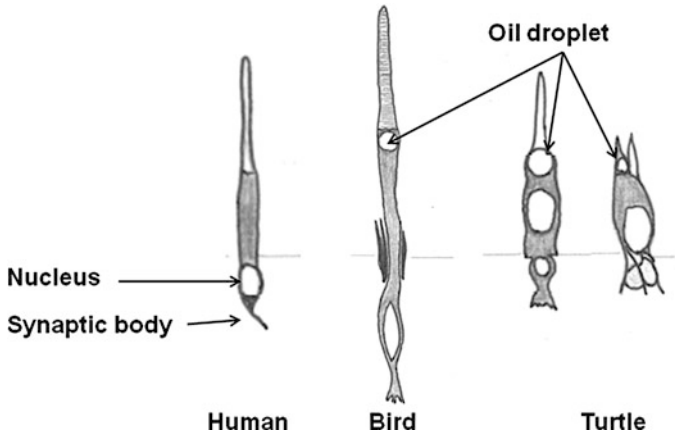


Fig. 8 Oil droplets of bird and reptile cone in comparison to human cone (Illustrated by Sibylle Schroer)

Being ectothermic, reptiles cannot tolerate cold climates in most cases and thus are mainly adapted to diurnal conditions and bright sunlight. Some clades associated with warm climate, such as geckos, show strong tendencies toward nocturnal activity (Perry and Fisher 2006). Sea turtles may be disorientated even by skyglow to the extent that their migration habits are affected (Salmon 2006). They orient on the reflection of starlit water surface to rapidly enter their habitat after hatching at beaches.

Color Spectra

The cones of most reptile and bird species contain brightly colored oil droplets that filter and transmit light to the visual pigment (Fig. 8). Oil droplets decrease cone quantum catch and reduce the overlap in sensitivity between spectrally adjacent cones. The benefit of oil droplets is an increase in the number of object colors that can be discriminated (Vorobyev 2003). The visual color discrimination is species dependent and crucial for communication and social interactions. Of their four specific wavelength maxima, one is sensitive in the UV spectrum, two in the mid-wavelength spectrum, and one to extra-long wavelengths (Bennett and Cuthill, 1994). Parent–offspring chromatic signaling is especially important for diurnal birds, but coloring of nocturnal little owl bills (*Athene noctua*) reveals as well information on owl body mass to adjust parental feeding strategies (Avilés and Parejo 2013). Individual host–prey signaling provides nutritional information. Fruit color communicates lipid content to ensure saturated nutrition supply and increase seed dispersal (Schaefer et al. 2014). Prey communicates further visual warning signals for being inedible or poisonous or disguises using camouflage color.

In the common lizard (*Zootoca vivipara*), Martin et al. (2015) found extended spectral sensitivity in the UV spectrum and the near infrared, which is a rare feature in terrestrial vertebrates. It enables discrimination of small differences in throat coloration and might have coevolved with the visual communication system.

Migrating bird species get attracted by the visible red light, the spectrum between 680 and 550 nm, or UV light. Poot et al. (2008), therefore, discuss the reduction of fatal injuries of migrating birds associated with ALAN by using light of the green and blue spectrum. It offers the highest sensitivity for human night vision but is outside the cone sensitivity of birds. However, Ogden (1996) debates that birds perceive color throughout the spectrum and that rather intensity than color matters. Ouyang et al. (2015) found that adult great tits nesting in white-light transects had higher corticosterone concentrations than individuals nesting in green light or dark control. Individuals in red light had higher corticosterone levels when they nested closer to the lamppost than individuals nesting farther away, a decline not observed in the green or dark treatment. This is fitness relevant because individuals with high corticosterone levels had fewer fledglings.

Timing and Duration of Illumination

The simplest solution to reduce fatal accidents with nighttime migratory bird is to turn off light, especially after midnight, when birds begin to descend from their peak migration altitude. The peak migratory periods are in spring and autumn (Krijgsveld et al. 2015); it seems to be a reasonable timing to turn off hazardous lighting for migratory birds after midnight. Where it is not possible to turn off the light, shielding can direct light downward. This is also an effective measure for many other species, e. g., for marine turtle protection to reduce disruption of visual cues for orientation at the nesting beach.

In urban agglomeration, ALAN interferes with the circadian rhythm of birds and suppresses melatonin even at low light levels (de Jong et al. 2016). Earlier dawn singing in the season is known for blackbirds (*Turdus merula*), robins (*Erithacus rubecula*), and great tits (*Parus major*), whereas blue tits (*Parus caeruleus*) start their dusk singing earlier (Da Silva et al. 2014). The song thrush (*Turdus philomelos*) males, conversely, are more likely to sing at dawn earlier in the season in naturally dark sites, compared to light-polluted sites. Short winter days may be perceived by the birds as longer spring days and cause males to sing earlier in the season provided weather conditions are clement. Birds living in the city center are exposed to a subjective daylength on average 50 min longer than conspecifics breeding in a nearby rural forest (Dominoni and Partecke 2015). The difference in subjective daylength between urban and rural birds is stronger in the beginning of the breeding season, being over 1 h in March, whereas in May the difference is reduced to 6 min.

Also the seasonal period is extended by ALAN. Urban blackbirds develop a functional reproductive system 19 days in advance of their forest counterparts. ALAN affects the timing of egg laying (Kempenaers et al. 2010) and advances of gonadal growth and testosterone production by up to 1 month. This seasonal advantage for urban birds might turn to the reverse, when weather conditions inhibit food availability (Dominoni et al. 2013).

In reptiles a circadian entrainment to photoperiod was as well observed for hatching synchronization of the species *Anolis sagrei* (Nash et al. 2015). Hatching times of the singly laid lizard eggs are synchronized in the morning at dawn. Most males hatch after the females. The impact of ALAN on this synchronization is presently unknown.

Responses to Light in Mammals

Eutherian mammals escaped the predominant diurnal predator activity in the Mesozoic era, also called the age of reptiles, in developing visual and nonvisual systems of photoreception, which are characteristic for nocturnal lifestyle (Gerkema et al. 2013). Hence, the predation pressure is considered to be the main evolutionary driver for nocturnal vision in mammals (Jacobs 2009). Hölker et al. (2010) discuss that this important evolutionary step is currently threatened by the unforeseen implications of the widespread use of artificial light. Most mammals have nocturnal, crepuscular, or arrhythmic activity patterns. Exceptions are some rodents, squirrels, and primates including humans, which are predominately diurnal and have adopted their visual senses to daylight.

Light Intensity

Most nocturnal mammals have few cones, e.g., bats and armadillos. These can easily become saturated by light intensities above 120 cd m^{-2} (light level at twilight), and the animals become temporarily blinded by brighter light exposure (Beier 2006).

Some, especially fast-flying bat species, benefit from artificial light sources due to improved foraging conditions. However, slow-flying species emerge later and appear to have an innate intolerance of lit conditions, even at relatively low light levels (Stone et al. 2015a). Light that spills on commuting routes or flyways can even cause avoidance behavior of some species. The illumination of roosts or the entrance to it can cause the bats to abandon roosts in the worst case (Stone et al. 2015b). Hale et al. (2015) observed that a common urban bat (*Pipistrellus pipistrellus*) select dark crossing roots at gaps (e.g., at a streets) in urban tree networks and that the success to cross depends on light intensity and width of the habitat gap.

The suppression of plasma melatonin in mammals can be affected at very low light levels and sensitivities are highly species dependent. The light intensity during the photophase is important to trigger the scotophase melatonin response. In general, nocturnally active species are rated as more sensitive to ALAN, in terms of melatonin suppression, than diurnally active animals (Reiter et al. 2011). For example, diurnally active rodents require irradiances in the order of $1.85 \mu\text{W cm}^{-2}$ (about 13 lx), whereas in nocturnal rodents irradiances of $0.005 \mu\text{W cm}^{-2}$ (about 0.03 lx) are sufficient for complete melatonin suppression (Deveson et al. 2000). Also animal groups with similar circadian rhythms can differ in their sensitivities to suppress melatonin. Goats, for example, are less sensitive to ALAN than sheep (Deveson et al. 2000).

Color Spectra

The circadian response to light in higher vertebrates, including humans, is triggered by blue light. Wavelengths around 480 nm are most effective. At this wavelength, the sensitivity is double compared to the one of green light spectrum around 555 nm. This circadian signal is not related to visual sensitivity but perceived by melanopsin photoreceptors (Brainard et al. 2001).

The basically nocturnal nature of mammals may have led to predominately dichromatic vision (Jacobs 2009). Primates developed the additional ability to sense red color. This trichromatic vision is discussed to optimize food finding in frugivorous animals (Vorobyev 2004). In some species, females developed trichromatic vision, whereas male vision stays dichromatic (Jacobs 1994).

UV vision in mammals is discussed as restricted to rodents and marsupials. However, a study on evolutionary history of 33 bat species spanning 65 million years discovered a more important role for UV vision than previously estimated (Zhao et al. 2009). Seven insectivorous species of bats, representing five genera and three families, were recently found to be sensitive to dim-light UV vision (Gorresen et al. 2015). The European mole (*Talpa europaea*) is mainly adapted to underground life and thus the anatomy of its visual system is subjected to an involution. However, the mole was recently found to perceive colors and to be sensitive in the UV spectrum (Glösmann et al. 2008).

Timing and Duration of Illumination

Reproduction of mammals living in temperate regions follows seasonality to ensure the birth of offspring in spring or summer, when optimal survival conditions are given. Species with a short incubation or gestation period such as hamsters and species with a circannual gestation period such as horses are long-day breeders, and their fertile period occurs in the springtime. Species with a gestation period around 5–6 months such as sheep and goats are short-day breeders and their breeding takes place in autumn. The pineal hormone melatonin is the common link between photoperiod and reproduction (Gerlach and Aurich 2000). The reproductive state is not determined by the absolute daylength but by the direction of change. Rams exposed to alterations of short and long days at 1-month intervals maintain a continuously high testicular activity (Gerlach and Aurich 2000).

In order to adopt to seasonal changes, the night-active mouse lemurs (*Microcebus murinus*) turn less active and cease eating when exposed to ALAN, compared to simulated moonlight treatment. The core temperatures, during night and the day resting time, are significantly higher under exposure to ALAN, in order to acclimatize to long-day photoperiod (Le Tallec et al. 2013). Altered behavior of two spiny mouse congeners exposed to ALAN differs in chronotype. The nocturnal common spiny mouse (*Acomys cahirinus*) decreases activity and foraging with ALAN. Probably due to increased predation risk, it restricts movement particularly in exposed microhabitats. The diurnal mouse neither expands its activity into the illuminated hours, leading to reduced overall activity and a relatively underexploited temporal niche, which may promote invasion of alien species that are less light sensitive (Rotics et al. 2011).

In different studies on mammals, there is convincing evidence that extended light exposure causes weight gain, even when calorie intake and physical activity are held constant (e.g., Fonken et al. 2013; Salgado-Delgado et al. 2010). When mammals escaped the diurnal predation pressure and conquered the nocturnal niche, they had to develop mechanisms to protect from low temperature at nighttime. Brown adipose tissue (BAT) is exclusively found in mammals. Its primary function is to produce

heat to adapt to ambient temperature changes and to maintain the balance between energy intake and energy expenditure by means of disposing the extra energy as heat. White adipose tissue (WAT) acts as an energy storage and cold acclimation in rodents has been shown to induce a transdifferentiation from WAT to BAT (Tam et al. 2012). Melatonin supplementation to small mammals promotes recruitment of BAT, thus increases the thermogenic capacity and activity. The suppression of melatonin due to exposure to ALAN after the onset of darkness is therefore discussed to lead to body weight gain and adipose storage both in human and other mammals (Tan et al. 2011).

Melatonin further acts on the immune system as stimulant, providing a pre-activated state for a more effective early immune response against external stressors (Carrillo-Vico et al. 2013). The review links melatonin to immune system stimulation against infections (bacteria, viruses, and parasites) and against autoimmunity (multiple sclerosis, rheumatoid arthritis, type 1 diabetes, etc.), for increased antibody titers after vaccination, for preventing organ rejection after transplantation, and for an altered immune response in senescent rodents. The chronic suppression of melatonin after the onset of darkness might come along with further medical costs, which urgently need to be considered when, e.g., reasoning shift works with efficiency for the gross national product.

Conclusion

The introduction of artificial light has caused an unprecedented transformation of nocturnal landscapes over large areas of the Earth (Kyba and Höölker 2013). However, the emission of light is rarely neutral and may have serious ecological and evolutionary implications for many organisms, from bacteria to mammals, and may reshape entire ecosystems (Kyba and Höölker 2013). The emitted signal of ALAN is in most cases underestimated and is not yet considered as a major pollutant, but the breadth of biological impacts of ALAN provides strong evidence that ALAN can be considered as one major stressor for organisms similar to noise, soil, water, and air pollution. The question arises if ALAN impacts entire nightscapes, then to what extent is it connected to biodiversity loss worldwide. Furthermore, the increasing use of illumination technology that emits a great part in the blue spectrum, e.g., fluorescent light and LED, is a rather critical trend. Blue light is a major circadian signal for higher vertebrates (including humans) and can substantially impact insect orientation.

In order to reduce the various effects on single species and thus cascading effects in communities and ecosystems, or on ALAN-induced fragmentation of habitats, it is necessary to consider all possible measures to identify ways in which practical steps can be taken to reduce environmental concerns (Schroer and Höölker 2014). Accordingly, ALAN needs to be directed to where it is needed, only used when necessary and in the lowest intensity required for its use. For example, the illumination of historical sites could be beautiful and of great benefit to humans, but at times when humans are sleeping, the lighting of sites is rather inadequate. An over-illumination

of sites could result in glare and the benefits of illumination might be reversed. When illuminating walls and natural rocks, inhabited by endangered species, the environmental impact needs to be considered besides the costs for energy consumption. During migrating periods of ALAN-sensitive fish or at times when insects like mayflies emerge in large quantities, the illumination of bridges should be reduced. In addition, the installation of artificial lights close to seminatural spaces such as green or blue urban areas should be avoided whenever possible and as far as the security and safety requirements will allow this. For the protection of biodiversity, natural darkness reserves are important, as well as the development of corridors to connect habitats of nocturnal species that avoid lit conditions. Reducing light pollution will reduce the necessary distance of urban estates to natural dark sky preserves (Aubé 2015).

Future Research

Future research on the environmental effects of ALAN is critically important to understand the influence of ALAN and how to use innovative lighting technologies in order to mitigate negative consequences for humans and nature.

Hölker et al. (2010a), Gaston et al. (2013), and Gaston et al. (2015) describe comprehensively the need of future research on the ecological impacts of light pollution condensed in four key issues:

1. To what extent does the disruption of natural light regimes by ALAN influence populations, communities, and ecosystems? Of concern is not only direct illumination but also skyglow and how the two interact.
2. At which thresholds of light intensity and duration does artificial lighting become light pollution with a significant and relevant ecological impact, and how do these thresholds depend on spectral composition?
3. What size of “dark refuges”, where the intensity and/or duration of artificial light fall below such thresholds, is necessary to maintain natural ecosystem processes?
4. What technologies and alternative lighting strategies can address the environmental disadvantages of current lighting practices in different natural areas or settlement types?

Outdoor illumination occurs predominantly with streetlighting, airports, sea ports, industrial areas, stadiums, and public service areas (Kuechly et al. 2012; Kyba et al. 2015). In most countries, streetlighting is based on international accepted levels (e.g., CIE 115 2010); for an overview see also Kyba et al. (2014). Brightness levels produced in rural and seminatural areas on country roads, cycle and footpaths can easily reach light levels that disturb many organisms, which are active at night. Furthermore, as the growth of cities and the development of lighting technologies continue, ALAN is increasingly modifying natural light regimes by encroaching on dark refuges in space, in time, and across wavelengths (Aubrecht et al. 2010).

Cathey and Campbell (1975) suggest for the well-being of plants to shut off streetlamps 2–4 h during the early part of the evening. This timing unfortunately conflicts with safety and security issues. But, with the knowledge of species-specific action spectra and considering the rapid introduction of new lighting systems, there is today a great opportunity to adjust ALAN in order to reduce any negative environmental impacts (Gaston et al. 2015). A first step into this direction is the development of indices to distinguish the impact of ALAN on, e.g., photosynthesis, melatonin suppression, or star visibility (e.g., Aubé et al. 2013). Furthermore, the labeling of lighting products for consumer information and awareness requires urgent investigations.

Often street and public places are better illuminated than our living rooms. Future research toward the development of sustainable illumination should allow us to better balance human requirements for lit spaces with the environmental needs for unlit spaces. The goal should be to provide the light needed for any given task while minimizing both the energy use and negative environmental side effects of the light (Kyba et al. 2014).

Last but not least, maps that describe the rapid changes in ALAN are urgently needed. By developing high-resolution maps of artificial light, it will be possible to analyze the degree to which different land uses are responsible for local light pollution, which is of particular interest for conservation management (Aubrecht et al. 2008; Kuechly et al. 2012). For example, the world atlas of artificial light sky brightness (Cinzano et al. 2001) allows to determine areas of good and bad lighting and to establish and protect refuges for ALAN-sensitive organisms.

The rising of public awareness of the broad range of practices and impacts is an important step, in order to reduce negative consequences for humans and nature by the implementation of advanced lighting technologies. The recently significant increase in public involvement in light pollution research goes into this direction (e.g., <http://lossofthenight.blogspot.de>, <http://www.stars4all.eu>, <http://tatortgewasser.de>, <http://www.citiesatnight.org>) and has proven that citizen scientists provide valuable research data, for example, for monitoring global night sky luminance (Kyba et al. 2013).

Acknowledgment We want to acknowledge the support by the European Cooperation in Science and Technology (COST) through the Action ES1204 LoNNe (Loss of the Night Network) and the national support by both the German Federal Ministry of Research and Technology (support code: 033L038A) and the Federal Agency for Nature Conservation (support code: 3514821700).

References

- Ahmad M, Cashmore AR (1996) Seeing blue: the discovery of cryptochrome. *Plant Mol Biol* 30 (5):851–861
- Altshuler DL (2001) Ultraviolet reflectance in fruits, ambient light composition and fruit removal in a tropical forest. *Evol Ecol Res* 3(7):767–778
- Ashmore MR (2005) Assessing the future global impacts of ozone on vegetation. *Plant Cell Environ* 28(8):949–964

- Aubé M (2015) Physical behaviour of anthropogenic light propagation into the nocturnal environment. *Philos Trans R Soc Lond B Biol Sci* 370(1667):2014.0117
- Aubé M, Roby J, Kocifaj M (2013) Evaluating potential spectral impacts of various artificial lights on melatonin suppression, photosynthesis, and star visibility. *Plos One* 8(7):e67789
- Aubrecht C, Elvidge CD, Ziskin D, Longcore T, Rich C (2008) “When the lights stay on” – a novel approach to assessing human impact on the environment. *Earth*. <http://www.earthzine.org/2008/12/31/when-the-lights-stay-on-a-novel-approach-to-assessing->
- Aubrecht C, Malanding J, Sherbinin De A (2010) Global assessment of light pollution impact on protected areas. Retrieved from <http://www.ciesin.columbia.edu/publications.html>
- Auldridge ME, Forest KT (2011) Bacterial phytochromes: more than meets the light. *Crit Rev Biochem Mol Biol* 46(1):67–88
- Avilés JM, Parejo D (2013) Colour also matters for nocturnal birds: Owllet bill coloration advertises quality and influences parental feeding behaviour in little owls. *Oecologia* 173(2):399–408
- Baker BJ, Richardson JML (2006) The effect of artificial light on male breeding-season behaviour in green frogs, *Rana clamitans melanota*. *Can J Zool* 84(10):1528–1532
- Beier P (2006) Effects of artificial night lighting on terrestrial mammals. In: Rich C, Longcore T (eds) *Ecological Consequences of Artificial Night Lighting*. Island Press, Washington, DC, pp 19–42
- Bennie J, Davies TW, Cruse D, Inger R, Gaston KJ (2015) Cascading effects of artificial light at night: resource-mediated control of herbivores in a grassland ecosystem. *Philos Trans R Soc Lond B Biol Sci* 370:20140131
- Bennett ATD, Cuthill IC (1994) Ultraviolet vision in birds : what is its function? *Vision Res* 34 (11):1471–1478
- Berson DM, Dunn FA, Takao M (2002) Phototransduction by retinal Ganglion cells that set the Circadian clock. *Science* 295(5557):1070–1073
- Bishop JE (1969) Light control of aquatic insect activity and drift. *Ecology* 50:371–380
- Boeuf G, Le Bail P-Y (1999) Does light have an influence on fish growth? *Aquaculture* 177 (1–4):129–152
- Brainard GC, Hanifin JP, Rollag MD, Greeson J, Byrne B, Glickman G, Gerner E, Sanford B (2001) Human melatonin regulation is not mediated by the three cone photopic visual system. *J Clin Endocrinol Metab* 86(1):433–436
- Briggs WR, Christie JM (2002) Phototropins 1 and 2: versatile plant blue-light receptors. *Trends Plant Sci* 7(5):204–210
- Brüning A, Hölker F, Wolter C (2011) Artificial light at night: implications for early life stages development in four temperate freshwater fish species. *Aquat Sci* 73(1):143–152
- Brüning A, Hölker F, Franke S, Preuer T, Kloas W (2015) Spotlight on fish: light pollution affects circadian rhythms of European perch but does not cause stress. *Sci Total Environ* 511:516–522
- Brüning A, Hölker F, Franke S, Kleiner W, Kloas W (2016) Impact of different colours of artificial light at night on melatonin rhythm and gene expression of gonadotropins in European perch. *Sci Total Environ* 543:214–222
- Buchanan BW (2006) Observed and potential effects of artificial night lighting on anuran amphibians. In: Rich C, Longcore T (eds) *Ecological consequences of artificial night lighting*. Island Press, Washington, DC, pp 192–220
- Bünning E, Moser I (1969) Interference of moonlight with the photoperiodic measurement of time by plants, and their adaptive reaction. *Proc Natl Acad Sci* 62(4):1018–1022
- Canavero A, Arim M (2009) Clues supporting photoperiod as the main determinant of seasonal variation in amphibian activity. *J Nat Hist* 43(47–48):2975–2984
- Carrillo-Vico A, Lardone P, Álvarez-Sánchez N, Álvarez-Sánchez A, Guerrero J (2013) Melatonin: buffering the immune system. *Int J Mol Sci* 14(4):8638–8683
- Carton L, Josef N, Lerner A, McCusker SD, Darmaillacq A-S, Dickel L, Shashar N (2013) Polarization vision can improve object detection in turbid waters by cuttlefish. *J Exp Mar Biol Ecol* 447:80–85

- Cathey HM, Campbell LE (1975) Security lighting and its impact on the landscape. *J Arboric* 1(1):181–187
- Chen M, Chory J, Fankhauser C (2004) Light signal transduction in higher plants. *Annu Rev Genet* 38:87–117
- CIE 115 (2010) Lighting of roads for motor and pedestrian traffic
- Cinzano P, Falchi F, Elvidge CD (2001) The first World Atlas of the artificial night sky brightness. *Mon Not R Astron Soc* 707:689–707
- Collin SP, Davies WL, Hart NS, Hunt DM (2009) The evolution of early vertebrate photoreceptors. *Philos Trans R Soc Lond Ser B Biol Sci* 364(1531):2925–2940
- Cronin TW, Caldwell RL, Marshall J (2010) Sensory adaptation: tunable colour vision in a mantis shrimp. *Nature* 411(6837):527
- Da Silva A, Samplonius JM, Schlicht E, Valcu M, Kempenaers B (2014) Artificial night lighting rather than traffic noise affects the daily timing of dawn and dusk singing in common European songbirds. *Behav Ecol* 25(5):1037–1047
- Dacke M, Nilsson D-E, Scholtz CH, Byrne M, Warrant EJ (2003) Insect orientation to polarized moonlight. *Nature* 424(6944):33–33
- Danks HV (2005) How similar are daily and seasonal biological clocks? *J Insect Physiol* 51(6):609–619
- Davies TW, Bennie J, Inger R, de Ibarra NH, Gaston KJ (2013) Artificial light pollution: are shifting spectral signatures changing the balance of species interactions? *Glob Chang Biol* 19(5):1417–1423
- Davies TW, James PD, Bennie J, Gaston KJ (2014) The nature, extent, and ecological implications of marine light pollution. *Front Ecol Environ* 12(6):347–355
- De Jong M, Jeninga L, Ouyang JQ, van Oers K, Spoelstra K, Visser ME (2016) Dose-dependent responses of avian daily rhythms to artificial light at night. *Physiol Behav* 155:172–179
- Deveson SL, Arendt J, Forsyth IA (2000) Sensitivity of goats to a light pulse during the night as assessed by suppression of melatonin concentrations in the plasma. *J Pineal Res* 177(1990):169–177
- Dominoni DM, Partecke J (2015) Does light pollution alter daylength? A test using light loggers on free-ranging European blackbirds (*Turdus merula*). *Philos Trans R Soc Lond Ser B Biol Sci* 370:20140118
- Dominoni DM, Quetting M, Partecke J (2013) Artificial light at night advances avian reproductive physiology. Artificial light at night advances avian reproductive physiology. *Proc R Soc B Biol Sci* 280:20123017
- Döring TF (2014) How aphids find their host plants, and how they don't. *Ann Appl Biol* 165(1):3–26
- Durrant J, Michaelides EB, Rupasinghe T, Tull D, Green MP, Jones TM (2015) Constant illumination reduces circulating melatonin and impairs immune function in the cricket *Teleogryllus commodus*. *Peer J* 3:e1075
- Eisenbeis G (2006) Artificial night lighting and insects: attraction of insects to streetlamps in a rural setting in Germany. In: Rich C, Longcore T (eds) *Ecological consequences of artificial night lighting*. Island Press, Washington, DC, pp 191–198
- Eisenbeis G, Eick K (2011) Studie zur Anziehung nachtaktiver Insekten an die Straßenbeleuchtung unter Einbeziehung von LEDs. *Natur Landschaft* 86(7):298–306
- Ekström P, Meissl H (2003) Evolution of photosensory pineal organs in new light: the fate of neuroendocrine photoreceptors. *Philos Trans R Soc Lond Ser B Biol Sci* 358(1438):1679–1700
- Fan X-X, Xu Z-G, Liu X-Y, Tang C-M, Wang L-W, Han X (2013) Effects of light intensity on the growth and leaf development of young tomato plants grown under a combination of red and blue light. *Sci Hort* 153:50–55
- Fonken LK, Lieberman RA, Weil ZM, Nelson RJ (2013) Dim light at night exaggerates weight gain and inflammation associated with a high-fat diet in male mice. *Endocrinology* 154(10):3817–3825

- Frank KD (2006) Effects of artificial night lighting on moths. In: Rich C, Longcore T (eds) *Ecological consequences of artificial night lighting*. Island Press, Washington, DC, pp 305–344
- Fulgione D, Trapanese M, Maselli V, Ripa D, Itri F, Avallone B, Van Damme R, Monti DM, Raia P (2014) Seeing through the skin: dermal light sensitivity provides cryptism in moorish gecko. *J Zool* 294(2):122–128
- Gaston KJ, Bennie J, Davies TW, Hopkins J (2013) The ecological impacts of nighttime light pollution: a mechanistic appraisal. *Biol Rev Camb Philos Soc* 88(4):912–927
- Gaston KJ, Visser ME, Hölker F (2015) The biological impacts of artificial light at night : the research challenge. *Philos Trans R Soc Lond Ser B Biol Sci* 370:20140133
- Geffen van KG (2015) Moths in illuminated nights -Artificial night light effects on moth ecology. Doctoral thesis. Wageningen University
- Gerkema MP, Davies WIL, Foster RG, Menaker M, Hut RA (2013) The nocturnal bottleneck and the evolution of activity patterns in mammals. *Proc R Soc B Biol Sci* 280(1765):20130508
- Gerlach T, Aurich JE (2000) Regulation of seasonal reproductive activity in the stallion, ram and hamster. *Anim Reprod Sci* 58(3–4):197–213
- Gerlach T, Sprenger D, Michiels NK (2014) Fairy wrasses perceive and respond to their deep red fluorescent coloration. *Proc R Soc B Biol Sci* 281(1787):2014.0787
- Giraud E, Verméglio A (2008) Bacteriophytochromes in anoxygenic photosynthetic bacteria. *Photosynth Res* 97(2):141–153
- Glösmann M, Steiner M, Peichl L, Ahnelt PK (2008) Cone photoreceptors and potential UV vision in a subterranean insectivore, the European mole. *J Vis* 8(4):23
- Gorresen MP, Cryan PM, Dalton DC, Wolf S, Bonaccorso FJ (2015) Ultraviolet vision may be widespread in bats. *Acta Chiropterologica* 17(1):193–198
- Hadderingh RH, Van Aerssen GHFM, De Beijer RFLJ, Van Der Velde G (1999) Reaction of silver eels to artificial light sources and water currents: an experimental deflection study. *Regul Rivers: Res Manage* 15(4):365–371
- Hagen O, Santos RM, Schlindwein MN, Viviani VR (2015) Artificial night lighting reduces firefly (Coleoptera : Lampyridae) occurrence in Sorocaba, Brazil. *Adv Entomol* 3(01):24–32
- Hale JD, Fairbrass AJ, Matthews TJ, Davies G, Sadler JP (2015) The ecological impact of city lighting scenarios: exploring gap crossing thresholds for urban bats. *Glob Chang Biol* 21(7):2467–2478
- Hankins MW, Peirson SN, Foster RG (2008) Melanopsin: an exciting photopigment. *Trends Neurosci* 31(1):27–36
- Hattar S, Liao H-W, Takao M, Berson DM, Yau K-W (2002) Melanopsin-containing retinal ganglion cells : architecture, projections, and intrinsic photosensitivity. *Science* 295(5557):1065–1071
- Heiling AM (1999) Why do nocturnal orb-web spiders (Araneidae) search for light? *Behav Ecol Sociobiol* 46(1):43–49
- Hölker F, Moss T, Griefahn B, Kloas W, Voigt CC (2010a) The dark side of light : a transdisciplinary research agenda for light. *Ecology And Society* 15(4):Art 13
- Hölker F, Wolter C, Perkin EK, Tockner K (2010b) Light pollution as a biodiversity threat. *Trends Ecol Evol* 12:681–682
- Hori M, Shibuya K, Sato M, Saito Y (2014) Lethal effects of short-wavelength visible light on insects. *Sci Rep* 4:7383
- Horváth G (2014) *Polarized light and polarization vision in animal sciences*, 2nd edn. Springer, Berlin/Heidelberg
- Horváth G, Csabai Z (2014) Polarization vision of aquatic insects. In: Horváth G, (eds) *Polarized light and polarization vision in animal sciences*. Springer Science & Business Media. Berlin/Heidelberg/New York, pp 113–145
- Horváth G, Kriska G, Malik P, Robertson B (2009) Polarized light pollution: a new kind of ecological photopollution. *Front Ecol Environ* 7(6):317–325
- Itoh MT, Hattori A, Sumi Y, Suzuki T (1995) Day-night changes in melatonin levels in different organs of the cricket (*Gryllus bimaculatus*). *J Pineal Res* 18(3):165–169

- Jacobs GH (1994) Variations in primate color vision: mechanisms and utility. *Evol Anthropol Issues News Rev* 3(6):196–205
- Jacobs GH (2009) Evolution of colour vision in mammals. *Philos Trans R Soc Lond B Biol Sci* 364 (1531):2957–2967
- Jahan SMH, Lee G-S, Lee S, Lee K-Y (2014) Acquisition of Tomato yellow leaf curl virus enhances attraction of *Bemisia tabaci* to green light emitting diodes. *J Asia Pac Entomol* 17(1):79–82
- Johansen NS, Vänninen I, Pinto DM, Nissinen AI, Shipp L (2011) In the light of new greenhouse technologies: 2. Direct effects of artificial lighting on arthropods and integrated pest management in greenhouse crops. *Ann Appl Biol* 159(1):1–27
- Johanson U, Gehrke C, Björn LO, Callaghan TV, Sonesson M (1995) The effects of enhanced UV-B radiation on a subarctic heath ecosystem. *R Swedish Acad Sci* 24(2):106–111
- Johnson SD, Andersson S (2002) A simple field method for manipulating ultraviolet reflectance of flowers. *Can J Bot* 80(12):1325–1328
- Juell J-E, Fosseidengen JE (2004) Use of artificial light to control swimming depth and fish density of Atlantic salmon (*Salmo salar*) in production cages. *Aquaculture* 233(1–4):269–282
- Kelber A, Roth LSV (2006) Nocturnal colour vision-not as rare as we might think. *J Exp Biol* 209 (5):781–788
- Kelber A, Vorobyev M, Osorio D (2003) Animal colour vision-behavioural tests and physiological concepts. *Biol Rev* 78(1):81–118
- Keller MM, Jaillais Y, Pedmale UV, Moreno JE, Chory J, Ballaré CL (2011) Cryptochrome 1 and phytochrome B control shade-avoidance responses in *Arabidopsis* via partially independent hormonal cascades. *Plant J Cell Mol Biol* 67(2):195–207
- Kempnaers B, Borgström P, Loës P, Schlicht E, Valcu M (2010) Artificial night lighting affects dawn song, extra-pair siring success, and lay date in songbirds. *Curr Biol* 20(19):1735–1739
- Kleinlogel S, White AG (2008) The secret world of shrimps: polarisation vision at its best. *PLoS One* 3(5):e2190
- Kleinteich A, Schneider JM (2011) Developmental strategies in an invasive spider: constraints and plasticity. *Ecol Entomol* 36(1):82–93
- Kolkovski S, Dabrowski K (1998) Off-season spawning of yellow perch. *Prog Fish Cult* 60 (2):133–136
- Krijgsveld KL, Fijn RC, Lensink R (2015) Occurrence of peaks in songbird migration at rotor heights of offshore wind farms in the southern North Sea. Report
- Kritsky MS (1984) The blue light responses in evolutionary studies. In: Senger H (ed) *Blue light Effects in biological systems*. Springer, Berlin/Heidelberg, pp 3–5
- Kuechly HU, Kyba CCM, Ruhtz T, Lindemann C, Wolter C et al (2012) Aerial survey and spatial analysis of sources of light pollution in Berlin, Germany. *Remote Sens Environ* 126:39–50
- Kummu M, De Moel H, Ward PJ, Varis O (2011) How close do we live to water? A global analysis of population distance to freshwater bodies. *PLoS One* 6(6):e20578
- Kurvers RHJM, Hölker F (2014) Bright nights and social interactions: a neglected issue. *Behav Ecol* 26(2):334–339
- Kyba CCM, Hölker F (2013) Do artificially illuminated skies affect biodiversity in nocturnal landscapes? *Landsc Ecol* 28(9):1637–1640
- Kyba CCM, Ruhtz T, Fischer J, Hölker F (2011) Lunar skylight polarization signal polluted by urban lighting. *J Geophys Res* 116(D24):1–7
- Kyba CCM, Wagner JM, Kuechly HU, Walker CE, Elvidge CD et al (2013) Citizen science provides valuable data for monitoring global night sky luminance. *Sci Rep* 3:1835
- Kyba CCM, Hänel A, Hölker F (2014) Redefining efficiency for outdoor lighting. *Energy Environ Sci* 7(6):1806–1809
- Kyba CCM, Garz S, Kuechly H, De Miguel A, Zamorano J, Fischer J, Hölker F (2015a) High-resolution imagery of earth at night: new sources, opportunities and challenges. *Remote Sens* 7 (1):1–23
- Kyba CCM, Tong KP, Bennie J, Birriel I, Birriel JJ et al (2015b) Worldwide variations in artificial skyglow. *Sci Rep* 5:8409

- Le Tallec T, Perret M, Théry M (2013) Light pollution modifies the expression of daily rhythms and behavior patterns in a nocturnal primate. *Plos One* 8(11):e79250
- Lerner A, Meltser N, Sapir N, Erlick C, Shashar N, Broza M (2008) Reflected polarization guides chironomid females to oviposition sites. *J Exp Biol* 21(22):3536–3543
- Lerner A, Sabbah S, Erlick C, Shashar N (2011) Navigation by light polarization in clear and turbid waters. *Philos Trans R Soc Lond B Biol Sci* 366(1565):671–679
- Litman BJ, Mitchell DC (1996) Rhodopsin structure and function. *Biomembranes A Multi-Volume Treatise* 2:1–32
- Lohman KJ (2010) Q&A: animal behaviour: magnetic-field perception. *Nature* 464(7292):1140–1142
- Longcore T, Rich C (2004) Ecological light pollution. *Front Ecol Environ* 2(4):191–198
- Lythgoe JN (1988) Light and vision in the aquatic environment. In: Atema J, Fay RR, Popper AN, Tavolga WN (eds) *Sensory biology of aquatic animals*. Springer, New York, pp 57–87
- MacQuarrie DW, Vanstone WE, Markert JR (1979) Photoperiod induced off-season spawning of pink salmon (*Oncorhynchus gorbuscha*). *Aquaculture* 18(4):289–302
- Martin G, Rojas LM, Ramirez Y, McNeil R (2004) The eyes of oilbirds (*Steatornis caripensis*): pushing at the limits of sensitivity. *Naturwissenschaften* 91(1):26–29
- Martin M, Le Galliard J-F, Meylan S, Loew ER (2015) The importance of ultraviolet and near-infrared sensitivity for visual discrimination in two species of lacertid lizards. *J Exp Biol* 218(3):458–465
- Massa GD, Drive AM, Lafayette W, Kim H, Wheeler RM, Mitchell CA (2008) Plant productivity in response to LED lighting. *Hortic Sci* 43(7):1951–1956
- Matthews RW, Matthews JR (2009) Light reception. In: *Insect behaviour*, 2nd edn. Springer, Dordrecht/Heidelberg/London/New York, pp 268–277
- Matzke EB (1936) The effect of street lights in delaying leaf-fall in certain trees. *Am J Bot* 23(6):446–452
- Mazza CA, Izaguirre MM, Zavala J, Ana LS, Ballaré CL (2002) Insect perception of ambient ultraviolet-B radiation. *Ecol Lett* 5(6):722–726
- McAlary FA, McFarland WN (1993) The effect of light and darkness on hatching in the pomacentrid *Abudefduf saxatilis*. *Environ Biol Fishes* 37(3):237–244
- McClung CR (2006) Plant Circadian rhythms. *Plant Cell* 18(4):792–803
- McMahon DG, Iuvone PM, Tosini G (2014) Circadian organization of the mammalian retina: from gene regulation to physiology and diseases. *Prog Retin Eye Res* 39:58–76
- Mège P, Ödeen A, Théry M, Picard D, Secondi J (2016) Partial Opsin sequences suggest UV-sensitive vision is widespread in Caudata. *Evol Biol* 43:109–118
- Moore MV, Pierce SM, Walsh HM, Kvalvik SK, Lim JD (2000) Urban light pollution alters the diel vertical migration of *Daphnia*. *Internationale Vereinigung Fur Theoretische Und Angewandte Limnologie Verhandlungen* 27(2):779–782
- Myrberg AA, Fuiman LA (2002) The sensory world of coral reef fishes. In: *Coral reef fishes: dynamics and diversity in a complex ecosystem*. Academic, San Diego, pp 123–148
- Nash J, Price J, Cox RM (2015) Photoperiodic hatching rhythms suggest Circadian entrainment of *Anolis sagrei* Eggs. *J Herpetol* 29(4):611–615
- Navarro-Barranco C, Hughes LE (2015) Effects of light pollution on the emergent fauna of shallow marine ecosystems: Amphipods as a case study. *Mar Pollut Bull* 94(1):235–240
- Nightingale B, Longcore T, Simenstad CA (2006) Artificial night lighting and fishes. In: Rich C, Longcore T (eds) *Ecological consequences of artificial night lighting*. Island Press, Washington, DC, pp 257–276
- Nordt A, Klenke R (2013) Sleepless in town – drivers of the temporal shift in Dawn song in Urban European Blackbirds. *Environ Res* 8(8):1–10
- Ogden LJE (1996) Collision course: the hazards of lighted structures and windows to migrating birds collision course. <http://digitalcommons.unl.edu/flap/3>

- Ollivier FJ, Samuelson DA, Brooks DE, Lewis PA, Kallberg ME, Komaromy AM (2004) Comparative morphology of the tapetum lucidum (among selected species). *Vet Ophthalmol* 7 (1):11–22
- Ouyang J, De Jong M, Hau M, Visser ME, Van Grunsven RHA, Spoelstra K (2015) Stressful colours: corticosterone concentrations in a free-living songbird vary with the spectral composition of experimental illumination. *Biol Lett* 11(18):20150517
- Pawson SM, Bader MF (2014) LED lighting increases the ecological impact of light pollution irrespective of color temperature. *Ecol Appl* 24(7):1561–1568
- Perkin EK, Hölker F, Richardson JS, Sadler JP, Wolter C, Tockner K (2011) The influence of artificial light on stream and riparian ecosystems: questions, challenges, and perspectives. *Ecosphere* 2(11): art122
- Perry G, Fisher RN (2006) Night lights and reptiles: observed and potential effects. In: Rich C, Longcore T (eds) *Ecological consequences of artificial night lighting*. Island Press, Washington, DC, pp 169–191
- Perry G, Buchanan BW, Fisher RN, Salmon M, Wise SE (2008) Effects of artificial night lighting on amphibians and reptiles in urban environments. *Urban Herpetol* 3:239–256
- Pévet P, Agez L, Bothorel B, Saboureau M, Gauer F, Laurent V, Masson-Pévet M (2006) Melatonin in the multi-oscillatory mammalian circadian world. *Chronobiol Int* 23(1–2):39–51
- Poot H, Ens BJ, De Vries H, Donners M, Wernand MR, Marquenie JM (2008) Green light for nocturnally migrating birds. *Ecol Soc* 13(2):47
- Prokopy RJ, Owens ED (1983) Visual detection of plants by herbivorous insects. *Annu Rev Entomol* 28(1):337–364
- Provencio I, Rodriguez IR, Jiang G, Pa W, Moreira EF, Rollag MD (2000) A novel human Opsin in the inner retina. *J Neurosci* 20(2):600–605
- Queval G, Issakidis-Bourguet E, Hoeberichts FA, Vandenbergh M, Gakière B et al (2007) Conditional oxidative stress responses in the Arabidopsis photorespiratory mutant *cat2* demonstrate that redox state is a key modulator of daylength-dependent gene expression, and define photoperiod as a crucial factor in the regulation of H₂O₂-induced cell death. *Plant J Cell Mol Biol* 52 (4):640–657
- Raap T, Pinxten R, Eens M (2015) Light pollution disrupts sleep in free-living animals. *Sci Rep* 5:13557
- Reiter RJ, Sanchez-Barcelo E, Mediavilla M, Gitto E, Korkmaz A (2011) Circadian mechanisms in the regulation of melatonin synthesis: disruption with light at night and the pathophysiological consequences. *J Exp Integrat Med* 1(1):13–22
- Riley JR, Reynolds DR, Farmery MJ (1983) Observations of the flight behavior of the Armyworm Moth, *Spodoptera-Exempta*, at an emergence site using radar and infrared optical techniques. *Ecol Entomol* 8:395–418
- Riley WD, Davison PI, Maxwell DL, Newman RC, Ives MJ (2015) A laboratory experiment to determine the dispersal response of Atlantic salmon (*Salmo salar*) fry to street light intensity. *Freshw Biol* 60(5):1016–1028
- Rockwell NC, Lagarias JC (2010) A brief history of phytochromes. *Chemphyschem* 11 (6):1172–1180
- Roden LC, Song H, Jackson S, Morris K, Carre IA (2002) Floral responses to photoperiod are correlated with the timing of rhythmic expression relative to dawn and dusk in Arabidopsis. *Proc Natl Acad Sci* 99(20):13313–13318
- Rodriguez-Romero J, Hedtke M, Kastner C, Müller S, Fischer R (2010) Fungi, hidden in soil or up in the air: light makes a difference. *Annu Rev Microbiol* 64:585–610
- Rotics S, Dayan T, Kronfeld-Schor N (2011) Effect of artificial night lighting on temporally partitioned spiny mice. *J Mammal* 62(1):159–168
- Sabbah S, Lerner A, Erlick C, Shashar N (2005) Under water polarization vision- a physical examination. *Recent Res Develop Exp Theoret Biol* 1:123–176

- Salgado-Delgado R, Angeles-Castellanos M, Sadari N, Buijs RM, Escobar C (2010) Food intake during the normal activity phase prevents obesity and circadian desynchrony in a rat model of night work. *Endocrinology* 151(3):1019–1029
- Salmon M (2006) Protecting sea turtles from artificial night lighting at Florida's oceanic beaches. In: Rich C, Longcore T (eds) *Ecological consequences of artificial night lighting*. Island Press, Washington, DC, pp 141–168
- Sanabria EA, Quiroga LB (2011) Change in the thermal biology of tadpoles of *Odontophrynus occidentalis* from the Monte desert, Argentina: responses to photoperiod. *J Ther Biol* 36(5):288–291
- Saunders DS (2002) *Insect clocks*. Elsevier, Amsterdam
- Schaefer HM, Valido A, Jordano P (2014) Birds see the true colours of fruits to live off the fat of the land. *Proc R Soc Lond B Biol Sci* 281(1777):20132516
- Schroer S, Hölker F (2014). Light pollution reduction. In: Karlicek R, Sun C-C, Zissis G, Ma R (eds) *Handbook of advanced lighting technology*. Springer International Publishing, Switzerland, pp. 1–17
- Shimmura T, Yoshimura T (2013) Circadian clock determines the timing of rooster crowing. *Curr Biol* 23(6):231–233
- Sinnadurai S (1981) High pressure sodium street lights affect crops in Ghana. *World Crops* 33:120–122
- Stack PA, Drummond FA (1997) Reproduction and development of *Orius insidiosus* in a blue light-supplemented short photoperiod. *Biol Control* 65(9):59–65
- Stone EL, Harris S, Jones G (2015a) Impacts of artificial lighting on bats: a review of challenges and solutions. *Mammalian Biology – Zeitschrift Für Säugetierkunde* 80(3):213–219
- Stone EL, Wakefield A, Harris S, Jones G (2015b) The impacts of new street light technologies : experimentally testing the effects on bats of changing from low- pressure sodium to white metal halide. *Philos Trans R Soc Lond Ser B Biol Sci* 370(1667):20140127
- Stuart SN, Chanson JS, Cox NA, Young BE, Rodrigues AS, Fischman DL, Waller RW (2004) Status and trends of amphibian declines and extinctions worldwide. *Science* 306(5702):1783–1786
- Stutte GW (2009) Light-emitting diodes for manipulating the phytochrome apparatus. *Hortic Sci* 44(2):231–234
- Tam CS, Lecoultré V, Ravussin E (2012) Brown adipose tissue: mechanisms and potential therapeutic targets. *Circulation* 125(22):2782–2791
- Tan D-X, Manchester LC, Fuentes-Broto L, Paredes SD, Reiter RJ (2011) Significance and application of melatonin in the regulation of brown adipose tissue metabolism: relation to human obesity. *Obesity Rev Off J Int Assoc Study Obesity* 12(3):167–188
- Tani A, Shiina S, Nakashima K, Hayashi M (2014) Improvement in lettuce growth by light diffusion under solar panels. *J Agricul Meteorol* 70(3):139–149
- Tsuboi H, Wada M (2011) Chloroplasts can move in any direction to avoid strong light. *J Plant Res* 124(1):201–210
- Van Geffen KG, Van Grunsven RHA, Van Ruijven J, Van Berendse F, Veenendaal EM (2014) Artificial light at night causes diapause inhibition and sex-specific life history changes in a moth. *Ecol Evol* 4(11):2082–2089
- Van Geffen KG, Van Eck E, De Boer RA, Van Grunsven RHA, Salis L et al (2015) Artificial light at night inhibits mating in a Geometrid moth. *Insect Conser Diversity* 8(3):282–287
- Van Grunsven RHA, Donners M, Boekee K, Tichelaar I, Van Geffen KG et al (2014) Spectral composition of light sources and insect phototaxis, with an evaluation of existing spectral response models. *J Insect Conser* 18(2):225–231
- Van Langevelde F, Ettema JA, Donners M, WallisDeVries MF, Groenendijk D (2011) Effect of spectral composition of artificial light on the attraction of moths. *Biol Conserv* 144(9):2274–2281

- Vänninen I, Pinto DM, Nissinen AI, Johansen NS, Shipp L (2010) In the light of new greenhouse technologies: 1. Plant-mediated effects of artificial lighting on arthropods and tritrophic interactions. *Ann Appl Biol* 157(3):393–414
- Veerman A (2001) Photoperiodic time measurement in insects and mites : a critical evaluation of the oscillator-clock hypothesis. *J Insect Physiol* 47(10):1097–1109
- Veerman A, Veenendaal RL (2003) Experimental evidence for a non-clock role of the circadian system in spider mite photoperiodism. *J Insect Physiol* 49(8):727–732
- Vera LM, Davie A, Taylor JF, Migaud H (2010) Differential light intensity and spectral sensitivities of Atlantic salmon, European sea bass and Atlantic cod pineal glands ex vivo. *Gen Comp Endocrinol* 165(1):25–33
- Vollsnes AV, Eriksen AB, Otterholt E, Kvaal K, Oxaal U, Futsaether CM (2009) Visible foliar injury and infrared imaging show that daylength affects short-term recovery after ozone stress in *Trifolium subterraneum*. *J Exp Bot* 60(13):3677–3686
- Vorobyev M (2003) Coloured oil droplets enhance colour discrimination. *Proc R Soc B Biol Sci* 270(1521):1255–1261
- Vorobyev M (2004) Ecology and evolution of primate colour vision. *Clin Exp Optom* 87(4–5):230–238
- Warrant E (2004) Vision in the dimmest habitats on earth. *J Comp Physiol A* 190(10):765–789
- Wehner R (1984) Astronavigation in insects. *Annu Rev Entomol* 29:277–298
- Wehner R, Müller M (2006) The significance of direct sunlight and polarized skylight in the ant's celestial system of navigation. *Proc Natl Acad Sci U S A* 103(33):12575–12579
- Wiltschko R, Stapput K, Thalau P, Wiltschko W (2009) Directional orientation of birds by the magnetic field under different light conditions. *J R Soc Interface* 7:163–177
- Winter Y, López J, Helversen O (2003) Ultraviolet vision in a bat. *Nature* 425(6958):612–614
- Zaidi FH, Hull JT, Peirson SN, Wulff K, Aeschbach D et al (2007) Short-wavelength light sensitivity of circadian, pupillary, and visual awareness in humans lacking an outer retina. *Curr Biol* 17(24):2122–2128
- Zhao H, Rossiter SJ, Teeling EC, Li C, Cotton JA, Zhang S (2009) The evolution of color vision in nocturnal mammals. *Proc Natl Acad Sci U S A* 106(22):8980–8985
- Zschokke S, Herberstein ME (2005) Laboratory methods for maintaining and studying web-building spiders. *J Arachnol* 33(2):205–213

Light Pollution Reduction

Methods to Reduce the Environmental Impact of Artificial Light at Night

Sibylle Schroer and Franz Hölker

Contents

Introduction	992
Direct the Light Where It Is Needed	993
Switch the Light Off Whenever It Is Not Needed	997
Dim the Light and Choose the Most Appropriate Illuminants, Color Spectra, and Filters	999
Shade the Light to Protect Your Neighbor	1004
Learn from Nature	1004
Conclusion and Directions for Future	1006
References	1007

Abstract

Artificial light at night is an irreplaceable technology for our society and its activities at nighttime. But this indispensable tool has detrimental side effects, which have only come to light in the past 10–20 years. This chapter reviews ways to implement technology in order to lower the impact of artificial light at night on nature and humans. Further, it provides guidelines for environmental protection and scientific approaches to reduce the increase in light pollution and discusses the urgent need for further research.

Measures to prevent obtrusive light and unintentional trespass into homes and natural habitats are mostly simple solutions like shielding luminaires and predominantly require awareness. Shades are another effective tool to reduce trespass from interior lights. Especially in greenhouses, the use of shades significantly reduces the contribution to skyglow. Artificial light should be switched off whenever it is not needed. Smart, flexible lighting systems can help to use artificial light with precision. The choice of the appropriate illumination has to be balanced by the

S. Schroer (✉) • F. Hölker
Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany
e-mail: schroer@igb-berlin.de

needs for optimal visibility, human well-being, environmental conservation and protection of the night sky. For visibility, conditions comparable to bright moonlit nights (0.3 lx) are sufficient. Low-level streetlights that produce only 1–3 lx at the surface meet the requirement of facial cognition. Although this light level might be too low for road safety, a consideration of maximum illumination levels in street lighting is recommended. The spectral power distribution of illuminants can impact several environmental parameters. For example, illuminants emitting short wavelengths can suppress melatonin in higher vertebrates (including humans), are attracting many insect species, and contribute in skyglow above average. Recent findings in different measures for energy efficiency of illuminants at scotopic or mesopic vision conditions compared to photopic conditions indicate that the assessment of lighting products needs fundamental revision. Further research is crucially needed to create refuges for light-sensitive species at night, to measure the impact of artificial light on nature, and also to monitor the improvements of light pollution-reducing measures.

Decrees in various regions have helped to lower the impact of artificial light at night significantly. Measures to reduce the impact of artificial light at night need to be carefully balanced with the surrounding environment. Thoughtful guidelines are crucial to reducing the rapid increase in sky brightness worldwide. These guidelines need to be made accessible for decision makers especially in areas which require new light installations.

Introduction

Light is necessary, is pleasant, and can increase safety and security. Especially in places lacking artificial light at night, the appreciation for new light installations can be tremendous. At many places around the world, it is dangerous to walk at night, simply because obstacles cannot be seen. Moreover, in many low- and middle-income countries with two-, three-, or four-wheeled or animal-drawn vehicles and pedestrians all sharing the same road space, traffic laws are more complex than that in high-income countries and often inadequately enforced. The roads may be poorly constructed and maintained, road signs and lighting inadequate, and driving habits poor (WHO 2014). Artificial light at night is needed to provide safety and to highlight the beauty of sites. However, where light is easily accessible and inexpensive, its use can be exaggerated and a great disturbance for others. Artificial light at night can provide safety and security, but this might not be enhanced by greater brightness. Light design can underline the beauty of landmarks at night, but it can also rob us of the tremendous beauty of the night sky and our view of the Milky Way. Artificial light after the onset of darkness can suppress melatonin production of higher vertebrates, thus disrupting circadian rhythms and dysregulating organisms. Besides the direct impact by point sources of light such as streetlights, skyglow is a landscape-scale phenomenon affecting very large areas. Indeed, skyglow now constitutes one of the most important alterations of the biosphere (Kyba and Hölker 2013). Today, more than 60 % of the world human population live under

light-polluted skies (artificial sky brightness $\geq 10\%$ of the natural night sky brightness above 45° of elevation on the horizon). Ninety-nine percent of the population in the USA and Europe cannot see a pristine night sky, and almost one-fifth of the world's landmass is affected by increased artificial sky brightness (Cinzano et al. 2001). Since light emission is believed to be increasing globally at a rate of 3–6 % per year (Narisada and Schreuder 2004; Hölker et al. 2010b), we have to expect an ongoing transformation of nightscapes in the future with the consequence that the luminance of a cloudy night in urban areas can be up to thousands of times brighter than natural (Kyba and Hölker 2013).

But what are the thresholds for light turning from a desired resource to pollution, e. g., disruption of human health and ecological communities? And how can we use artificial light in a way that maximizes its economic and social benefits while minimizing its negative, unintended ecological and health impacts? Rules for the reduction of light pollution can be simple and need our attention, which can be increased by improving knowledge about the perception of light. An improved understanding about the impacts of artificial light at night could result in more advanced regulations and guidelines and the development of adaptive and context-dependent lighting concepts. These will help countries, regions, and cities to define thresholds for light pollution while maintaining the beneficial aspects of artificial light at night (Hölker et al. 2010a).

Recently, calls for a more environmental-friendly use of light have been increased, and different research groups have been analyzing the costs of artificial lights at night. A growing number of articles have pointed out options for solutions:

- Habitat conservation in protecting species-rich areas from artificial light.
- Determine thresholds and upper limits for light emission.
- Environment-specific light brightness adapted to the surrounding area.
- Time control of light emission and diurnal adapted color spectrum.
- Reducing trespass of lighting.

(Cinzano 2002; Longcore and Rich 2006; Navara and Nelson 2007; Eisenbeis and Hänel 2009; Hölker et al. 2010b; Bruce-White and Shardlow 2011; Gaston et al. 2012; Hölker 2013; Dick 2014).

In this chapter, options to minimize detrimental effects of artificial light at night on the environment are described in five steps.

Direct the Light Where It Is Needed

Artificial light associated with streets was found to be a major source for zenith-directed light emission (Kuechly et al. 2012). Streetlight is intended to illuminate traffic and infrastructure, but obtrusive light into adjacent homes, the environment, or even the universe should be prevented. Light that shines beyond the target area creates glare, without the benefit of illumination.

The Commission Internationale de l'Éclairage (CIE 1997) recommends a maximum illuminance on the flat surface of a window to limit light trespass into a

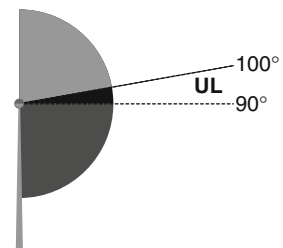
building and a maximum upward light ratio to reduce skyglow. For the regulation of light emission, the CIE developed four environmental zones with E1 being the strictest category for national parks and dark landscapes. The obtrusive light is regulated for luminaires in E1 with 0 % direct emission into the sky, and the vertical illuminance permitted into windows should be limited to 1 lx. In 2007, the CIE released a recommendation for road transport lighting in developing countries (CIE 2007). Unfortunately, in this document the ratio for obtrusive light is only mentioned in a short paragraph. Light emissions from areas adjacent to protected lands have environmentally critical side effects, which need careful regulation (Aubrecht et al. 2010). Since 2001, the National Park Service of the USA analyzes the nightly sky quality in about 100 US national parks and found increased sky brightness for almost every park (National Park Service 2012). Why should developing countries need to repeat the same mistakes we already learned from? Understanding the detrimental effects of light pollution is of major importance to protect the value of nature as the developing world installs necessary streetlights. Recommendations for the protection of the environment by shielding artificial light should be prioritized and information made easily accessible for decision makers.

The Institution of Lighting Professionals (ILP) (formerly called the Institution of Lighting Engineers, ILE) nicely illustrated the design of luminaires to reduce light pollution (ILP 2011). Globe lamps without shields should be banned, and fully shielded luminaires that direct the light downward to reduce glare should be preferred. For minimizing skyglow upward light output (UL) between 90° and 100° is the most critical (Fig. 1).

The angle of the luminaires should be $\leq 70^\circ$ (Fig. 2a); higher mounting of luminaires is recommended, because this allows lower beam angles, which prevent glare. It is recommended that the double-asymmetric beams are designed in such way that the front glazing is kept parallel to the surface (Fig. 2b), as well as modern, well-controlled projector-type luminaires, which can be aimed very precisely. Upward directed light should be avoided, but if inevitable, the use of shield baffles and louvers can reduce the spillover around the target structure (Fig. 2c).

The ILP complies with the four environmental zones developed by the CIE and supplies them with much higher upper limits for obtrusive light (Table 1). In 2011 the institution added another zone, E0, for protected areas, where natural darkness is desired. They recommend embedding the E0 zone in an E1 zone with 0 % obtrusive light (ILP 2011).

Fig. 1 Critical luminaire angles for minimizing skyglow. Critical angles for upward light output (UL) are between 90° and 100° (Reillustrated by Anna Rothmund with kind permission from the ILP (2011))



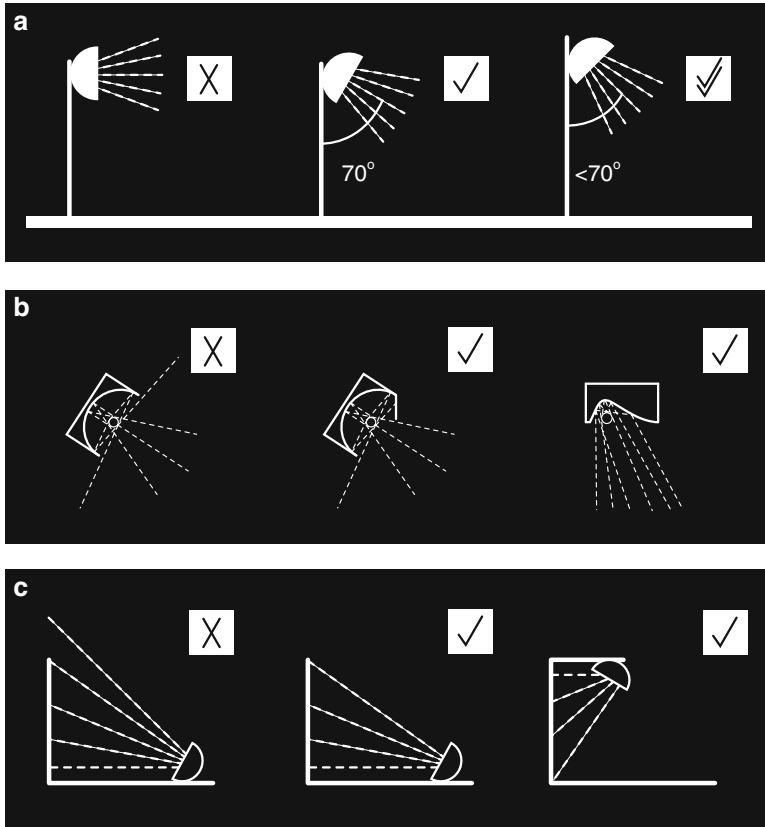


Fig. 2 Design of luminaires to reduce light pollution. (a): Keep the beam angle below 70°. (b): Minimize the upward light output. (c) For vertical illumination, reduce the spill-over around the target structure (Reillustrated by Anna Rothmund with kind permission from the ILP (2011))

Table 1 Zone rating according to the CIE (1997) and ILP (2011) with different upward light ratio limits for luminaire installations

Zone rating	ULR (%) CIE	ULR (%) ILE/ILP	Surrounding	Lighting environment
E0 (ILP 2011)	–	0	Protected	Dark
E1	0	0	Natural	Intrinsically dark
E2	0–5	2.5	Rural	Low district brightness
E3	0–15	5	Suburban	Medium district brightness
E4	0–25	15	Urban	High district brightness

ULR upward light ratio of the installation is the maximum permitted percentage of luminaire flux for the total installation that goes directly into the sky

Examples of Acceptable / Unacceptable Lighting Fixtures



Fig. 3 The good and the bad examples of unshielded and comparable fully shielded luminaires (Reprinted with kind permission from the IDA (IDA/IESNA 2011), Copyright Bob Crelin/BobCrelin.com)

The International Dark-Sky Association (IDA) together with the Illuminating Engineering Society of North America (IESNA) developed the Model Lighting Ordinance in 2011. They recommend the use of fully shielded luminaires everywhere (Fig. 3). Using similar lighting zones to the ILP, they recommend downward

directed light with only low glare in zones 0–2 and the minimization of upward directed light in the zones 3 and 4 (IDA/IESNA 2011).

Gaston et al. (2012) recommend focusing reflectors to enhance the efficiency of luminaires and to direct the light to where it is needed. For sensitive areas, useful protection from light emission can also be provided by walls or plantings of vegetation, such as hedges.

Dick (2014) demands not only full cutoff luminaires, which still have a tolerance threshold of 10 % light to be emitted within 10° in the horizon, but sharp cutoff luminaires, which lower this 10 % even further to 1 %.

Best practice lighting installations have been estimated to reduce skyglow at the zenith up to 90 % and up to 82 % over the entire sky when viewed from a distant location (Duriscoe et al. 2014). The IDA has spent 2 years working with the United States National Park Service to analyze current outdoor lighting practices in national parks and to develop best practice standards that can be used in environmentally sensitive areas. This knowledge will be published in the form of a CIE report, which forms the basis of the recommendation for the CIE Technical Committee 5–27: “Artificial lighting and its impact on the natural environment” (Parks 2013).

Methods to reduce obtrusive light and glare have been issued regionally by many governments, for example:

- The Sky Law – Canary Islands, 1988
- Chilean Decree for Light Pollution Regulation, 1998
- Visual Lombardy Law no. 17, 2000
- Light pollution laws in several US states, about 2000
- Light Pollution Law Republic of Slovenia, 2007

In 2012, the government of South Korea has enacted and executed the “Light Pollution Prevention Act” (Cha et al. 2014). Other states and regions worldwide propose similar laws against light pollution. The positive result of the Lombardy Law was measured by Falchi (2011), who observed no increase in sky brightness in two studied sites in Lombardy despite an almost doubling in the outdoor-installed flux density over 12 years.

Switch the Light Off Whenever It Is Not Needed

All advice associated with light pollution recommends switching off artificial light at night, whenever it is not used (CIE 1997; ILE 2005; ILP 2011; IDA/IESNA 2011).

Today, curfew hours (defined by the authority when outdoor lighting has to be reduced or extinguished) are recommended, when artificial light is not required for safety, security, or enhancement of the nighttime scene. Also, seasonal changes demand different requirements on public lighting especially at mid- to high latitudes. During summer months the need for artificial lighting is reduced in evenings and mornings with long hours of daylight. During nights with higher temperatures, more artificial light is needed than during nights of the colder winter months, as more

people spend time outside at night in the summer. Opportunities for switching off or reducing the light for certain hours should be regulated using a needs-based approach.

However, when the illumination is completely switched off, no light is available for the few people that might need to use the street at late hours, and classic places of fear are created (Pauen-Höppner and Höppner 2013). With the reduction of light points, the illumination uniformity on streets is significantly lowered. Thus, obstacles, pedestrians, or cyclists may become hidden in shadows between the light points. A major technological innovation and alternative to curfew hours was made with the development of intelligent streetlight control systems, which are able to control the light output based on usage and occupancy. These systems also sense the velocity of movement and illuminate a certain number of streetlights ahead and fewer behind (Elejoste et al. 2013; Li and Zhang 2013; Juntunen et al. 2013). Further research on smart, flexible systems is required to adjust the light depending on road conditions. For example, snow or rain produces more reflectance; therefore, reduced light regimes are required to lower the glare. For many regions this technology is still not affordable. Hopefully, the development and rising demand of this technology will make these systems affordable for many communities in the near future.

Gaston et al. (2012) recommend switching lights off when significant nocturnal biological activity could be disturbed, such as when foraging, breeding, or dispersal/migratory activities are occurring. Many species require alternating light and darkness to maintain their circadian clock or as a signal for seasonal-dependent forms of behavior, including bud burst, flowering, dormancy, and leaf abscission in plants and reproduction, migration, and diapause in animals. Other species avoid light. A substantial proportion of global biodiversity is nocturnal (30 % of all vertebrates and >60 % of all invertebrates), and for these organisms, highly developed senses are required, often including specially adapted eyesight (Hölker et al. 2010a). Davies et al. (2012) found that streetlighting has an effect on the community composition of invertebrates. Predators such as harvestman (Opiliones), spiders, and ground beetles are affected by the light. Using presence detection sensors for park illumination would leave habitat dark for most of the night and might be a major step forward in the preservation of functional ecosystems. Areas that are species-rich and in the proximity of water, where many invertebrates emerge, especially need our consideration to be dedicated as natural darkness refuges (Perkin et al. 2011, 2014). Further research is urgently necessary to determine times and places for protection for light-sensitive species.

The protection or creation of naturally dark areas has received attention in the context of establishing International Dark Sky Places (www.darksky.org). But many other areas need protection from the increasing transformation of nightscapes. The removal of unnecessary lighting points could become a critical component for the conservation of protected habitats. Further, the proposal to consider the introduction of nighttime light pollution to an area should become a routine component of environmental impact assessments (Bruce-White and Shardlow 2011).

In France, a new light pollution law implemented in 2010 defined four objectives: biodiversity, energy consumption, discomfort, and preservation of the starry night. In 2012, a decree was issued to turn off shop signs and illuminated advertisements in the time between one and six in the morning. Since July 2013, interior lights of nonresidential buildings must be turned off an hour after the last person leaves. Lights in shop windows and building facades must be turned off by 1 a.m. at the latest. Due to this curfew decree, energy savings of 2 terawatt hours (equal to the consumption of 750,000 households) and a cut of carbon dioxide emission by 250,000 t per year are expected (Waks 2013).

Dim the Light and Choose the Most Appropriate Illuminants, Color Spectra, and Filters

Where and when lighting cannot be switched off due to human activities, the illumination level might be lowered. Moonlight in most cases is sufficient for clear visibility. In the absence of artificial light, full moonlight under clear skies gives an illumination of about 0.1–0.3 lx, a clear starry sky of about 0.001 lx, and an overcast night sky of 0.00003–0.0001 lx (Rich and Longcore 2006). Dick (2014) therefore demands the use of artificial light below 3 lx, except for roads with higher speed limits.

Without question, road lighting has significant safety benefits. Upgrading or improving streetlighting has resulted in a reduction of about 35 % in car crashes (Wanvik 2009; Elvik 1995). However, the safety augmentation due to lighting is greatly discussed due to unsound statistics. Combining statistical and analytical data, Bullough et al. (2013) present a maximal 13 % increase in safety due to improved lighting and argue that in former publications most weather conditions and safety influencing variables were not considered. Pauen-Höppner and Höppner (2013) demonstrate for Berlin that streets having a relatively low level compared to those with a rather high level of illumination showed no significant differences in accident loads. But the compensation of motorists' glare, i.e., the headlights of oncoming cars, is of higher importance for safety than the illumination of the traffic. Furthermore, in studies about the prevention of crime, the positive effects of artificial light are based on indications with biased statistical analyses (Marchant 2004, 2005, 2006). Today, no well-established dose–response relationship to lighting parameters exists, from which one can deduct thresholds of lighting for safety and security (Jackett and Frith 2013).

When determining the energy efficiency and efficacy of lighting products, appropriate measurements must take the visual effectiveness of the light produced into account. This is typically assessed using the photopic eye response function and expressed in terms of the total luminous flux in lumens per watt of electrical energy input.

Diverse organisms have sensitivities in different parts of the light spectrum. Humans see differently at different light levels. For photopic vision, which is used mainly during the day, at high light levels, the human eye uses cones, which provide

excellent color discrimination ability. During dark, moonless nights, scotopic vision is utilized; the eye then uses rods to process light. Colors are not distinguishable under these conditions. At most nighttime light levels and especially for visibility on roads, mesopic vision is used. The eye switches from cones to rods for vision at light levels from 0.001 to 3 cd/m² (Purkinje shift). The luminous efficacy in the mesopic range is significantly different from the efficacy in the photopic range. He et al. (1997) show that at 0.1 cd/m², a conventional metal halide lamp was 60 % more efficacious for mesopic, off-axis visual tasks than high-pressure sodium lamps with a 20 % higher lumen rating. Bisketzis et al. (2004) calculate uniformity and brightness for high-pressure sodium and metal halide lamps at photopic and mesopic conditions and indicate that unacceptable values for brightness and uniformity at photopic conditions can become acceptable at mesopic conditions. Findings in this field of research might fundamentally change the nighttime illumination of streets in the future.

The chromatic properties of light are determined by the light's spectral power distribution, which is the amount of energy the light contains at every wavelength in the visible spectrum. Many organisms have sensitivities in the UV range such as insects and fish (Perkin et al. 2011, 2014), e.g., moths actively congregate around light sources with a high amount of short wavelengths (Eisenbeis and Hänel 2009). UV light is invisible for humans. Higher vertebrates are mainly sensitive to wavelengths between 400 and 700 nm. The sensitivity to a light's spectrum differs with age. Brainard et al. (1997) demonstrate significant differences in lenticular transmission among age groups in the shorter wavelength ranges. The lenses from humans aged 50–59 years transmit significantly less blue (440 nm) and green (540 nm) light than the lenses from humans between the ages of 20–29 years. The widespread use of metal halide and light-emitting diodes (LEDs) with high proportion of blue light might therefore be counterproductive for the increasing proportion of elderly drivers.

However, today the most efficient LEDs are blue with a nominal wavelength ranging from 440 to 480 nm, which is far from natural solar radiation. In order to expand the blue spectrum into a broadband distribution, phosphorous material can be placed between the blue LED and the observer resulting in almost white light (Aubé et al. 2013). Preventing the blue component under 530 nm from reaching the eye by means of filters blocking wavelengths also preserves nocturnal melatonin production in humans and higher vertebrates (Kayumov et al. 2005). Gaston et al. (2012) recommend using filters incorporated into lighting design in order to improve the spectral quality of the light and thus the spectral composition of nighttime light pollution and reduce its negative environmental consequences. The use of filters to achieve warm-white light, however, lowers the efficiency of LEDs by about 30 %. A warm-white color temperature with an efficacy of about 82 lm/W was developed using fluoride and oxyfluoride phosphors (i.e., (Sr,Ca)₃(Al,Si)O₄(F,O):Ce³⁺ yellow-green phosphor and K₂TiF₆:Mn⁴⁺ red phosphor) (Setlur et al. 2010). Further research for the optimization of LED filters is urgently needed.

Falchi et al. (2011) recommend a total ban of the outdoor emission of light at wavelengths shorter than 540 nm to reduce the adverse health effects of decreased melatonin production and circadian rhythm disruption in humans and animals.



Fig. 4 Light dome over Osnabruck, Germany, and its luminance (cd/m^2) measured from about 20 km distance, at the observatory “Oldendorfer Berg.” Copyright Andreas Hänel

They refer to high-pressure sodium with relatively low emissions in this spectral range as a threshold. Lamps with a spectrum below 540 nm that emit energy flux larger ($\geq 15\%$) than that emitted by the standard high-pressure sodium lamps on a basis of equal photopic output should not be installed outdoors.

Another phenomenon of light pollution is skyglow (Fig. 4), which is caused by atmospheric scattering of light by small particles like aerosols and molecules (Kyba and Hölker 2013). As light moves through the atmosphere, most of the longer wavelengths pass straight through. However, much of the shorter wavelength light is absorbed by the gas molecules. The absorbed blue light is then radiated in different directions. It gets scattered all around the sky. This effect is therefore called the “blue sky” effect and due to Rayleigh scattering. The amount of Rayleigh scattering that occurs for a light beam depends upon the particle size and inversely of the wavelength, meaning that shorter wavelength radiation will scatter in the atmosphere more than longer wavelength radiation such as green and red light (Benenson et al. 2002). Aubé et al. (2013) prefer monochromatic light from low-pressure sodium lamps because this kind of spectral power distribution is easy to filter out for visible astronomy and its color is not very efficient in terms of atmospheric scattering and thus skyglow (Aubé et al. 2013). The illuminants that pollute the most under clear sky conditions are those with a strong blue emission, such as metal halide and white LEDs. Measurements of skyglow from blue-rich light sources differ significantly under photopic compared to scotopic conditions, i.e., at cloudless nighttime, particularly at larger distances from the light source (Luginbuhl et al. 2013). At scotopic vision conditions, the sky luminance by blue-rich light sources is much higher than that measured under photopic vision.

Luginbuhl et al. (2013) therefore see a higher threat for increased skyglow by the recent development to replace high-pressure sodium lighting with LEDs than previously assumed. Cloud coverage as well can dramatically amplify the luminance of the sky (Kyba et al. 2011). A sky brightness amplification factor of 10 was measured in urban areas, in Berlin, and of 2.8 for rural areas. But the brightness due to clouds increases more with long than with short wavelengths (Kyba et al. 2012).

Several systems for representing the colorimetric properties of light sources and surfaces have been developed, which reduce the full spectral information to a small set of numbers. The most common are CIE chromaticity coordinates (Fig. 5), correlated color temperature, color rendering indices, and various color appearance models developed by the CIE and other organizations (CIE 2004, 2007).

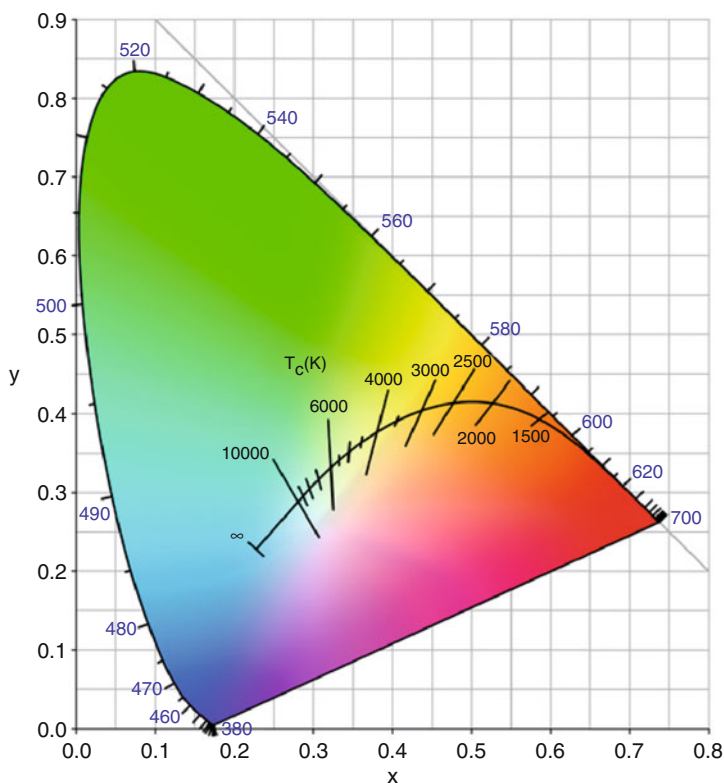


Fig. 5 The CIE color space chromaticity diagram with wavelengths in nanometers, developed in 1931, is still used today as the standard to define colors. The diagram is a two-dimensional display of colors with the same intensity (*brightness*), according to observations of color measurements by humans. A complete display of colors is actually three dimensional, where the z-axis displays the brightness. The *black line* within the diagram is the Planckian locus, the path that a *black body* color will take through the diagram as the *black body* temperature changes, and lines crossing the locus indicate constant correlated color temperature in Kelvin

Table 2 Level of impact of some commercial lamps adopted from indices and parameters for melatonin suppression, photosynthesis, and scotopic vision developed by Aubé et al. (2013)

Lamp type	Melatonin suppression	Photosynthesis	Scotopic vision
LED 5,000 K – filtered	+	++	+
LED 5,000 K	++	++	++
LED 4,000 K	++	++	++
LED 2,700 K -filtered	–	++	+
LED 2,700 K	+	++	++
HPS	+	++	–
LPS	–	+	+
Incandescent (3,000 K)	+	+++	++
Metal halide	++	++	++
Halogen	+	+++	++

Values – low impact (indices <0.1), + medium impact (indices >0.1 < 0.4), ++ high impact (indices >0.4 < 0.7), and +++ highest measured impact (>0.7)

In order to favor the design of new environmentally and health-friendly light devices, Aubé et al. (2013) developed three new parameters or indices for the measurement of spectral impact of lamps: melatonin suppression, photosynthesis, and scotopic vision (Aubé et al. 2013). The group found different potential impact indices on melatonin suppression, on photosynthesis, and on star visibility for each analyzed lamp. See Table 2 for an overview of these indices and a rough assessment of several commercial illuminants. Atmospheric scattering under clear or cloudy skies generally reduces the value of the indices in comparison to direct light, except for metal halide lamps for which the effect is opposite (Aubé et al. 2013). Unfiltered LEDs with a high color temperature (high-blue spectral power distribution) show a more than two times higher impact on melatonin suppression and scotopic vision compared to high- or low-pressure sodium lamps. The research group therefore recommends reducing newly installed LED luminous flux by a factor of 2.5 compared to the high-pressure sodium output.

The environmental impact of artificial light at night depends strongly on the illuminant and can be reduced by full cutoff luminaires and further by the ground under light points. Asphalt and concrete reflects less short wavelength radiation compared to long wavelength. Long wavelength sources such as low- or high pressure sodium lamps will have more light reflected by these surfaces compared to high-blue content lamps (Falchi et al. 2011). Light absorbent surfaces, such as grass or other vegetation, can minimize this effect.

Different species perceive color differently. Therefore, the choice of the illuminant's color should strongly depend on the surrounding habitat and occurring species. The broader the spectrum of light, the more organisms can be affected (Davies et al. 2012). As white light has the broadest spectrum, most species are affected by it. Research on color perception of humans as well as flora and fauna is imperative in order to understand how light has to change during the night to adapt to the nightly rhythm with the lowest impact on non-target organisms while providing enough visibility.

Last but not least, to calculate the ecological impact factor of a lamp, the lifetime and recycling possibilities need to be taken into account.

Shade the Light to Protect Your Neighbor

Some people might have problems with the prospect of a decree to close shades for private homes after a certain time. Methods to reduce light pollution by shading are not supposed to restrict residents, but to increase the awareness of commercial users of interior light or architects to consider detrimental effects of light emission at nighttime. For example, office departments or commercially used building floors can contribute particularly to light pollution with brightly lit windows, while not even presenting an attractive sight. It would only require a matter of shading lit office windows after dark, to reduce the contribution on light emission.

Sky lights are very attractive, because they allow indoors natural daylight. At nighttime, however, these windows can become unshielded lamps with 100 % obtrusive light emission. Globular sky lights especially turn unintentionally into great unshielded luminaires. The use of shades at nighttime could stop unintended light emission through windows easily, without eliminating the benefit of the daylight.

Supplemental lighting is necessary to induce flowering of some ornamentals in greenhouses in the tropics or for the extension of the vegetable growing season in temperate zones. These horticultural lighting applications cover large areas worldwide. High-intensity discharge (HID), high-pressure sodium, and metal halide lamps are the most common lighting products in horticulture. The intention of greenhouses is to let the greatest possible amount of sunlight pass through the glass during the daytime. On the other hand, artificial light at night can also pass to the outside and contribute to light pollution. In the Netherlands, huge areas are occupied by greenhouse productions. In order to limit the massive light emission, regional laws restrict the illumination from 6 p.m. to midnight in winter months and from 8 p.m. to 2 a.m. in summer months (Sabelis 2013). The objective is to reduce the light emissions of greenhouses by 100 % by 2018. Today, Dutch legislation requires an opaque screen that reduces light transmission of the greenhouse wall and roof by 95 % and supplementary light is limited to 15,000 lx, unless light emission is totally prevented (van't Ooster et al. 2008). The investigation of greenhouse shades yields additional benefits of energy saving, increased light intensity (van't Ooster et al. 2008), and climatic balancing (Albright et al. 2000). Further research is needed to measure the impact of decreased number of insects attracted to the greenhouse illumination at nighttime for insecticide treatments.

Learn from Nature

Enormous amounts of spiders accumulating at artificial lights at night do not only increase costs for cleaning the cobwebs but also are a sign of disturbed ecosystems.

In Hamburg, Germany, the disregard for the ecosystem creates immense economic costs. A new city center developed with modern architecture for offices and homes, with broad window fronts to the water, illuminated at nighttime, not only attracted realtors but also *Larinioides sclopetarius* – the bridge spider. This spider feeds on emerging insects from the water and is attracted by artificial light at night. The huge illuminated waterfront offers ideal conditions for the spider, to which it responds with shorter generation cycles and increased numbers of offspring (Kleinteich and Schneider 2011) – a great nuisance to the residents.

Evening and morning hours are the times with the highest demand for artificial outdoor light. These times unfortunately collide with the highest biological activity of many species. It is our responsibility to adapt outdoor lighting to the needs for the surrounding habitat.

Plants are highly sensitive to light. Trees, which are still fully or partly foliated in temperate regions in November or December, indicate delayed leaf fall (Fig. 6). This effect often happens when trees are directly illuminated by streetlights. The photosynthesis action spectrum, representing the efficiency of each wavelength in inducing photosynthesis in potential plants, shows two peaks: one in the blue region at around 450 nm and the other in the red part of the spectrum at around 660 nm. Lamps having a significant emission around these wavelengths bear a greater risk of interfering with photosynthesis, e.g., white LEDs with their blue wavelength peak at about 450 nm (Aubé et al. 2013). The photosynthesis indices for the choice of illumination should be considered in order to protect sensitive trees in urban areas



Fig. 6 Four horse chestnut trees (*Aesculus hippocastanum*) completely defoliated except in the light cone of a mercury vapor streetlight (Image taken in Berlin, Germany, end of November 2012, by Schroer)

(Cathey and Campbell 1975), because these trees are threatened by frost and subsequently many organisms that rely on the signal of leaf senescence for diapausing.

Further research is urgently needed to understand the cognition of artificial light for sensitive species. For example, it has been shown that avian collisions with communication towers tend to be much lower when using flashing rather than non-flashing (steady-burning) lights (Gehring et al. 2009).

Research is also needed for the sustainable design of outdoor light to develop luminaires that have a spectral power division comparable to natural light either at day- or at nighttime. The nightly outdoor illumination of the future might follow diurnal changes like copying sunset, the blue hour (the period of twilight each morning and evening), fire flames, and moonlight. Bioluminescence might be another great source for lighting designers to copy natural light with low lumen output.

The greatest step toward the protection of natural night sites, however, will be to learn and teach the value of darkness. In regions with low infrastructure, valuing natural dark areas could become economically important, by encouraging a recreation factor for tourism.

Conclusion and Directions for Future

Methods to reduce the detrimental effects of artificial light at night on nature can be inexpensive and easy to implement. Light should be directed to where it is needed and obtrusive light and glare prevented as much as possible. The light should be dimmed to the lowest level possible. The choice of the illuminant is significant and it should be harmonized with the needs of the surrounding habitat. Whenever great quantities of light are needed, for example, in greenhouses, using shades will protect invertebrates and birds and thus the surrounding ecosystems. We need to learn from nature and create areas of protected darkness. We will be rewarded with tremendous views of our home galaxy the Milky Way, massive amounts of energy savings, and healthy recreation.

So far, efforts to protect nature from light pollution have reduced the effects of or at least slowed the increase in light pollution (Falchi 2011; Duriscoe et al. 2014). Regional regulations have greatly helped in achieving this progress. Further research will significantly change the nighttime illumination for the benefit of ecosystems. Measurements of the improvements achieved by recent implementation of best practice standards will indicate the most appropriate measures and thus optimize future guidelines for the prevention of light pollution.

The choice of the right lighting product is of major importance for the environment. Today the predominant arguments for the decision on streetlight illumination are investment costs and energy efficiency. But the optimal choice is by far more complex. Short wavelengths of certain illuminants, e.g., metal halide or LEDs, can increase skyglow dramatically, might suppress melatonin, and thus disrupt circadian rhythms. Additionally, this blue-rich light increases the glare for elderly people and

thus decreases visibility. To exploit the energy efficiency of cold-white LEDs, appropriate filters are needed that change the spectral power distribution and create warm-white light. Further research might indicate ways to optimize spectral and energy efficiency requirements for LEDs.

Today, the efficiency of illuminants is measured using photopic vision. Much more knowledge is needed to create measurements for mesopic or scotopic vision, because these are the predominant human visions at nighttime. Gaining knowledge in the field of mesopic and scotopic vision might significantly change the luminous efficacy measure of many luminaires. Fundamental rethinking is needed on the required brightness of streetlighting. Only 1–3 lx at the surface is necessary to create visibility and sufficient for facial cognition. Therefore, the development of lighting products with low light output or the ability to be dimmed to 1–3 lx needs further consideration.

The development of intelligent streetlight systems is a major step forward to safer artificial light output at night. These systems have to be made affordable for communities and further developed to be able to react to weather conditions, seasonal changes, and biological activity.

The perception of artificial light and the effect of variable spectral power distribution need to be considered and adapted to the surrounding environment and sensitive species. We need a better understanding of at what times which species need protection and how to create adequate dark refuges.

In passing on the knowledge about the value of natural darkness, we create sustainable standards for recreation and thus our benefit, especially in areas where light installations are lacking today, like in many developing countries. Knowledge of the detrimental effects of artificial light and ways in which to avoid these problems need to be passed on with the lighting technology. This knowledge could help decision makers to avoid mistakes that have occurred in currently illuminated areas.

Acknowledgments We want to acknowledge the support by the European Cooperation in Science and Technology (COST) through the Action ES1204 LoNNe (Loss of the Night Network), and the national support by both the German Federal Ministry of Research and Technology (support code: 033L038A) and the Federal Agency for Nature Conservation (support code: 3514821700).

References

- Albright LD, Both AJ, Chiu AJ (2000) Controlling greenhouse light to a consistent daily integral. *Trans ASAE* 43(2):421–431
- Aubé M, Roby J, Kocifaj M (2013) Evaluating potential spectral impacts of various artificial lights on melatonin suppression, photosynthesis, and star visibility. *PLoS One* 8(7):e67798
- Aubrecht C, Stojan-Dolar M, De Sherbinin A, Jaiteh M, Longcore T, Elvidge C (2010) Lighting governance for protected areas and beyond—Identifying the urgent need for sustainable management of artificial light at night. *Earthzine*.
- Benenson W, Harris WJ, Stocker H, Lutz H (2002) *Handbook of physics*. Springer, New York, p 376
- Bisketzis N, Polymeropoulos G, Topalis FV (2004) A mesopic vision approach for a better design of road lighting. *WSEAS Trans Circuits Syst* 3(5):1380–1385

- Brainard GC, Rollag MD, Hanifin JP (1997) Photic regulation of melatonin in humans: ocular and neural signal transduction. *J Biol Rhythms* 12(6):537–546
- Bruce-White C, Shardlow M (2011) A review of the impact of artificial light on invertebrates. Buglife – The Invertebrate Conservation Trust, Peterborough
- Bullough JD, Donnell ET, Rea MS (2013) To illuminate or not to illuminate: roadway lighting as it affects traffic safety at intersections. *Accid Anal Prev* 53:65–77
- Cathey HM, Campbell LE (1975) Effectiveness of five vision-lighting sources on photo-regulation of 22 species of ornamental plants. *J Am Soc Hortic Sci* 100:65–71
- Cha JS, Lee JW, Lee WS, Jung JW, Le KM, Han JS, Gu JH (2014) Policy and status of light pollution management in Korea. *Light Res Technol* 46(1):78–88
- CIE (1997) Guidelines for minimizing sky glow. Publication No 126-1997. CIE, Vienna
- CIE (2004) Technical report: colorimetry, 3rd edn. Publication No 15-2004. CIE, Vienna
- CIE (2007) Technical report: road transport lighting in developing countries. Publication No 180-2007. CIE, Vienna
- Cinzano P (2002) Technical measures for an effective limitation of the effects of light pollution. In: Cinzano P (ed) Light pollution and the protection of the night environment. Proceedings of the IDA regional meeting “Venice: let’s save the night”. ISTIL, Thiene, pp 193–205
- Cinzano P, Falchi F, Elvidge CD (2001) The first world atlas of the artificial night sky brightness. *Mon Not R Astron Soc* 328(3):689–707
- Davies TW, Bennie J, Gaston KJ (2012) Street lighting changes the composition of invertebrate communities. *Biol Lett* 8(5):764–767
- Dick R (2014) Applied scotobiology in luminaire design. *Light Res Technol* 46:50–66
- Duriscoe DM, Luginbuhl CB, Elvidge CD (2014) The relation of outdoor lighting characteristics to sky glow from distant cities. *Light Res Technol* 46:35–49
- Eisenbeis G, Hänel A (2009) Light pollution and the impact of artificial night lighting on insects. In: Eisenbeis G, Hänel A (eds) Ecology of cities and towns. Cambridge University Press, Cambridge. Books online <http://dx.doi.org/10.1017/CBO9780511609763.016>
- Elejoste P, Angulo I, Perillos A, Chertudi A, Zuazola IJG, Moreno A, Azpilicueta L, Astrain JJ, Falcone F, Villadangos J (2013) An easy to deploy street light control system based on wireless communication and LED technology. *Sensors* 13(5):6492–6523
- Elvik R (1995) Meta-analysis of evaluations of public lighting as accident countermeasure. In: Transportation research record no 1485. Transportation Research Board/National Research Council, Washington, DC, pp 112–123
- Falchi F (2011) Campaign of sky brightness and extinction measurements using a portable CCD camera. *Mon Not R Astron Soc* 412:33–48
- Falchi F, Cinzano P, Elvidge CD, Keith DM, Haim A (2011) Limiting the impact of light pollution on human health, environment and stellar visibility. *J Environ Manage* 92(10):2714–2722
- Gaston KJ, Davies TW, Bennie J, Hopkins J (2012) Reducing the ecological consequences of night-time light pollution: options and developments. *J Appl Ecol* 49(6):1256–1266
- Gehring J, Kerlinger P, Manville AM II (2009) Communication towers, lights, and birds: successful methods of reducing the frequency of avian collisions. *Ecol Appl* 19:505–514
- He Y, Rea M, Bierman A, Bullough J (1997) Evaluating light source efficacy under mesopic conditions using reaction times. *J Illum Eng Soc* 26:125–138
- Hölker F (2013) Lichtverschmutzung und die Folgen für Ökosysteme und Biodiversität. In: Held M, Hölker F, Jessel B (eds) Schutz der Nacht – Lichtverschmutzung, Biodiversität und Nachtlandschaft. BfN-Skripten 336, Bonn. pp 73–76
- Hölker F, Moss T, Griefahn B, Kloas W, Voigt CC, Henckel D, Hänel A, Kappeler PM, Völker S, Schwöpe A, Franke S, Uhrlandt D, Fischer J, Klenke R, Wolter C, Tockner K (2010a) The dark side of light – a transdisciplinary research agenda for light pollution policy. *Ecol Soc* 15(4), 13
- Hölker F, Wolter C, Perkin EK, Tockner K (2010b) Light pollution as a biodiversity threat. *Trends Ecol Evol* 25:681–682
- IDA/IESNA (2011) Model Lighting Ordinance (MLO) with users guide. <http://www.darksky.org/outdoorlighting/mlo>

- ILE (2005) Guidance notes for the reduction of obtrusive light. <http://www.britastro.org/dark-skies/pdfs/ile.pdf>
- ILP (2011) Guidance for the reduction of obtrusive light, GN01. <https://www.theilp.org.uk/documents/obtrusive-light/>
- Jackett M, Frith W (2013) Quantifying the impact of road lighting on road safety – A New Zealand Study. *IATSS Res* 36:139–145. <http://dx.doi.org/10.1016/j.iatssr.2012.09.001>
- Juntunen E, Tetri E, Tapaninen O, Yrjänä S, Kondratyev V, Sitomaniemi A, Siirtola H, Sarjanoja EM, Aikio J, Heikkinen V (2013) A smart LED luminaire for energy savings in pedestrian road lighting. *Light Res Technol*. DOI: 10.1177/1477153513510015
- Kayumov L, Casper RF, Hawa RJ, Perelman P, Chung SH, Sokalsky S, Shapiro CM (2005) Blocking low-wavelength light prevents nocturnal melatonin suppression with no adverse effect on performance during simulated shift work. *J Clin Endocrinol Metabol* 90(5): 2755–2761
- Kleinteich A, Schneider JM (2011) Developmental strategies in an invasive spider: constraints and plasticity. *Ecol Entomol* 36(1):82–93
- Kuechly HU, Kyba CCM, Ruhtz T, Lindemann C, Wolter C, Fischer J, Hölker F (2012) Aerial survey and spatial analysis of sources of light pollution in Berlin, Germany. *Remote Sens Environ* 126:39–50
- Kyba CCM, Hölker F (2013) Do artificially illuminated skies affect biodiversity in nocturnal landscapes? *Landsc Ecol* 28:1637–1640
- Kyba CCM, Ruhtz T, Fischer J, Hölker F (2011) Cloud coverage acts as an amplifier for ecological light pollution in urban ecosystems. *PLoS One* 6(3):e17307
- Kyba CCM, Ruhtz T, Fischer J, Hölker F (2012) Red is the new black: how the colour of urban skyglow varies with cloud cover. *Mon Not R Astron Soc* 425(1):701–708
- Li XY, Zhang WQ (2013) Design of intelligent street light control system based on ZigBee. *Appl Mech Mater* 299:71–74
- Longcore T, Rich C (2006) In: Synthesis Rich C, Longcore T (eds) *Ecological consequences of artificial night lighting*. Island Press, Washington, DC. pp 413–430
- Luginbuhl CB, Boley PA, Davis DR (2013) The impact of light source spectral power distribution on sky glow. *J Quant Spectrosc Radiat Transf*. <http://dx.doi.org/10.1016/j.jqsrt.2013.12.004i>
- Marchant PR (2004) A demonstration that the claim that brighter lighting reduces crime is unfounded. *Br J Criminol* 44:441–447. <http://bjc.oupjournals.org/cgi/content/abstract/44/3/441>
- Marchant PR (2005) What works? A critical note on the evaluation of crime reduction initiatives. *Crime Prev Community Saf Int J* 7(2):7–13
- Marchant PR (2006) Shining a light on evidence based policy: street lighting and crime. *Crim Justice Matters* 62(1):18–45
- Narisada K, Schreuder D (2004) *Light pollution handbook*, vol 322, Astrophysics and space science library. Springer, Dordrecht
- National Park Service (2012) *Natural sounds and night skies. Managing Lightscapes*. <http://www.nature.nps.gov/night/management.cfm>
- Navara KJ, Nelson RJ (2007) The dark side of light at night: physiological, epidemiological, and ecological consequences. *J Pineal Res* 43:215–224
- Parks B (2013) Ecological responsible outdoor lighting guidelines. In: Krop-Benesch A, Kyba CCM, Hölker F (eds) *ALAN – first international conference on artificial light at night*, Berlin, 28–30 Oct, p 95
- Pauen-Höppner U, Höppner M (2013) Öffentliche Beleuchtung – mehr Licht heißt nicht mehr Sicherheit. In: Held M, Hölker F, Jessel B (eds) *Schutz der Nacht – Lichtverschmutzung, Biodiversität und Nachtlandschaft*. BfN-Skripten 336, Bonn. pp 10–108
- Perkin EK, Hölker F, Richardson JS, Sadler JP, Wolter C, Tockner K (2011) The influence of artificial light on stream and riparian ecosystems: questions, challenges, and perspectives. *Ecosphere* 2(11):122. doi:10.1890/ES11-00241.1
- Perkin EK, Hölker F, Tockner K (2014) The effects of artificial lighting on adult aquatic and terrestrial insects. *Freshw Biol* 59(2):368–377

- Sabelis I (2013) Lichtverschmutzung durch Gewächshäuser in den Niederlanden. In: Held M, Hölker F, Jessel B (eds) Schutz der Nacht – Lichtverschmutzung, Biodiversität und Nachtlandschaft. BfN (Federal Agency for Nature Conservation)-Skripten 336, Bonn. pp 181–184
- Setlur AA, Radkov EV, Henderson CS, Her JH, Srivastava AM, Karkada N, Kishore MS, Kumar NP, Aesram D, Deshpande A, Kolodin B, Grigorov LS, Happek U (2010) Energy-efficient, high-color-rendering LED lamps using oxyfluoride and fluoride phosphors. *Chem Mater* 22 (13):4076–4082
- van't Ooster A, van Henten EJ, Janssen EGON, Bongaerts E (2008) Use of supplementary lighting top screens and effects on greenhouse climate and return on investment. *Acta Horticult* 801:645–652
- Waks L (2013) Regulation of light pollution in France. In: Krop-Benesch A, Kyba CCM, Hölker F (eds) ALAN – first international conference on artificial light at night, Berlin, 28–30 Oct, pp 181–184
- Wanvik A (2009) Effects of road lighting: an analysis based on Dutch accident statistics 1987–2006. *Accid Anal Prev* 41:123–128
- WHO (2014) World Health Organization. International travel and health. http://www.who.int/ith/other_health_risks/injuries_violence/en/index.html

Further Reading

- Bennie J, Davies TW, Duffy JP, Inger R, Gaston KJ (2014) Contrasting trends in light pollution across Europe based on satellite observed night time lights. *Sci Rep* 4:3789
- CIE (2003) Guide on the limitation of the effect of obtrusive light from outdoor lighting installations. Publication No 150. CIE, Vienna
- Longcore T, Rich C (2004) Ecological light pollution. *Front Ecol Environ* 2(4):191–198
- Mizon B (2002) Light pollution: responses and remedies. Springer, London
- Posch T, Hölker F, Uhlmann T, Freyhoff A (2014) Das Ende der Nacht: Lichtsmog: Gefahren-Perspektiven-Lösungen. Wiley-VCH, Weinheim

Part VIII

Conventional Light Sources

Incandescent Lamps

Maxime F. Gendre

Contents

Introduction	1015
Fundamental Mechanism and Properties of Incandescence	1015
Incandescence and Energy Levels in the Solid State	1016
The Blackbody	1017
Radiative Properties of Actual Incandescent Materials	1018
Physics and Technology of Tungsten Filament Lamps	1023
Main Limiting Factor in the Operation of Incandescent Lamps	1023
Gas Fill	1024
Filament Coiling and Thermal Management	1027
Overall Optical Emission and Energy Balance	1028
Tungsten Metallurgy and Filament Production	1030
Standard Incandescent Lamps for General Lighting Applications	1033
General Structure and Design	1033
Standard Lamps for General Lighting Service	1034
Standard Reflector Lamps	1036
Tungsten Halogen Lamps	1037
The Halogen Cycle	1038
General Structure and Design	1039
Tungsten Halogen Burners for General Lighting	1040
Jacketed Tungsten Halogen Lamps for General Lighting	1042
Reflector Tungsten Halogen Lamps for General Lighting	1043
Incandescent Lamps for Special Applications	1045
Standard Lamps for Vibration Service and Traffic Signals	1045
Automotive Lamps	1047
Stage and Studio Lamps	1047
Infrared Radiators	1049
Calibration and Reference Lamps	1051

M.F. Gendre (✉)
Helmond, The Netherlands
e-mail: mfgendre@lampreview.net

Further Improvements of the Incandescent Lamp Technology	1053
Infrared Conservation	1053
High-Pressure Xenon Incandescent Lamps	1056
Life Expectancy and Failure Mechanisms	1057
Balance Between Efficacy and Service Life	1057
Normal Failure Mechanisms	1057
Survival Rate	1059
Anomalous and Nonpassive Failure Mechanisms	1060
Alternative Developments and Conclusion	1061
References	1062

Abstract

Discovered in 1802 by H. Davy, the phenomenon of incandescence is the oldest practical mean of light generation from electricity. In this process, optical emission arises from the constant energy change of electrons in hot solid materials, resulting in a continuous electromagnetic spectrum with a temperature-dependent Planckian wavelength distribution. Incandescence is implemented in lamps by driving an electric current through a thin filament made of tungsten, a refractory metal chosen for its high melting point (3695 K) and low vapor pressure (1 Pa at 3477 K). In order to limit thermal losses and material evaporation, lamps are in most cases filled with a protective gaseous atmosphere, and the tungsten wire is wound into a compact coil or coiled coil configuration. In order to ensure a stable filament structure at high temperature, the metal is doped with potassium or rhenium so as to promote the most favorable crystallographic structure.

When operated in a neutral Ar-N₂ or Kr-N₂ atmosphere, the filament temperature lies in the 2600–2800 K range, resulting in 5–8 % energy conversion into visible light. Better performances are obtained with a higher gas fill pressure combined with a tungsten-bromine cycle which prevents tungsten deposition onto the lamp wall. Due to a higher bulb temperature requirement (800–1000 K), halogen lamps are made with a refractory glass bulb in a very compact configuration. With a fill pressure reaching 3 bars, filaments can be operated in the 2800–3200 K range, resulting in 7–13 % energy efficiency.

Lamps for general lighting applications are both of the standard and halogen types and are made for an isotropic or a directed emission of light. Standard gas-filled lamps feature a 3.5–20 lm W⁻¹ efficacy with a 1000 h average service life, optimized for the most economical lamp usage. Standard halogen lamps have a 9–25 lm W⁻¹ efficacy with a mean service life reaching up to 10,000 h. Incandescent lamps are also made in a wide variety of configurations with different filament structures and temperatures so as to address specific lighting needs in traffic signals, automotive applications, on stages and studios, for infrared processing, and for instrument calibration. Finally, the most recent and refined lamp designs integrate an infrared mirror for energy conservation, resulting in compact general lighting sources with up to 35 lm W⁻¹ efficacy, or feature a novel wafer-sealed bulb construction and a 5 bar xenon fill yielding 18.8 lm W⁻¹ in a compact low-wattage package for automotive applications.

However, incandescent lamps are plagued by two intrinsic limitations, the first of which is a lumen efficacy constrained by the nature of the light emission mechanism and by the maximum filament temperature permitted by technology. The latter constitutes the second intrinsic limitation as the filament life is mostly limited by the formation and growth of local hot spots on the tungsten wire as a result of material evaporation and diffusion. These two limitations result in a relatively poor life-efficacy balance compared to other light source technologies. For this reason, incandescent lamps are being progressively phased out in a number of lighting applications as more efficient technological alternatives emerge.

Introduction

Discovered over two centuries ago, incandescence is the oldest form of controlled electrical energy conversion process into light. First demonstrated in 1802 by H. Davy, this phenomenon was successfully implemented in practical light sources simultaneously by T.A. Edison and J.W. Swan in 1879 (Bright 1949). In the course of their 135 years of market presence, incandescent lamps have been used in every lighting applications, and although this technology is no longer dominant in the present market, these still have the widest range of product design variations and are still light sources of choice in many areas. An intrinsically simple construction and operation associated with many favorable light technical properties are the prime reasons for the durable presence of this technology in a constantly changing lighting market.

The science and technology of incandescent light sources are treated in detail in this chapter, with the fundamental mechanism and characteristics of incandescence covered in the first part, followed by the physics and technology of filament lamps. The next three sections deal with the characteristics of practical lamps designed for general and special lighting applications. The subject of the tungsten-halogen cycle and its implementation in lamps is also covered. Section “[Further Improvements of the Incandescent Lamp Technology](#)” presents two of the most significant and recent improvements in the technology. Finally, the aspect of lamp service life and failure is treated in the last section.

Fundamental Mechanism and Properties of Incandescence

The phenomenon of incandescence is at the heart of the light sources considered here, and its properties have a large influence on the design and characteristics of practical lamps. This section deals with the fundamental processes underlying this type of light emission and covers the factors defining and influencing the optical properties of materials.

Incandescence and Energy Levels in the Solid State

The mechanism of incandescence is an energy conversion process occurring in solid-state bodies, which results in the emission of electromagnetic energy as a consequence of heating. This light emission arises from the continuous energy changes within the population of free and bound (valence) electrons present in the material and subjected to thermal motion. Although energy levels of individual atoms are of discrete nature, solid-state materials are characterized by energy bands arising from Pauli's exclusion that forces individual energy levels of valence electrons to displace and split when atoms are brought close together, as shown in Fig. 1 (Hummel 1993). The vast multiplicity of energy states allows a multitude of electron energy transitions which then result in a continuous light emission spectrum.

The energy distribution of the emitted photons depends on the characteristics of the particle interactions within the material, and the nature and energy range of the energy exchange processes between ions, electrons, and photons. The shape of the light emission spectrum thus depends on the energy band structure of the material and on how fully reversible these processes are, which are functions of the nature and structure of the incandescent radiator.

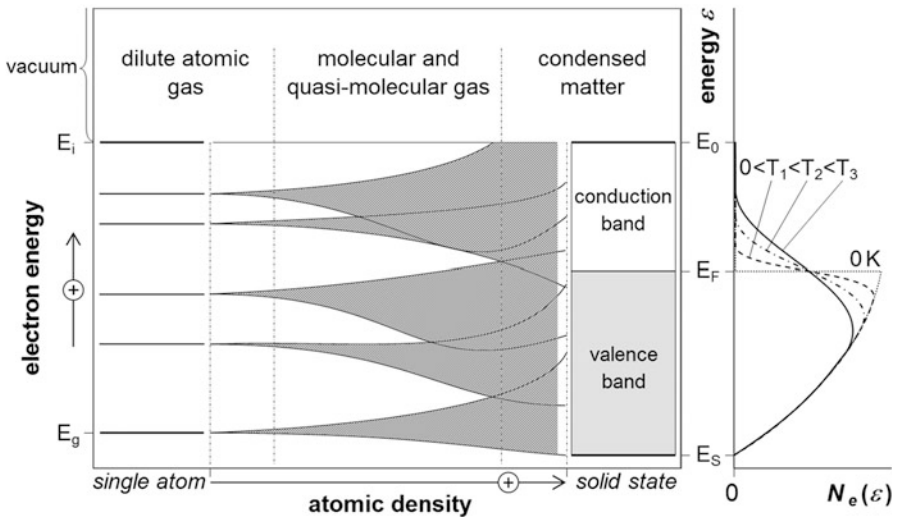


Fig. 1 Schematic representation of the evolution and transformation of the energy levels of metal atoms as a function of density, leading to an energy continuum in the solid state. E_s , E_F , and E_0 are the minimum electron rest energy, the Fermi energy level, and the electron energy at vacuum level, respectively. The electron energy distribution $N_e(\epsilon)$, shown at right for different body temperatures, is defined as the product of the Fermi-Dirac distribution by the energy state density distribution of the material

The Blackbody

The blackbody is an ideally absorbing medium where all energetic exchange processes occurring at microscopic scale are fully reversible over an infinite energy range. The resulting emission spectrum has a characteristic distribution which is described by Planck’s law of blackbody radiation (Vukceвич 1992), derived in 1900 from the first formulated hypothesis of the quantum nature of light interactions with matter:

$$B_{bb}(T, \lambda) = \frac{2hc^2}{\lambda^5} \cdot \left(\exp\left(\frac{hc}{\lambda k_b T}\right) - 1 \right)^{-1} \tag{1}$$

where B_{bb} is the spectral radiance (in $J s^{-1} m^{-3} sr^{-1}$), h is Planck’s constant, k_b is Boltzmann’s constant, and c is the velocity of light in vacuum. The variables λ and T are the wavelength (in [m]) and the body temperature (in [K]), respectively. Planck’s distributions plotted in Fig. 2 for different body temperatures are similar to the spectral distribution of light emitted by actual incandescent lamps.

Since the energy distribution of photons depends directly on the thermal energy distribution of material particles they interact with, the mean energy of the radiated light thus increases at higher body temperatures. There are two consequences to this causality, the first of which is a quartic proportionality between the total radiated power density $\partial P_{rad} / \partial A$ and temperature, a relation described by the Stefan-Boltzmann law which is derived by integrating Eq. 1 over the wavelength and solid angle Ω :

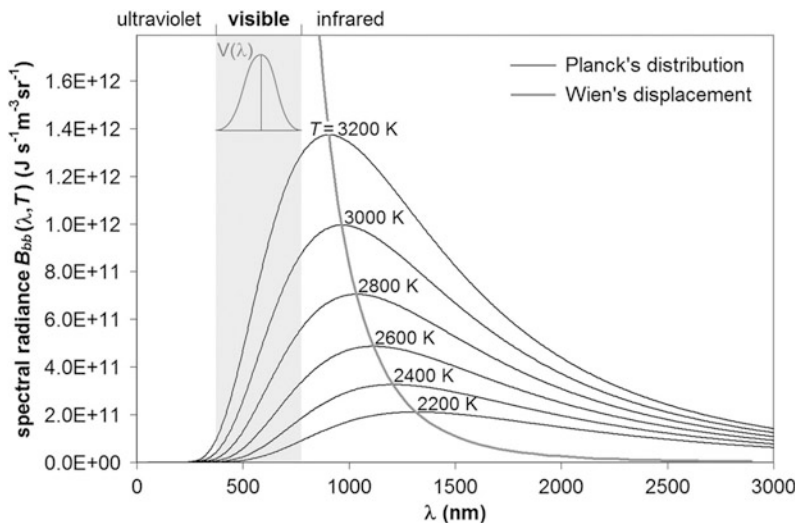


Fig. 2 Evolution of the wavelength at peak spectral emission and of the spectral radiance distribution of a blackbody as a function of its temperature. The visible domain is shown in the shaded area with the eye’s photopic sensitivity curve $V(\lambda)$

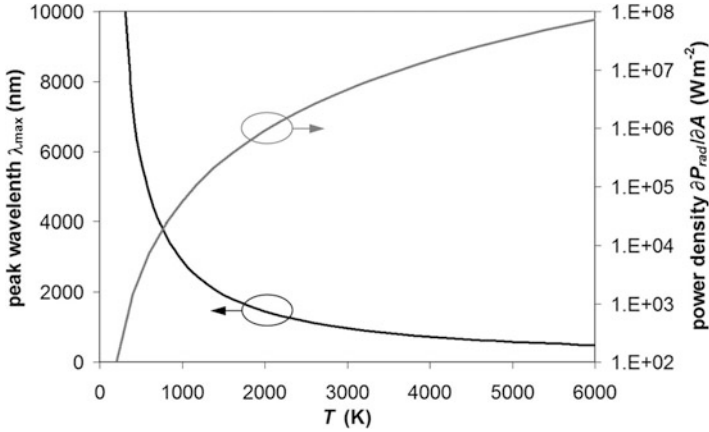


Fig. 3 Evolution of the wavelength at peak spectral emission and of the total radiated power flux density of a blackbody as a function of its temperature

$$\frac{\partial P_{\text{rad}}}{\partial A} = \frac{1}{2} \int_0^{2\pi} \int_0^{\infty} B_{bb}(T, \lambda) \, d\lambda \, d\Omega = \frac{2\pi^5 k_b^4 T^4}{15 c^2 h^3} = \sigma_{sb} T^4 \quad (2)$$

The total optical power output is P_{rad} in Eq. 2, while ∂A is the unit surface area and σ_{sb} is Stefan's constant ($5.670 \cdot 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$). The second consequence is a displacement of the emission peak from the radio wave to the X-ray domain as the body temperature increases. This behavior, shown in Fig. 2, is described by Wien's displacement law which is the nontrivial wavelength solution λ_{max} to the condition $\partial B_{bb}/\partial \lambda = 0$ at any temperature, excluding $\lambda = 0$ and ∞ for obvious reasons (Vukcevic 1992):

$$\lambda_{\text{max}} = \frac{hc}{5 k_b T} = \frac{K_W}{T} \quad (3)$$

where K_W is Wien's constant, equal to $2.877 \cdot 10^{-3} \text{ m K}$. The evolutions of λ_{max} and $\partial P_{\text{rad}}/\partial A$ as a function of temperature are presented in the graph of Fig. 3.

Radiative Properties of Actual Incandescent Materials

Electronic Properties and Spectral Emissivity

In metals, the combination of a high charge density with overlapping valence and conduction bands results in a very high interaction frequency between ions, electrons, and photons, whose related energy exchange processes are balanced over a wide energy range (Hummel 1993; Mandel and Wolf 1995). The system is therefore in local thermodynamic equilibrium, and the characteristics of the radiation in the metal bulk are properly described by Planck's theory. However, close to the

material surface, within the skin depth at optical frequency, the thermal equilibrium is broken due to the net escape of photons. The local radiation properties then depart from those of Planck's blackbody (Waymouth 1987).

The spectral radiance distribution B_m which results from this situation is related to that of a blackbody by an emissivity coefficient ε_{em} such that $B_m(T, \lambda, \theta) = B_{bb}(T, \lambda, \theta) \cdot \varepsilon_{em}(T, \lambda, \theta)$ (Vukcevic 1992). This factor is a function of temperature T , of wavelength λ , and of the emission angle θ . Kirchhoff's radiation law relates the emissivity to the optical properties of the radiating body and more particularly to its absorbance A (i.e., the fraction of absorbed radiation) such that $\varepsilon_{em}(T, \lambda, \theta) = A(T, \lambda, \theta) = 1 - R(T, \lambda, \theta)$, where R is the material spectral reflectance (Kauer 1965). Since the emissivity depends very little on the emission angle, this variable is usually ignored in practical calculations.

There exist two types of thermal radiators which have different emissivity characteristics: gray bodies, with a constant emissivity, and selective emitters, whose emissivity is a function of wavelength (Kauer 1965). In both cases ε_{em} is below unity and it results that the total radiated power density is lower than that from a blackbody. To account for this change, the Stefan-Boltzmann law expressed in Eq. 2 needs to be rewritten by introducing the average value $\langle \varepsilon_{em} \rangle$ of the material emissivity:

$$\frac{\partial P_{\text{rad}}}{\partial A} = \langle \varepsilon_{em} \rangle \cdot \sigma_{sb} T^4 \quad (4)$$

The light emission properties of incandescent materials thus depend on their optical characteristics and more particularly on the spectral reflectance of the solid-vacuum surface interface (Waymouth 1989). This characteristic is a function of the material refractive index n and optical absorption coefficient α (Kauer 1965; Köstlin and Frank 1983–1984; Hummel 1993):

$$R = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2} \quad (5)$$

where k is the extinction coefficient defined as $k = \alpha \cdot \lambda / (4\pi)$. Both absorption coefficient and refractive index are functions of wavelength and temperature. The refractive index and the extinction coefficient are related by the normalized dispersion equations:

$$n^2 - k^2 = \varepsilon_{bc} \cdot \left(1 - \frac{1 + (\gamma/\omega_p)^2}{(\omega/\omega_p)^2 + (\gamma/\omega_p)^2} \right) \quad (6)$$

$$2nk = \varepsilon_{bc} \cdot \frac{(\gamma/\omega_p) \cdot (1 + (\gamma/\omega_p)^2)}{(\omega/\omega_p) \cdot ((\omega/\omega_p)^2 + (\gamma/\omega_p)^2)} \quad (7)$$

with ε_{bc} a relative permittivity constant accounting for the polarizability of bound charges in the material. The electronic properties are characterized by n_e and m_e^* which are the density and effective mass of the free electrons, respectively, while q_e is the unit charge. γ is the electron oscillation damping factor defined as $\gamma = q_e / (\mu_e m_e^*)$ with μ_e the electron mobility in the material. Finally, the angular frequency ω is defined such that $\omega = 2\pi f$, with f the oscillation frequency, while ω_p is the characteristic electron oscillation frequency corresponding to the limit at $n^2 - k^2 = 0$:

$$\omega_p = \sqrt{\frac{n_e q_e^2}{\varepsilon_0 \varepsilon_{bc} m_e^*}} - \gamma^2 \quad (8)$$

where ε_0 is the absolute dielectric permittivity of vacuum. ω_p is a constant which defines the optical properties of the material relative to the impinging electromagnetic radiation of frequency ω . In situations where $\omega > \omega_p$, the free electrons cannot respond to the applied alternating electric field, and the material has an essentially dielectric optical behavior. In the opposite case of $\omega < \omega_p$, the free electrons respond to the oscillating electric field, and the absorption and reflection of the incident electromagnetic wave ensue. In this case the material has metal-like optical properties. The part of the absorbed electromagnetic energy which is dissipated as heat in the material is defined by the damping constant γ .

Due to the very high charge density in metals, their electron plasma frequency $\omega_p/2\pi$ lies in the X-UV domain, and the most important factor defining the material reflectivity (and thus its spectral emissivity) in the visible and infrared domains is the (γ/ω_p) ratio, i.e., the interaction of the free electrons with the surrounding ionic lattice.

Overall Efficacy of Incandescent Sources

A common measure of light output is the lumen, which is the value of the emitted spectral power distribution integrated over the photopic lumen sensitivity function of the human eye $V(\lambda)$ (as defined by CIE 1931). It thus results that the perceived light output from an incandescent lamp depends on the temperature and emissivity of the thermal radiator. Moreover, the overall lumen efficacy η_{lm} of practical light sources is also influenced by non-radiative losses resulting from heat conduction (P_{cd}) and convection (P_{cv}) (Kauer 1965):

$$\eta_{lm}(T) = \frac{\int_0^\infty B_{bb}(T, \lambda) \varepsilon_{em}(T, \lambda) V(\lambda) d\lambda}{\int_0^\infty B_{bb}(T, \lambda) \varepsilon_{em}(T, \lambda) d\lambda + P_{cd} + P_{cv}} \quad (9)$$

Equation 9 shows that there are three ways to increase the lamp efficacy: increasing the radiator temperature, reducing the heat losses, and using a selective emitter having a higher emissivity in the visible than in the infrared.

The maximum operating temperature and the spectral emissivity properties of practical radiators depend on the nature of the material used, while the heat losses are affected by both lamp and radiator designs, as detailed in the next chapter. As far as the blackbody is concerned, the maximum efficacy is 95 lm W^{-1} , reached at a temperature around 6000 K and for zero P_{cd} and P_{cv} values (Kauer 1965). However, since there is no electrically conducting material with a melting point above 4000 K, the performances of practical lamps are thus limited by the reduced maximum operating temperature of available materials.

Characteristics of Tungsten as a Source of Thermal Radiation

Thermal radiators implemented in practical lamps are made of tungsten because of its favorable thermomechanical properties, and the metal’s optical characteristics have been the subject of extensive investigations (Langmuir 1916a; Elenbaas 1972; Vukceвич 1992), whose main results are presented here.

Tungsten is characterized by a 55 % optical reflectance in the visible which, following Kirchhoff’s radiation law, results in an average optical emissivity limited to 0.45 in this spectral range. With a $\langle \epsilon_{em} \rangle$ value decreasing to around 0.35 in the near infrared and to 0.1–0.2 toward 10 μm (Fig. 4), tungsten is only a mildly selective thermal radiator. This particular characteristic is caused by the combination of a high charge carrier density and a strong electron oscillation damping factor (Kauer 1965; Waymouth 1987).

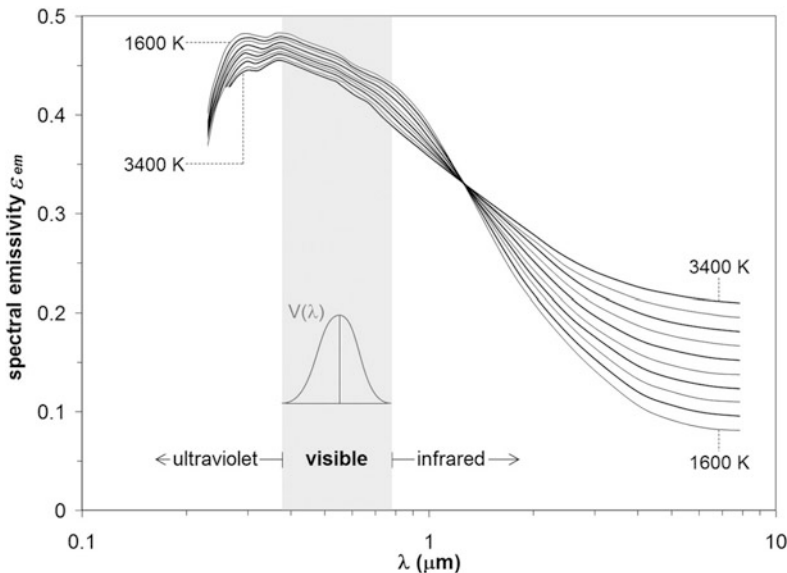


Fig. 4 Spectral emissivity distribution of tungsten as a function of temperature between 1600 and 3400 K, given in 200 K intervals (Data from Elenbaas 1972). The visible domain is shown in gray with the eye’s photopic sensitivity curve $V(\lambda)$ superimposed to the area

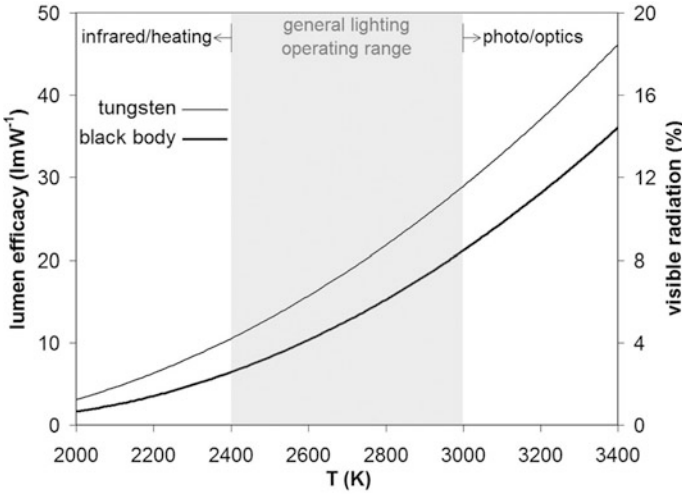


Fig. 5 Lumen efficacy and percentage of radiated visible light as a function of temperature for a blackbody and for a flat tungsten radiator in vacuum. The typical operating temperature ranges are shown for the three main classes of incandescent lamps

The temperature evolution of the emissivity depends on how the electron properties change with the temperature. The flattening of the spectral emissivity distribution at higher temperature observed in Fig. 4 arises primarily from an increased γ value in Eqs. 6 and 7 caused by the stronger and more frequent interactions between free electrons and the crystal lattice (Kauer 1965). The evolution of the metal emissivity as a function of temperature was determined empirically in the 200–1270 nm range (Vukceвич 1992):

$$\varepsilon_{em}(\lambda, T) = K_a(\lambda) - K_b(\lambda, T) \cdot \frac{T - 1600}{1000} \tag{10}$$

where T is expressed in kelvin and K_a and K_b are wavelength- and temperature-dependent factors whose values are given in Vukceвич’s book. In the case of tungsten, the average emissivity $\langle \varepsilon_{em} \rangle$ is well approximated by the following expression (Levin 1966):

$$\langle \varepsilon_{em} \rangle = 0.4756 - 0.2073 \cdot \frac{T}{10^4} \tag{11}$$

Given the linear relationship in Eq. 11, the Stefan-Boltzmann law expressed in Eq. 4 then becomes proportional to T^5 for tungsten. This increase in temperature slope characteristic can be also observed in Fig. 5 in the case of the lumen efficacy and the fraction of visible light emitted from a flat tungsten radiator compared to a blackbody.

In most applications the tungsten radiator is operated between 2400 and 3000 K, defined mostly by life-efficiency trade-off considerations. In this case about 4–12 % of the input power is converted into visible light, as shown in Fig. 5. The selective emission properties of tungsten result in a lumen efficacy which is 37–63 % higher than that from a blackbody radiator. However, as it will be shown in the next chapter, these figures are lower in practical light sources. In lamps intended for special applications, the operating temperature can be as high as 3400 K or as low as 1800 K depending on the intended use.

Physics and Technology of Tungsten Filament Lamps

In practical lamps, the emission of light by incandescence is achieved by bringing a refractory metal filament to a very high temperature via ohmic heating. Tungsten is used as the incandescent radiator material because of its high melting point of 3695 K and low vapor pressure at high temperature (1 Pa at 3477 K) (van den Hoek 2010). However, because this metal has a low electrical resistivity ($85 \cdot 10^{-6} \Omega \text{ cm}$ at 2800 K), filament wires have to be made with a very small diameter and used in long sections, especially in mains-voltage applications. In common 230 V–60 W household lamps, this wire has a 35 μm diameter for a 1 m length and is formed into a few centimeter-long filament structures.

The efficient and stable operation of incandescent lamps depends directly on the delicate balance between the electrical, optical, thermal, and mechanical properties of filaments. These properties and the operating conditions must be optimal and tightly controlled over time in order to obtain the desired lamp characteristics. This section provides details on all critical aspects of filament operation requirements and characteristics.

Main Limiting Factor in the Operation of Incandescent Lamps

Material evaporation and other processes described in section “[Normal Failure Mechanisms](#)” limit the practical operating temperature of filaments to a level well below that of the melting point of tungsten. In his study of filament failures, E.J. Covington showed that lamp life is inversely proportional to the initial evaporation rate of the filament. He also found that the relation between the tungsten evaporation rate Γ_{We} and temperature T is well approximated by the Arrhenius equation (Covington 1973a):

$$\Gamma_{\text{We}} = K_{\text{We}} \cdot \exp\left(\frac{-E_{\text{We}}}{R_{\text{g}} \cdot T}\right) \quad (12)$$

where K_{We} is a constant, E_{We} is the activation energy, and R_{g} is the universal gas constant. In his study of tungsten evaporation at 2380–3330 K in vacuum, Covington found $K_{\text{We}} = 1.869 \cdot 10^9 \text{ g cm}^{-2} \text{ s}^{-1}$ and $E_{\text{We}} = 875 \text{ kJ mol}^{-1}$. It thus results that

Γ_{we} doubles every 1.7–2.2 % increase in temperature. This characteristic has a strong influence on the durability of incandescent filaments because lamp life L is inversely proportional to the tungsten evaporation rate (Vukcevic 1992):

$$L = K_L \cdot \frac{D_F}{\Gamma_{we}(T)} \quad (13)$$

with D_F the filament wire diameter and K_L a constant whose value depends on the wire diameter (larger wires result in a higher coefficient) and on the filament environment. The value of K_L ranges from 0.56 to 0.99 g cm⁻³ in the case of vacuum lamps between 6 and 25 W, respectively (Vukcevic 1992). From the data provided, it is derived that L decreases by half every 1.9 % increase in filament temperature. For this reason practical thermal radiators are operated between 2400 and 3400 K in order to provide a life-efficiency balance which best matches the intended applications in visible lighting (Valin and Magnien 1991). However, the need for higher filament efficiencies led to measures aimed at reducing the tungsten evaporation rate so as to enable higher operating temperatures without affecting the lamp life. The first of such measures consists in surrounding the filament with a protective gaseous atmosphere.

Gas Fill

Requirements

The first introduction of a gas fill atmosphere in incandescent lamps dates back from 1883 when Edison tried to reduce the evaporation of carbon filaments. This first approach was not successful due to excessive thermal losses and the high risk of filament arcing (Suits and Way 1960). A more thorough investigation on gas-filled incandescent lamps was initiated by I. Langmuir in 1910 at GE. It was soon established that for a proper lamp operation, its atmosphere must meet certain requirements such as a low chemical reactivity, a low thermal conductivity, and a high electrical breakdown potential (Moore 1958). Gases which comply with these conditions are of the noble gas family and nitrogen.

Reduction of the Tungsten Evaporation Rate

The presence of a gaseous atmosphere strongly impedes on the transport of evaporated filament materials because of the frequent atomic collisions which scatter most of the tungsten atoms back toward their source. In his investigations, W. Elenbaas found that the net tungsten loss rate is a factor ~ 500 lower in the presence of a 1.27 bar argon atmosphere than in vacuum at constant filament temperature (Elenbaas 1972).

The reduction in filament evaporation rate depends on the density, radius, and mass of the atoms constituting the lamp atmosphere. Heavier and larger gas atoms are more effective at backscattering evaporated materials because of a higher probability of

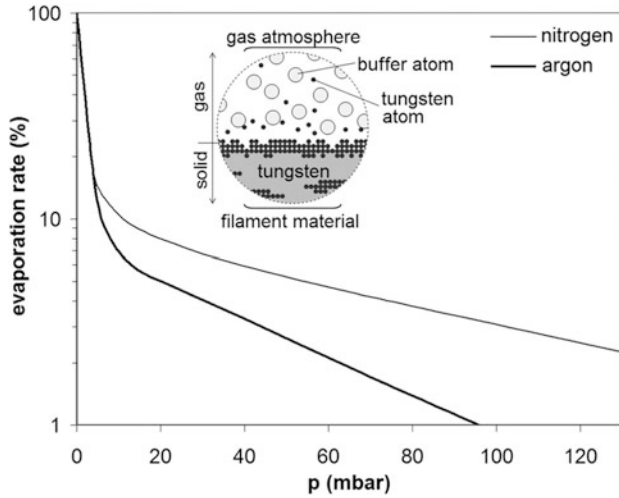


Fig. 6 Relative evaporation rate of a hot tungsten filament as a function of the nature and pressure of the buffer gas (Data and insert from Brons 1987). The insert is a magnified representation of the filament-gas interface with the processes of tungsten evaporation and diffusion in a gaseous atmosphere

atomic collision. Figure 6 shows that argon (39.9 amu) is more effective at reducing the tungsten evaporation rate than the lighter nitrogen molecules (28.0 amu).

The increase in fill pressure results also in a higher rate of atomic collisions, which leads to a significant reduction in net tungsten evaporation rate with a more pronounced effect from heavier gases, as shown in Fig. 6. It thus results from Eq. 13 that lamp life is positively affected by higher fill pressures and heavier gas fills. For this reason certain lamps are filled with high-pressure xenon so as to provide a better life-efficiency balance (see section “High-Pressure Xenon Incandescent Lamps”). The positive impact of the gas fill on lamp life is also evident from the K_L value relative to 40–1000 W argon-nitrogen lamps, which ranges from 25.7 to 42.1 g cm⁻³ and is about 20–25 times greater than for vacuum lamps (Vukceвич 1992).

The interactions between the gas fill and incandescent filaments were the subject of extensive investigations in order to better understand the factors affecting filament life and to optimize lamp designs (Elenbaas 1963, 1972; Covington 1973a; Vukceвич 1992). However, the presence of an atmosphere around the filament results in significant gas thermal conduction and convection losses, issues which were addressed in separate studies.

Thermal Conduction Losses in the Gaseous Atmosphere

The first significant research work on heat transfer mechanisms through gases at high temperature was initiated in 1910 by I. Langmuir (GE), who subsequently introduced the concept of a stationary gas film attached to material surfaces in order to properly compute heat transfer losses from hot bodies (Suits and Way 1960). The theory proved successful, evidences of the gas film (later called a Langmuir film) were

Table 1 Constant parameters in Eq. 15 for the most common gases used in standard incandescent lamps (Vukceвич 1992)

Constant	Nitrogen	Argon	Krypton	Xenon
K_1 ($\text{W m}^{-1}\text{K}^{-5/2}$)	$2.27 \cdot 10^{-3}$	$1.16 \cdot 10^{-3}$	$6.84 \cdot 10^{-4}$	$4.11 \cdot 10^{-4}$
K_2 (K)	1158	566	674	711

found experimentally (Covington 1968), and the theory was then refined further (Elenbaas 1963, 1972; Schirmer et al. 1967).

The formation of the Langmuir film occurs as a consequence of the increase in gas viscosity at very high temperature, causing a strong reduction in gas flow velocity (Suits and Way 1960). The thickness of the resulting stationary gas film is independent of the filament wire diameter and temperature, but is a function of the gas fill nature, pressure, and temperature. In a 1 bar argon atmosphere, the typical film thickness is around 2.5 mm (Dérivé 1965).

The lack of gas convection flow in the Langmuir film results in a thermal energy transfer from the filament through the stationary gas layer which occurs only via a conduction process. The thermal power P_c flowing from the incandescent filament surface is thus best described by Fourier's law of conduction applied to a cylindrical geometry:

$$P_c = 2\pi r L_F \kappa(T) \cdot \frac{dT}{dr} \quad (14)$$

where r is the position along the radial direction, L_F is the filament length, and $\kappa(T)$ is the heat conduction coefficient of the gas atmosphere. The latter coefficient has a significant impact on the magnitude of P_c and is best expressed using the empirical relationship given below (Vukceвич 1992), which is valid in the 1000–3500 K range:

$$\kappa(T) = \frac{T^{3/2}}{2} \cdot \frac{5K_1K_2 + 3K_1 \cdot T}{(K_2 + T)^2} \quad (15)$$

The constants K_1 and K_2 , whose values are given in Table 1, were found empirically.

The thermal conductivity decreases with gases of increasing molecular weight, in the left-to-right order of species listed in Table 1. At 2800 K, the conductivity of nitrogen, argon, krypton, and xenon is 152.3, 85.2, 49.4, and 29.5 $\text{mW m}^{-1}\text{K}^{-1}$, respectively. Since neon and helium are much lighter, they have a much higher κ value (222.6 and 773.9 $\text{mW m}^{-1}\text{K}^{-1}$ at 2800 K, respectively) which prevents their use in efficient incandescent lamps. Argon is the most commonly used gas because it provides a good balance between cost, efficiency, and service life. Heavier noble gases significantly reduce the filament heat losses and enable more efficient lamps (Thouret et al. 1975, Valin and Magnien 1991).

Gas Mixture and Optimum Fill Pressure

Because of their more favorable properties, noble gases are preferably used for the lamp atmosphere. However, a 5–10 % fraction of nitrogen is added in order to limit the risk of arcing arising from the filament thermionic emission (Clapp 1950) and from the spurious release of easily ionizable species from the tungsten material (Fax et al. 1971; Horacek 1980).

The gas fill pressure is adjusted so as to provide the best balance between filament life and thermal losses (Coaton and Marsden 1997). The latter is affected by gas convection processes which are stronger in larger lamps and enhance the overall heat losses. For this reason, large lamps are usually filled at a lower pressure so as to reduce the strength of gas convection (Dérivé 1965; Brons 1987). According to Brons the optimum fill pressure for a standard incandescent lamp filled with a 95 % Ar/ 5 % N₂ mix is around 80 mbar. In practice the cold fill pressure is usually around 0.8–0.9 bar so as to reduce the risk of arcing (Moore 1958), and the pressure in smaller lamps can be as high as 5 bars.

Filament Coiling and Thermal Management

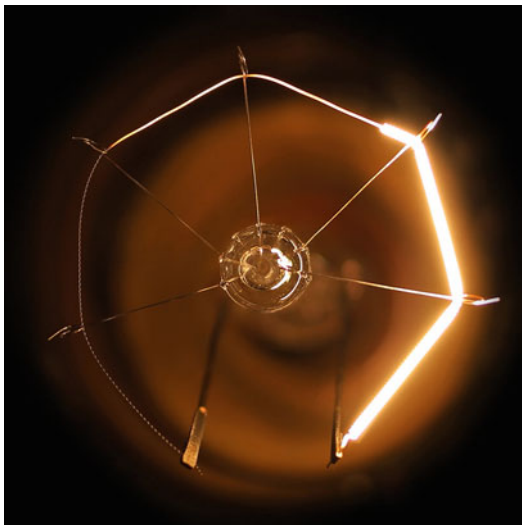
Because most of the temperature drop between the incandescent filament and its surroundings occurs within its Langmuir film (where $\Delta T > 2000$ K), the overall heat losses can be estimated using Fourier's law of thermal conduction (Vukcevic 1992). Considering the filament as a material cylinder of radius R_F and temperature T_F , surrounded by a stationary gas sheath of total radius R_L and with an outer edge temperature T_L , the overall thermal power P_c conducted away per filament length L_F is:

$$\frac{P_c}{L_F} = 2\pi \cdot \frac{\int_0^{T_F} \kappa(T)dT - \int_0^{T_L} \kappa(T)dT}{\ln(R_L/R_F)} = \frac{2\pi K_1}{K_2 + T} \cdot \frac{T_F^{5/2} - T_L^{5/2}}{\ln(R_L/R_F)} \quad (16)$$

The values of the constant parameters K_1 and K_2 are the same as those given in Table 1. It is clear from Eq. 16 that the total heat loss from the incandescent filament is linearly proportional to its length while its diameter has a much lesser influence. It is from this insight that Langmuir proposed in 1912 to coil the filament so as to reduce its length by a factor 5–10 and decrease the free surface area by a factor 2 (Langmuir 1916b; Fischer et al. 1975).

The specially designed lamp shown in Fig. 7 illustrates the impact of filament coiling on its thermal balance. The fraction of heat conduction loss is so significant in the straight filament wire section at left that it is barely hot enough to radiate visible light. On the opposite side, the significantly increased linear power dissipation in the coiled coil filament section results in a much lower heat conduction loss fraction, which raises the temperature and luminance of the filament.

Fig. 7 Operation of a 60 W–220 V gas-filled incandescent lamp having different filament sections: straight wire (*left*), single coiled wire (*center*), and coiled coil wire (*right*) (Lamp design by J.D. Hooker)



The introduction in 1913 of the gas-filled single coil filament raised the lamp efficacy to 10–15 lm W^{-1} from the 7–8 lm W^{-1} figure of vacuum lamps with straight wire filament (van den Hoek 2010). The coiled coil filament lamp, released in 1933, raised the efficacy by a further 20 % (Moore 1958; Stoer 1986).

Overall Optical Emission and Energy Balance

Impact of Filament Coiling on the Spectral Emissivity Properties

When a tungsten wire is wound into a coil, its spectral emission properties change and become intermediate between those of a blackbody and those of the straight wire (Vukceвич 1992). This modification was first reported and investigated in the mid-1910s by I. Langmuir (1915) who correctly identified the multiple light reflections between the filament windings as the cause. Further analyses showed that these reflections increase the overall optical absorbance of the inner side of the filament coil, which raises the local emissivity. This characteristic thus causes the surface brightness of incandescent filaments to be higher inside the coil than outside, as shown in Fig. 8.

The spectral emissivity of a tungsten filament is thus a function of its structure and more particularly of the coil pitch (Vukceвич 1992). Light emitted from more compact tungsten coils are more subjected to internal reflections, which results in a higher average emissivity and in a less selective emission. The spectral emissivity of the most compact coiled coil filaments can reach a peak value of 0.68, around 0.2 higher than that of a straight tungsten wire (Brett et al. 1981).

Fig. 8 Close-up view of a 1000 W–24 V coiled coil tungsten filament during operation, showing the increased emissivity inside the coiled structure. The temperature of the tungsten wire is constant across its diameter, so the visible brightness change does not originate from a temperature difference but is caused by multiple light reflections



In any cases the emissivity of tungsten filaments remains always lower in the infrared than in the visible, resulting in a more efficient production of visible light than with a blackbody at the same temperature (Coaton and Marsden 1997). Nevertheless, in the normal operating temperature range of 2400–3000 K, the peak of spectral emission lies in the near infrared (1150–835 nm), and only a small fraction (3–10 %) of the emitted light lies in the visible domain.

However, the emission properties of tungsten coils still result in a 10–30 % higher lumen efficacy than from a blackbody radiator, while the color temperature of the emitted light is also 50–100 K higher than the physical body temperature (Waymouth 1987; Valin and Magnien 1991).

Power Balance of Incandescent Lamps

In gas-filled incandescent lamps, the power input is dissipated via three loss channels (Hoegler and McGowan 1984; Brons 1987; Vukceвич 1992): the production of radiation (visible and infrared), the heating of the lamp atmosphere, and the heating of the lead and support wires. Radiation is the largest loss channel with 75–90 % of the lamp power converted into light. The second loss channel is the heat transfer via the lamp atmosphere, through which 10–20 % of the input power is dissipated. Finally, the power losses from the lead and support wires are limited to 1–5 % depending on the lamp design. Figure 9 presents the typical power fraction of four loss channels for three types of standard lamps.

Not shown in this figure are additional losses in the lamp caused by optical absorptions in the glass bulb and by light shading by the base. These losses amount to 6.5 % in standard lamps and do not affect the balance given in Fig. 9 because this process changes only the nature of the overall losses as seen from outside the lamp (Hoegler and McGowan 1984).

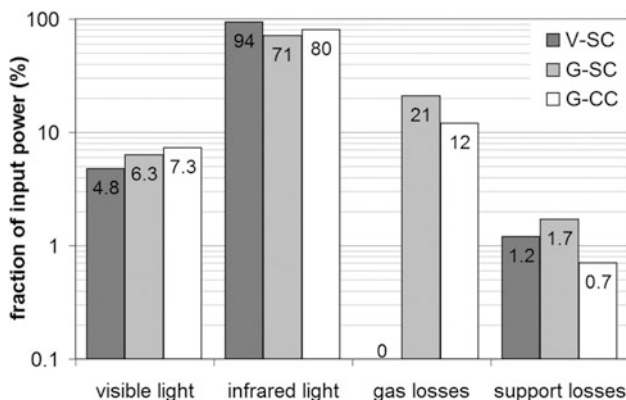


Fig. 9 Typical power distribution in three types of standard incandescent lamps: vacuum type with single coil filament (V-SC), gas filled with single coil filament (G-SC), and gas filled with coiled coil filament (G-CC) (Data from Brons 1987)

Tungsten Metallurgy and Filament Production

Limitations in the Tungsten Mechanical Properties

Despite its favorable properties, tungsten was not used in the first incandescent lamps due to major difficulties in fabricating long and thin wires. This metal is intrinsically brittle in its crystallized state, and the first wires produced in 1903 by A. Just and F. Hanaman were so fragile that they found only a limited use in practical light sources (Stoer 1986; Bright 1949). However, their improved light emission performances drove research efforts on tungsten metallurgy so as to develop wires with more favorable mechanical properties.

Development and Manufacturing of the Ductile Tungsten Wire

Tungsten becomes ductile at high temperature and as opposed to most other metals, its mechanical deformation results in the material becoming malleable at room temperature (Smithells 1936). With this knowledge, W.D. Coolidge (GE) began his research work on this metal in 1905 and successfully drew the first ductile tungsten wires 3 years later (Coolidge 1913; Moore 1958). The new wire proved essential for the mechanized production of coiled filaments.

The current fabrication process of ductile tungsten wires is not so different from the six-step procedure devised by Coolidge over a century ago (Smithells 1936, Henderson and Marsden 1972):

- The first step is the extraction of pure tungsten metal from scheelite (CaWO_4) and wolframite ($(\text{Fe},\text{Mn})\text{WO}_4$) which are first treated into tungsten trioxide, followed by a reduction process in hydrogen at high temperature.
- The tungsten powder is then mixed with a binder (water, glycerin, or paraffin wax) and is pressed into fragile bars.

- The tungsten bars are pre-sintered in hydrogen so as to improve the cohesion between metal grains.
- A full sintering process is then applied via Joule heating at 3300 °C in hydrogen in order to increase the tungsten density to 85–95 % of full density, which shrinks the metal bar by about 15 %.
- The diameter of the sintered bar is then reduced via a swaging process. The bars are first heated to 1400–1600 °C and are then pulled through a die which yields 10,000 percussions per minute and reduces the rod diameter by 12 % per pass.
- Swaged tungsten wires are then lubricated with graphite and are pulled through several tungsten carbide dies of successively decreasing diameters and at a temperature decreasing from 1000 °C to 400 °C. At certain stages of the drawing process, the wire is annealed in order to optimize its mechanical properties.

The final wire diameter depends on the material composition and the intended use of the thread. Wires can be made as fine as 10 µm with this process and smaller diameters are obtained with chemical or electrolytic etching or via ion sputtering in argon plasmas (Smithells 1936).

The swaging and drawing processes elongate the tungsten crystals along the wire's axis, which result in a fibrous microstructure (Fig. 10, upper micrograph) that significantly increases the tensile strength and ductility of the metal (Waymouth 1987). However, the recrystallization of pure tungsten above 1300 °C results in the nucleation of strain-free equiaxed grains which can be large enough to compromise



Fig. 10 Fibrous structure of ductile tungsten wire formed after the swaging and drawing processes (*upper* micrograph). Long interlocking grains formed in doped tungsten material following recrystallization above 2000 °C (*lower* micrograph) (Gendre 1999, courtesy of prof. O.J. Gregory). The space between each unit mark in the scale *inserts* is 10 µm long

the mechanical strength of incandescent wires. This issue was serious enough that it prompted the development in the 1910s of tungsten materials having superior mechanical properties at high temperature in order to ensure a stable filament structure throughout the lamp life (Horacsek 1980).

Control of the Crystallographic Structure of Tungsten

The deformation, or sagging, of early ductile tungsten wires is caused by the offsetting of tungsten crystals above 2300 °C (van den Hoek 2010), a mechanism driven by the migration of dislocations (Brons 1987). This issue became particularly problematic after the introduction of the gas-filled coiled filament lamps in 1913, because it compromised the structural stability of the thermal radiator. The solution came with the non-sag tungsten wire, which was first produced in the mid-1910s by A. Pacz (GE) (Pacz 1922) by adding a specific dose of potassium to the batch of tungsten oxide at the very beginning of the wire fabrication process (Smithells 1936; Suits and Way 1960). Lamps using the improved tungsten wire were then first introduced around 1917 and featured a much more stable coiled filament.

To this day potassium remains the most important tungsten dopant and is added to the metal oxide powder as an aluminosilicate, mostly in the form of a mixture of Al_2O_3 , K, and SiO_2 , resulting in the so-called sag-resistant AKS tungsten (Smithells 1936; van den Hoek and Jack 1990). During the reduction and sintering processes, about 90 % of the dopants are evaporated, leaving around 30–50 ppm of potassium in the material (Horacsek 1980; Waymouth 1987). Part of the remaining dopes is trapped in the metallic form inside many microbubbles which are dispersed in the metal bulk. The low solubility of the metallic dopants in tungsten prevents the sintering of these microbubbles, whose lasting presence is directly responsible for a change in crystallographic structure (Horacsek 1980; van den Hoek and Jack 1990).

During the drawing process these bubbles are stretched into dopant-filled microcavity lines which, during the annealing process, break up into rows of 5–50 nm bubbles aligned along the wire's axis (Horacsek 1980). During recrystallization these bubbles block the migration of dislocations and force a parallel growth of the crystal grains whose aspect ratio becomes larger than 10:1, with a length exceeding the wire diameter (Waymouth 1987). The resulting interlocked crystal structure (Fig. 10, lower micrograph) presents very large grain boundaries that are predominantly aligned along the wire axis, which impedes on diffusion creep and makes the wire resistant to sagging (Brons 1987; van den Hoek and Jack 1990).

Rhenium is also an important doping element used to improve further the mechanical properties of tungsten at high temperature (Carlson and Hurd 1965). Rhenium-doped tungsten wires are usually used in filaments for applications where lamps are subjected to shocks and vibrations.

Properties of Doped Tungsten Wires and Coil Fabrication

The presence of dopants increases the recrystallization temperature of tungsten from 1300–1400 °C to about 2000 °C (Horacsek 1980). This change facilitates the production of ductile tungsten wires since the material can be processed at higher

temperatures in order to obtain the most optimal material properties. The resulting wires are ductile enough for the filament fabrication, while the increase in grain creep resistance following recrystallization above 2000 °C effectively stabilizes the filament structure (Horacek 1980; van den Hoek and Jack 1990).

The filament coil is formed by winding the ductile tungsten wire around a mandrel made of molybdenum or iron. In the case of coiled coil filaments, the first coil assembly is also wound around a second larger mandrel. During this process the tungsten wire is heated to around 1700 °C so as to prevent the buildup of mechanical stress. This annealing step is optimized so as to prevent recrystallization, which would otherwise compromise the handling and mounting of the filament in later stages of the lamp manufacture (Henderson and Marsden 1972; van den Hoek and Jack 1990). Finally, the tungsten filament is released from its support by dissolving the mandrel material in an acid bath.

Standard Incandescent Lamps for General Lighting Applications

The standard general lighting service (GLS) lamp is the most widely used type of incandescent source and comes in a large variety of forms and shapes so as to address a wide range of applications. This section presents their general design and the characteristics of standard and reflector lamps.

General Structure and Design

GLS lamps contain the filament inside a hermetically sealed glass bulb in a typical arrangement, shown in Fig. 11, which in most cases has a single-ended configuration with a screw-fit Edison (E) or a bayonet base (B). A relatively large bulb is used in order to limit the surface accumulation of evaporated tungsten. Inside, the filament is clamped at both ends to current-carrying lead-in wires and is supported by one or several molybdenum wires. The electrical feedthroughs consist of copper-clad (*Dumet*) wires which form gastight seals with the stem glass. The electrical connection is then made to the base terminals via one or two fuses, which protect the lamp from arcing at the end of life (see section “[Anomalous and Nonpassive Failure Mechanisms](#)”).

The filament operates typically at 2400–2600 K in vacuum and at 2600–2800 K in a gas atmosphere. At the exception of krypton lamps, all mains-voltage lamps at or below 25 W are of the vacuum type in order to limit thermal losses. In all cases single coil or coiled coil filaments are used so as to facilitate lamp production.

The atmosphere in standard gas-filled lamps consists of an argon-nitrogen or a krypton-nitrogen mixture at a cold pressure of 0.8–0.9 bar. When krypton is used, either the lumen efficacy is raised by 10–20 % with the same service life as that of argon lamps or the life is doubled with an efficacy increased 5 % from that of argon lamps (Valin and Magnien 1991).

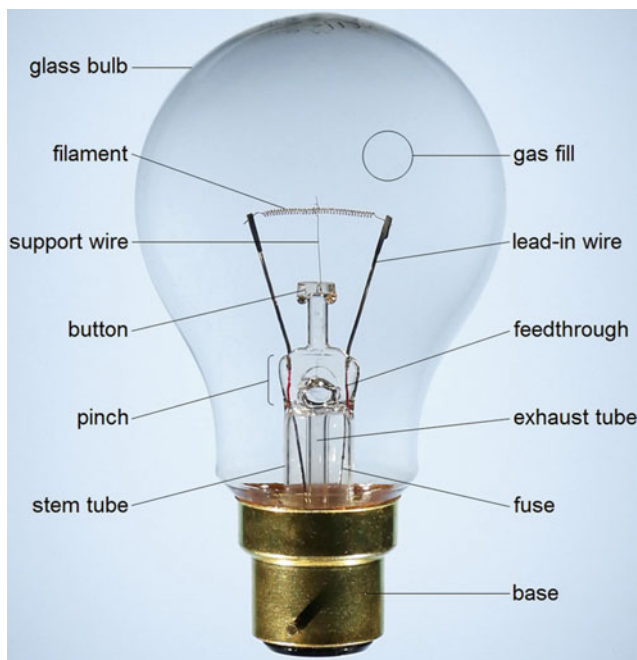


Fig. 11 Structure of a standard 60 W–120 V general lighting service (GLS) lamp (105 × 60 mm) made with a soda-lime silicate glass bulb attached to a B22 base. It also features a coiled coil tungsten filament and an Ar-N₂ fill atmosphere

The level of impurities in lamps must be less than 10 ppm so as to ensure their proper operation (van den Hoek and Jack 1990). To this end, easily oxidizable materials, such as zirconium, phosphorus, or barium, are flashed inside the lamp in the last stage of manufacturing in order to remove residual oxygen, water vapor, and hydrocarbon (Dérivé 1965).

At the exception of certain krypton, miniature, and special lamps, the average life expectancy is set to 750–1000 h so as to provide the most economical operation (see section “[Balance Between Efficacy and Service Life](#)”). Over the course of life, the light output decreases by 20 and 10 % in vacuum and gas-filled lamps, respectively (Dérivé 1965). This is a result of the continuous filament evaporation which causes bulb darkening, uniform in vacuum lamps, and localized in gas-filled lamps due to the convective transport of tungsten (Brons 1987).

Standard Lamps for General Lighting Service

Although most GLS lamps feature a pear-shaped (PS) or an arbitrarily shaped (A) bulb (Fig. 12(1, 2) and (3, 4), respectively), other shapes are also used: tubular (T) in (5, 9–11), straight sided (S) in (6), mushroom (M) in (7), and elliptical (E) in

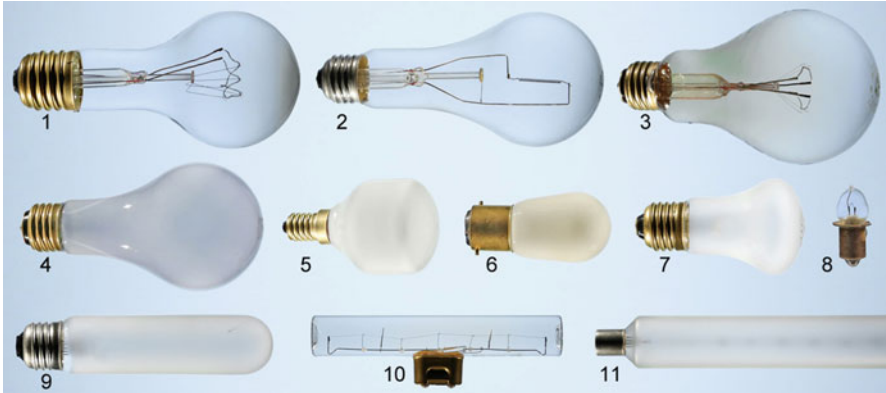


Fig. 12 Range of standard incandescent lamps with uniform light distribution: 300 W–230 V with filament in wreath configuration (1), 200 W–120 V with axial filament (2), 60/100 W–220 V with dual filaments in opposite half-ring arrangement (3), 50/100 W–120 V with light silica coat (4), 40 W–220 V with thick pyrophosphate coat (5), 5 W–220 V with opal glass (6), 60 W–130 V krypton lamp with inside-etched bulb (7), 3 W–6 V krypton miniature lamp (8), 40 W–120 V with inside-etched tubular bulb (9), 15 W–120 V with central base (10), and 60 W–220 V in double-ended configuration with external diffuse coating (11). The last three lamps are made with an axial linear single-coiled filament

Table 2 Characteristics of standard mains-voltage incandescent lamps for general lighting service

Lamp type	Power (W)	Voltage (V)	Flux (lm)	Efficacy (lm W ⁻¹)	Color temp (K)	Lifetime (kh)
Vacuum (clear bulb)	2–25	100–240	7–200	3.5–8	2400–2600	1.0–4.0
Argon filled (clear bulb)	40–1000	100–240	430–20,000	10.5–20	2600	1.0
Argon filled (diffuse bulb)	40–200	100–240	410–3000	10.3–15	2600	1.0
Krypton filled (diffuse bulb)	25–150	100–240	215–2250	8.6–15	2600–2700	1.0–2.0

(8). Because of its reduced volume, the M bulb is used primarily in krypton lamps. For certain applications lamps are also made with a double-ended structure, as shown in (11). The typical characteristics of the most common types of mains-voltage GLS lamps are provided in Table 2.

The filament structure varies also with the lamp type, and three of the most usual configurations are shown in Fig. 12(1–3), with the wreath arrangement (1) the most commonly used. The axial filament configuration (2) improves the lamp lumen maintenance when operated vertically because of the optimized convective

transport of evaporated tungsten (Pearson et al. 1956). The dual filament configuration (3) enables a step-variation in lamp power and output flux without affecting its efficacy and light color temperature.

Lamps are designed either with a clear (Fig. 12(1–3, 8, 10)) or a diffuse finish (4–7, 9, 11), the latter being used to reduce glare when the source is in direct view. This finish is implemented in different forms: with a thin silica coat for minimal light scattering (4), with an inside-etched bulb (7, 9), or with an outer diffuse coat (11). Minimum glare is obtained with a thick silica or pyrophosphate coat (5) or with an opal glass (6).

Standard Reflector Lamps

There are two types of reflector lamps: sources for forward beam projection (Fig. 13 (1–5, 8)) and lamps with backward light reflection (6, 7). Four main types of reflector configurations are used in the first case, with the standard R type being the most common (1). Higher light collection efficiencies are realized with elliptical (ER) and double reflector (NR) configurations (2 and 3, respectively). Finally, the highest light intensities are realized with precision-shaped parabolic reflector (PAR) lamps (4). The characteristics of these lamps are given in Table 3. Finally, a less common type of lamp features a standard glass bulb which is half coated with a reflecting material (5, 8).



Fig. 13 Range of standard reflector lamps for general lighting service: 150 W–235 V R125 80° (1), 50 W–220 V ER95 35° (2), 75 W–230 V NR95 35° (3), 150 W–240 V PAR121 30° with cold dichroic mirror (4), 100 W–220V krypton lamp with titanium dioxide coating (5), 150 W–120 V with inside-etched bulb and aluminum crown mirror (6), 60 W–230 V with gold crown mirror (7), and 25 W–220 V double ended with clear tubular bulb and silver side mirror (8)

Table 3 Characteristics of standard incandescent reflector lamps for general lighting service

Lamp type	Power (W)	Voltage (V)	Intensity (kcd max)	Efficiency (cd-sr/W max)	Beam angle (degrees)	Lifetime (kh)
Standard reflector (R)	25–300	120–230	0.12–6.0	1.9–10.1	25–80	1.0
Elliptical reflector (ER)	50–120	120–130	0.5–1.2	2.9	35	1.0
Double reflector (NR)	25–150	230	0.21–4.2	1.8–3.0	20–30	1.0
Pressed parabolic (PAR)	25–500	12–230	1.2–46	2.0–12.7	7–40	2.0

Reflector lamps with a backward light reflection design are made with a standard bulb with a metallized crown section (Fig. 13(6, 7)). These crown mirror lamps are often used in combination with a secondary parabolic reflector for the forward projection of a tightly controlled beam with very limited spill light.

Commonly used light-reflecting materials are either metals or oxides. Metal films include aluminum, silver, and gold, with the first metal used the most. Silver is preferred when a higher reflectivity is needed, while gold is used only in specialized cases. Oxide materials include titania and multilayer interference (dichroic) mirrors. The latter consists of 20–50 stacked layers of transparent dielectric materials, such as TiO₂, SiO₂, Ta₂O₅, or ZnS, having alternatively low and high refractive index (Beesley et al. 1963; Yuge 1995). Light interferences occurring within this structure result in well-defined and adjustable optical properties with complementary light transmission and reflection spectra (Law 1965). Because of low intrinsic optical losses, the peak transmittance and reflectance can reach near 100 % (Köstlin and Frank 1983–1984). For this reason dichroic mirrors are used in high-performance sources (Fig. 13(4)) where they can significantly reduce the infrared output in the direction of the projected light beam.

Although the finish of reflector lamps is most often clear, a diffuse coating or an inside-etched glass bulb can be used to reduce glare and homogenize the projected light beam. PAR lamps are also often made with a front glass plate having a stippled or lensed pattern in order to adjust the projected beam to the desired specifications.

Tungsten Halogen Lamps

A tungsten-halogen cycle effectively prevents the deposition of filament material onto the bulb and enables more compact lamp designs. This technology is treated in details in this section with the lamp chemistry described in the first part, followed by the structure and design of halogen lamps in the second part, and finally various types of tungsten-halogen lamps are reviewed in the last three parts.

The Halogen Cycle

Although inert lamp atmospheres are instrumental in limiting the filament evaporation, tungsten is still continuously lost and accumulates onto the lamp bulb over time. This particular problem is solved with a chemically active halogen atmosphere which reacts with tungsten and forms tungsten halide and oxyhalide compounds following the left-to-right reaction described in Eq. 17, where X is a halogen species. The volatile oxyhalide molecules do not condense and are brought back to the filament by convection currents (Moore and Jolly 1962; Dettingmeijer et al. 1975):



As oxyhalide molecules migrate toward the incandescent filament, the rise in temperature causes a gradual molecular decomposition which releases halide and oxide compounds first and then metallic tungsten atoms (Fig. 14). The full dissociation of oxyhalides (i.e., right-to-left reaction in Eq. 17) occurs typically around the edge of the Langmuir film, where the radial temperature gradient is the strongest (Elenbaas 1972).

The direction and rate of net tungsten transport depend on the gradients in the total effective tungsten pressure $\Sigma p(W)$. Since tungsten compounds and atoms diffuse from high to low $\Sigma p(W)$ zones, the lamp chemistry is then adjusted in such a way that $\Sigma p(W)$ is maximum at the wall so as to ensure that tungsten is always returned to the filament.

However, since most oxyhalides dissociate at a temperature below that of incandescent filaments, the halogen cycle is inoperative inside the Langmuir film (Dettingmeijer et al. 1975). It results that the longitudinal diffusion of tungsten at

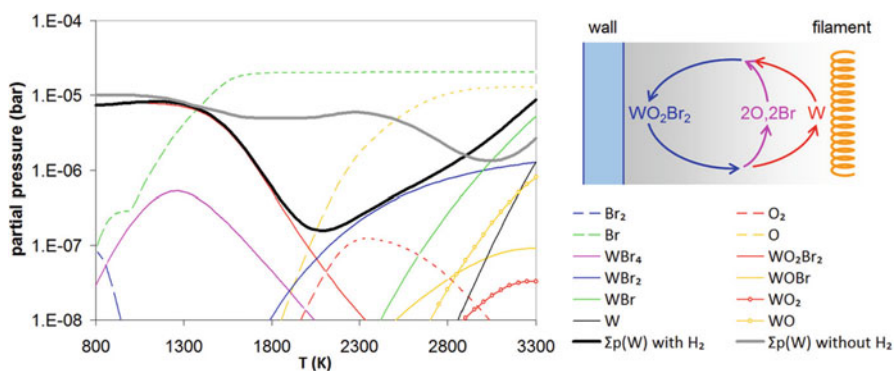


Fig. 14 Calculated partial pressures of main species present in a halogen lamp filled with 5 bar argon with 1 % dibromomethane (CH_2Br_2), with a residual oxygen pressure of 1 Pa and a wall and cold spot temperature set at 750 K. The summed elemental tungsten pressure corresponding to this system is drawn with the *thick black line*, while that for the hydrogen-free oxygen-bromine system is plotted with the *thick gray line*. A simplified schematic of the halogen cycle is inserted in the *top right corner*

the filament surface remains unimpeded (Elenbaas 1972), and the lamp temperature and chemistry must be properly adjusted so as to avoid local tungsten erosions and accumulations.

The temperature evolution of $\Sigma p(W)$ is mostly defined by the formation enthalpy of various tungsten compounds and the density of oxide molecules in the gas phase. The latter depends on the volatility of condensed tungsten oxides and on the lamp cold spot temperature (Dettingmeijer et al. 1975). For these reasons, the halogen cycle characteristics are strongly influenced by the lamp design and its fill chemistry, which must be both precisely tuned for a proper lamp operation.

The first tungsten-halogen lamps were filled with iodine because of its low chemical reactivity (Zubler and Mosby 1959) and this halogen was eventually replaced by bromine, which has more favorable characteristics (T'Jampens and van de Weijer 1966). Because bromine is aggressive toward tungsten below 1970 K, hydrogen is added so as to form stable hydrogen bromine compounds and reduce the free halogen density below 2200 K. This chemical change results in a dip in $\Sigma p(W)$ between 1400 and 2200 K (Fig. 14), which reduces the metal erosion rate in colder filament and lead-in wire parts (Elenbaas 1972; van den Hoek and Jack 1990).

Bromine is usually dosed as dibromomethane (CH_2Br_2), mixed with a neutral buffer gas in a 1:99 ratio (Elenbaas 1972). Oxygen, which is needed for the formation of volatile oxyhalides, is already present in the lamp as an impurity. Any excess oxygen reacts with tungsten and forms tungsten oxide which condenses at the cold spot location. Once sealed, the internal bromine and oxygen pressures are around 10^3 and 10^{-3} Pa, respectively (T'Jampens and van de Weijer 1966; Dettingmeijer et al. 1975).

General Structure and Design

Halogen lamps are designed with a compact quartz or aluminosilicate glass bulb which is close enough to the filament for an 800–1000 K wall temperature, as required for a proper tungsten-bromine cycle operation (Clark 1966; T'Jampens and van de Weijer 1966). The two most common halogen lamp designs are the compact single-ended and the linear double-ended burners, shown in Fig. 15. The former type features a compact filament suitable for a wide voltage range (3–240 V), while linear lamps have a longer axial filament for mains-voltage operation only. The filaments are connected to electrical feedthroughs which consist of tungsten wires in aluminosilicate glass lamps or of molybdenum foils in quartz lamps.

The atmosphere of halogen lamps consists of a mixture of dibromomethane and a neutral buffer gas which is usually an argon-nitrogen mix. Because of their small volumes, the use of heavier, more expensive noble gases does not have a significant impact on production cost, and a krypton buffer is often used when better lamp performances are needed.

The robustness of halogen lamps and the limited convection flow inside their small bulbs enable cold fill pressures as high as 3 bars. As a result, filaments are operated in a higher temperature range of 2800–3200 K, with an average around

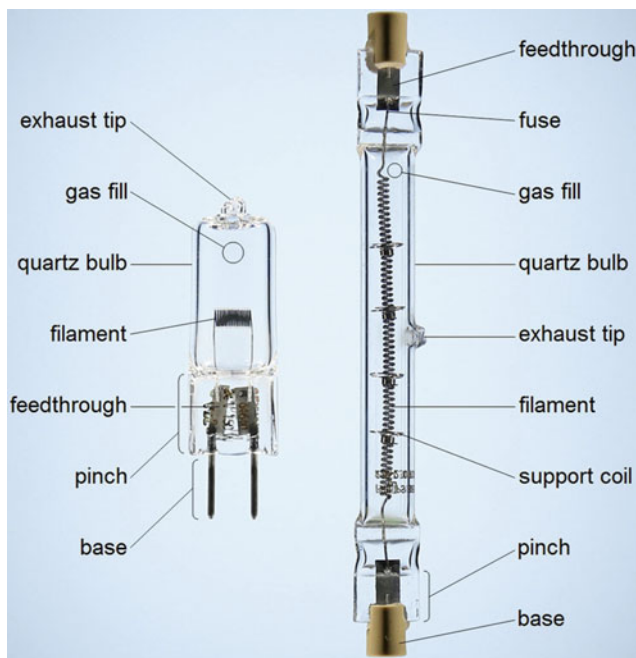


Fig. 15 Structure of two of the most common tungsten-halogen lamps used in general lighting: a low-voltage single-ended capsule at *left* (150 W–24 V, 58 × 16 mm) and a high-voltage double-ended lamp at *right* (1000 W–230 V, 12 × 114 mm). Both lamps are made of quartz

3000 K, resulting in higher efficacies (Dérivé 1965; Waymouth 1987). This also causes the enhanced emission of actinic light, which is usually filtered out using a UV-absorbing quartz bulb material doped with cerium or titanium oxide.

The higher fill pressure results also in an extended service life due to the reduction in tungsten evaporation rate. Standard halogen lamps are usually rated for 2000 h, and some lamps with a higher efficacy have a halved life expectancy, while others can reach up to 10,000 h with a reduced filament temperature.

Tungsten Halogen Burners for General Lighting

Bare halogen burners aimed at general lighting purposes are particularly compact sources which permit the use of small and efficient luminaires. These lamps are made with a double- or a single-ended design, each targeting different applications. Most halogen lamps are filled with an argon-nitrogen buffer, while some energy-saving types employ a krypton-based fill which raises the lumen efficacy by 20 %. The data relative to these lamps are provided in Table 4.

Table 4 Characteristics of standard tungsten-halogen lamps for general lighting service

Lamp type	Power (W)	Voltage (V)	Flux (lm)	Efficacy (lm W ⁻¹)	Color temp (K)	Lifetime (kh)
Linear double ended	48–2000	100–240	775–48,400	14–24	2900–3000	1.5–2.0
Low voltage single ended	5–150	6–24	60–3200	12–22	2800–3000	2.0–4.0
Mains voltage single ended	18–500	100–240	204–9500	9–20	2800–2950	1.5–2.0

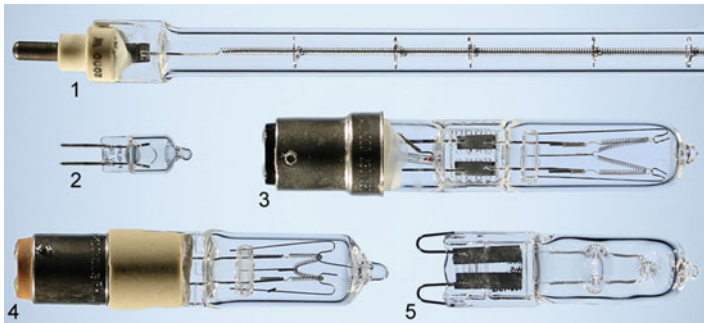


Fig. 16 Range of tungsten-halogen burners for general lighting applications: 2000 W–240 V with a double-ended linear design (1), 5 W–12 V with a hard glass bulb (2), 150 W–220 V with internal filament supports and fuse (3), 100 W–115 V with an insulating ceramic support (4), and 75 W–230 V in the compact single-ended configuration (5)

Double-ended lamps of the type shown in Fig. 16(1) are exclusively designed for mains-voltage operation and are most often used in combination with deep parabolic reflectors for precision floodlighting applications. The tungsten filament has an axial coiled or coiled coil configuration, supported on several points by either springs or local pinches in the quartz vessel. These lamps can be operated at all positions when the filament supports are mechanically fixed in the lamp.

The smaller single-ended burner construction simplifies the use of halogen lamps in compact luminaires and is made in two variants: low-voltage capsules for a 6–24 V operation (Fig. 16(2)) and mains-voltage lamps (100–240 V), provided with a dedicated base above 75 W (3–5).

For any given power input, the low-voltage halogen capsule can deliver 30–60 % more light than its mains-voltage counterpart on account of the lower thermal losses from the smaller low-voltage filament. Moreover, the more compact light source enables a more precise optical control of the emitted light. Many of these lamps are also designed with a low-pressure fill so as to reduce the risk of nonpassive failure and allow a use in open luminaires.

Despite their lower performances, mains-voltage single-ended lamps are popular in domestic lighting applications because these do not require any transformer for their operation. These lamps are also provided with internal fuses so as to limit the risk of arcing or of nonpassive failure at the end of life (see section “[Anomalous and Nonpassive Failure Mechanisms](#)”).

Jacketed Tungsten Halogen Lamps for General Lighting

Halogen lamps made with a glass jacket and mounted with a standard base can retrofit standard GLS lamps while providing a 10–30 % efficacy improvement and a doubled service life. Although the jacket increases the lamp size, it also protects the burner against external elements and can be made to contain burner fragments in case of nonpassive failure. All these lamps are designed for mains-voltage operation and are mainly classified into high-wattage and low-wattage sources, whose characteristics are given in Table 5.

High-power jacketed lamps are mostly used in floodlighting applications and are exclusively made with a double-ended burner housed inside a tubular borosilicate glass jacket (Fig. 17(1)). This particular configuration enables a 200 K higher operating temperature than in standard double-ended lamps, resulting in a 4–8 % efficacy increase. However, this also causes a higher hydrogen partial pressure at the burner wall which requires the outer jacket to be filled with hydrogen at the same pressure so as to prevent a net loss of this gas (T’Jampens and van de Weijer 1966).

Due to their limited light output, low-power jacketed halogen lamps are primarily intended for domestic lighting applications. Their design is closely related to that of standard GLS lamps and includes three types of halogen burners: quartz vessels with a single- and double-ended structure (Fig. 17(2, 3) and (4), respectively) and single-ended hard glass vessels (5, 8). Lamps provided with the last kind of burner are sometimes made with a thick-walled outer jacket (5) so as to protect from the higher risk of nonpassive failure at the end of life.

Low-wattage retrofit lamps below 150 W (Fig. 17(3–8)) are made in nearly the same range of shapes and finishes as their standard GLS counterparts (see section “[Standard Lamps for General Lighting Service](#)”). A neodymium glass jacket is sometimes used in order to modify the emission spectrum for some specific applications (4). Certain energy-saving retrofit lamps with a krypton-filled burner are designed with a 10–30 % lower power rating than the GLS lamps they replace

Table 5 Characteristics of standard jacketed halogen lamps for general lighting service

Lamp type	Power (W)	Voltage (V)	Flux (klm)	Efficacy (lm W^{-1})	Color temp (K)	Lifetime (kh)
High wattage with linear burner	300–2000	230–240	6.0–50.0	20–25	3000–3200	2.0
Low-wattage retrofit (clear bulb)	18–250	120–240	0.2–4.2	11–17	2800–2900	1.5–2.0

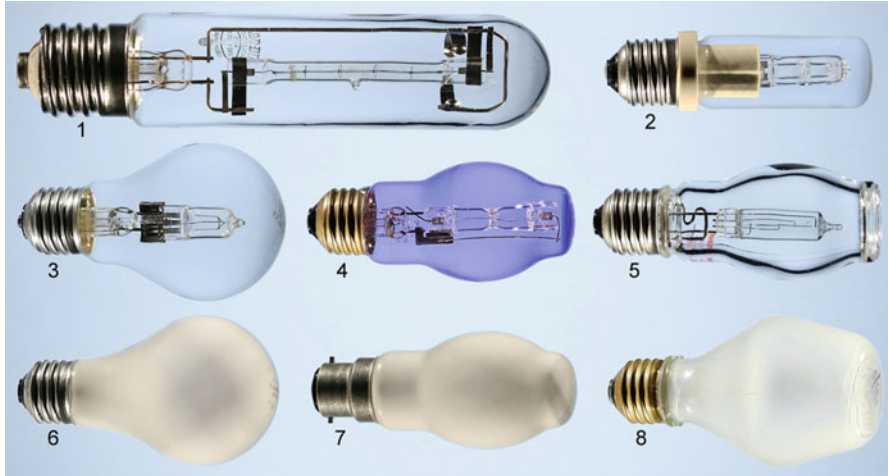


Fig. 17 Range of standard jacketed tungsten-halogen lamps for general lighting applications: 300 W–230 V with linear burner (1), 150 W–230 V with compact design (2), 42 W–220 V with standard A bulb (3), 60 W–120 V with neodymium glass jacket (4), 50 W–230 V with hard glass burner and thick-walled jacket (5), 45 W–120 V with light silica coat (6), 100 W–230 V with thick pyrophosphate coat (7), and 60 W–120 V with inside-etched thick-walled jacket (8)

while providing a similar light output. Finally, low-wattage lamps rated at 150 W and above have a compact design optimized for a limited heat transfer to the base (2).

Reflector Tungsten Halogen Lamps for General Lighting

With their compact filament and improved performances, halogen sources are ideally suited for the efficient projection of light, and as shown in Fig. 18, there is a large selection of halogen reflector lamps designed for this purpose. These are usually classified into either low-voltage sources for a 6–24 V operation with a transformer, or mains-voltage lamps for an operation at 100–240 V. Their characteristics are given in Table 6.

Most of these lamps have a precision-shaped reflector, sometimes made with various faceted patterns for the homogenization of the projected light beam. The reflecting coating consists either of a metal, such as aluminum or silver for a full-spectrum reflection and to reduce the heat load behind lamps, or of a cold dichroic mirror which reduces the amount of projected infrared light by 80 %. Reflector lamps have an open or closed front configuration depending on requirements with regard to nonpassive failure containment and UV protection. The front glass window can have a diffuse finish or a lensed pattern for the adjustment of the output beam angle.

Mains-voltage lamps are all built around single-ended burners of the type described in previous section. Those designed for the retrofit of standard GLS



Fig. 18 Range of tungsten-halogen reflector lamps for general lighting: 75 W–240 V PAR121 11° (1), 40 W–230 V NR50 30° (2), 50 W–230 V MR51 40° (3), 60 W–110 V MR51 40° (4), 10 W–12 V with side mirror coat (5), 35 W–12 V with axial mirror coat (6), 35 W–6 V AR111 3° (7), 35 W–12 V MR35 10° (8) and 50 W–12 V MR51 24° (9) with cold dichroic mirror, 20 W–12 V MR51 38° with lensed front window (10), 50 W–12 V MR51 38° with 4200 K dichroic mirror (11), 50 W–12 V MR51 38° with *blue* dichroic front filter (12), and 50 W–240 V MR63 with *yellow* dichroic mirror (13)

Table 6 Characteristics of standard tungsten-halogen reflector lamps for general lighting service

Lamp type	Power (W)	Voltage (V)	Intensity (kcd max)	Efficiency (cd·sr/W max)	Beam angle (degrees)	Lifetime (kh)
Standard/double reflector (R/NR)	18–70	120–230	0.21–2.0	2.5–6.8	20–43	2.0–3.0
Pressed parabolic reflector (PAR)	20–250	12–230	0.68–40.0	1.2–9.3	5–35	2.0–4.0
Compact (MR) mains voltage	18–75	110–230	0.28–2.5	1.4–7.8	20–50	1.0–3.0
Spun aluminum reflector (AR)	20–100	6–24	0.9–50.0	2.3–13.4	4–45	1.0–3.0
Compact (MR) low voltage	10–65	12	0.3–15.0	3.7–18.5	9–60	2.0–10.0

reflector lamps are provided with a standard base and are usually made with PAR, R, and NR bulbs (see section “[Standard Reflector Lamps](#)”), as shown in Fig. 18(1, 2). Non-retrofit lamps have dedicated bases and feature a more compact construction (3, 4) for an integration in small and lightweight luminaires.

Due to their higher efficacy and luminance, low-voltage halogen reflector lamps deliver superior performances in a smaller volume than their mains-voltage

counterparts. These lamps are also made in a wider variety of configurations, with the simplest one consisting of half-coated halogen burners (Fig. 18(5, 6)). However, sources provided with an external reflector are usually preferred due to their superior optical properties.

Excellent optical control of light is achieved with spun aluminum reflectors (AR) which enable the projection of very narrow beams with an angle as small as 4° while delivering some of the highest peak intensities. These lamps are often provided with a burner cap (Fig. 18(7)) so as to reduce spill light and prevent glare.

The highest overall optical efficiency is achieved in compact parabolic reflector (MR) lamps made with a diameter of 35 and 51 mm, as shown in Fig. 18(8, 9). Compared to their mains-voltage equivalents, these low-voltage sources deliver 2.5–6 times higher peak light intensity with an overall lumen efficacy increased by a factor 1.4–3.3. Most MR lamps are provided with a cold dichroic mirror so as to limit the heat exposure of illuminated objects. Sources with narrow beam angles are usually made with clear burner and front window (9), while wider beams are sometimes obtained with a lens pattern on the front window (10) or with an externally etched halogen burner (11).

Some MR reflector lamps operate at a reduced filament temperature so as to reach a service life of 10,000 h. In order to keep the light color temperature at 3000 K, a dichroic mirror designed with a lower red reflectance is used. However, the lamp efficacy and beam light intensity are consequently reduced by 20–25 %.

The same kind of dichroic mirror is also used to increase the light color temperature of standard MR lamps to 4200 K (Fig. 18(11)), but this results in 40–70 % lower peak intensity and a factor 1.8–3.2 lower overall efficacy compared to equivalent standard lamps. A more selective optical filtering is also used for the production of light with highly saturated colors. Such color filtering is applied to both low- and mains-voltage lamps where the dichroic filter is either on the front glass plate or used as the primary reflector, as shown in Fig. 18(12) and (13), respectively.

Incandescent Lamps for Special Applications

The incandescent light source technology is used in many other applications than just general lighting. This section presents details about five of the most important lamp types specifically designed for specialty applications such as vibration service and traffic signals, automotive lighting, stage and studio lighting, infrared processing, and instrument calibration.

Standard Lamps for Vibration Service and Traffic Signals

Lamps subjected to constant vibrations or used in applications requiring very low early failure rates feature a reinforced construction which includes a rhenium-doped tungsten filament supported on many points (Fig. 19(1, 2, 4, 5)). The high support

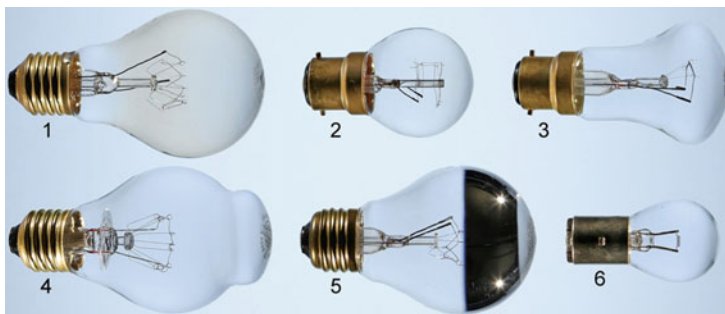


Fig. 19 Incandescent lamps for vibration service and traffic signaling applications: 60 W–220 V Ar-N₂ filled (1), 25 W–220 V vacuum lamp (2), both with a large number of filament supports, 55 W–230 V (3) and 69 W–130 V (4) with krypton fill, 51 W–130 V with ring crown mirror (5), and 45 W–10 V krypton lamp (6)

Table 7 Characteristics of incandescent lamps for vibration service and traffic signals

Lamp type	Power (W)	Voltage (V)	Flux (lm)	Efficacy (lm W ⁻¹)	Color temp (K)	Lifetime (kh)
Vibration service (vacuum and Ar-N ₂)	25–200	120–230	170–2500	6.8–12.5	2400	1.0–1.5
Traffic signal (vacuum and Ar-N ₂)	40–70	120–230	230–380	5.4–5.8	2300	8.0
Traffic signal (Kr) mains voltage	54–150	120–230	330–1275	6.1–8.5	2400	8.0
Traffic signal (Kr) low voltage	20–60	10–40	250–800	10.0–15.6	2400	8.0–15.0

losses combined with the low filament temperature in long-life lamps result in low efficacies (see data in Table 7). In order to limit thermal losses, these lamps have an Ar-N₂ fill only above 60 W. However, the efficacy is improved by 30 % with a krypton fill in certain mains-voltage lamps (3, 4) and in compact low-voltage sources (6).

Due to high reliability requirements, mains-voltage traffic signal lamps have an 8000 h service life with 2 % failure at 3000 h, while low-voltage lamps reach about 15,000 h with less than 2 % failure at 4400 h. Some of these lamps are provided with an AT bulb shape (Fig. 19(4)) so as to facilitate handling. Others are made with a ring crown mirror (5) which increases the optical control of light and enables a 17–33 % power reduction in existing systems.

Vibration service lamps are generally aimed at general lighting applications where lamp reliability requirements are less stringent than in traffic signaling. The lamps consequently have a much shorter life expectancy matching or slightly exceeding that of their standard GLS counterparts (see Table 7).

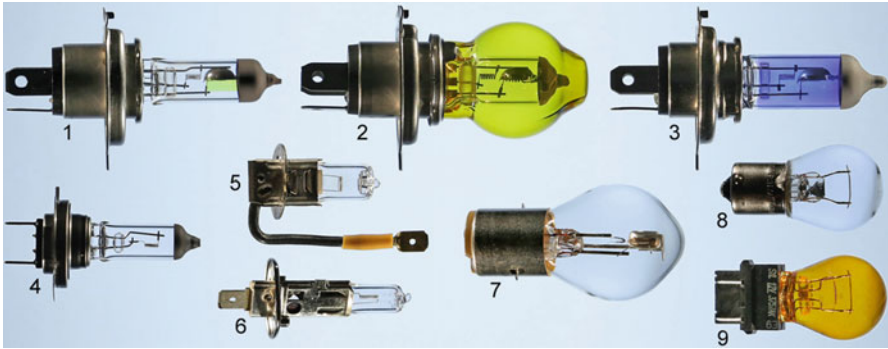


Fig. 20 Range of automotive incandescent lamps: 60/55 W–12 V quartz tungsten-halogen headlight lamp (1), 75/70 W–24 V with yellow filtering glass cover (2), 80/80 W–12 V with blue dichroic filter (3), 55 W–12 V with single filament (4), 55 W–12 V with lateral filament (5), 55 W–12 V with axial filament (6), 45/40 W–12 V standard headlight lamp (7), 21 W–12 V standard stoplight lamp (8), and 7/27 W–12 V with dual filaments and colored lacquer (9)

Automotive Lamps

Low-voltage incandescent lamps cater for a wide range of applications in automotive lighting, the most important of which is headlighting. Headlight sources are both of the halogen and of the standard types (Fig. 20(1–6) and (7), respectively) and feature compact vibration-resistant filaments operated around 3100 K for maximum photometric performances (up to 26 lm W^{-1} and 2 kcd cm^{-2}). Their service life is consequently short: 250–400 h for halogen lamps and around 60 h for standard sources.

All headlamps are provided with a prefocus base, and most of them feature a dual filament configuration (Fig. 20(1–3, 7)) for the projection of low and high beams from a single luminaire. The filament for the low beam mode is shielded so as to provide a well-defined horizontal cutoff in the projected beam pattern. Most dual-filament halogen lamps have a clear finish (1), a tinted glass cover (2), or a dichroic filter coating (3) so as to meet specific lighting needs. Headlamps with a single filament (4–6) are designed for compact, highly efficient luminaires.

Automotive lamps aimed at other applications, such as turn lights, stop markers, or interior lighting, are all of the standard type (Fig. 20(8, 9)). Due to less demanding photometric requirements, these lamps are operated at a lower temperature and have a service life reaching up to 1000 h. Although most lamps have a single filament inside a clear bulb, certain applications require a dual filament source or lamps with a colored finish.

Stage and Studio Lamps

Stage and studio lamps are all of the halogen type featuring tungsten filaments with a design optimized for the efficient projection of light in accent and floodlighting

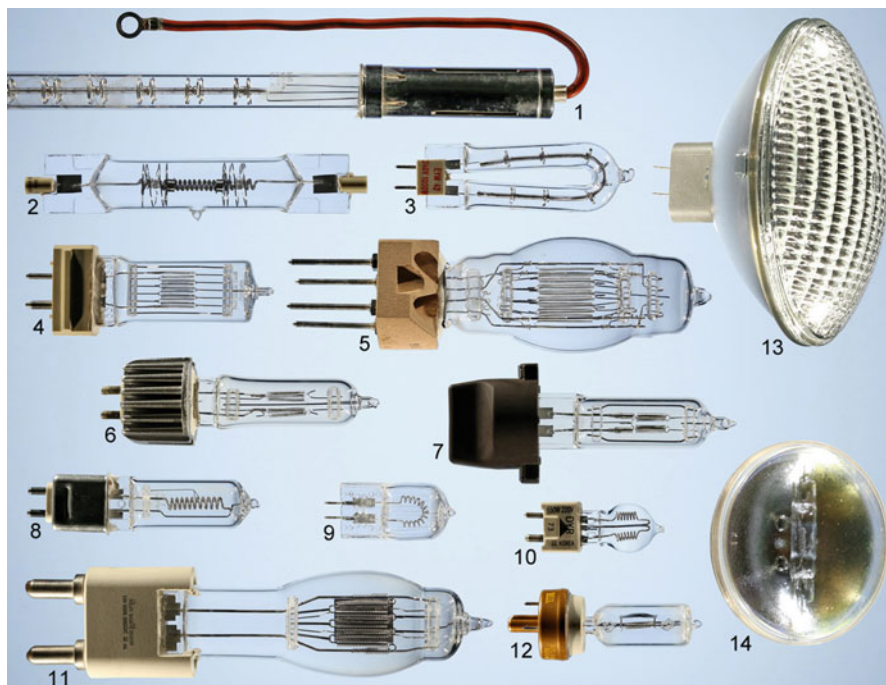


Fig. 21 Range of tungsten-halogen lamps for stage and studio lighting: 5 kW–220 V 3200 K (1), 2 kW–240 V 3200 K with CC filament (2), 1 kW–240 V 3300 K with U-shaped burner (3), 2 kW–220 V 3200 K with monoplane C filament (4), 5/5 kW–240 V 3200 K with dual filament construction (5), 575 W–240 V with barrel-shape C filament configuration (6), 750 W–77 V with compact box-shape quad C filament arrangement (7), 1 kW–240 V 3200 K with axial CC filaments (8), 650 W–120 V 3400 K with U-shaped CC filament (9), 650 W–220 V 3200 K with series-connected CC filaments (10), 5 kW–120 V 3200 K (11) and 500 W–220 V 3200 K (12) both with proximity reflector, 1 kW–120 V PAR203 3200 K (13), and 650 W–120 V PAR114 with a 5000 K dichroic front filter (14). C and CC refer to single-coiled and coiled coil filaments, respectively

applications. Most lamps are designed for mains-voltage operation and are made in three different configurations shown in Fig. 21(1–14): linear double ended (1,2) for floodlighting, compact single ended for floodlighting (3–5) or for precision light projection (6–10), and compact and jacketed reflector (11–14) for efficient and precision floodlighting.

Certain single-ended lamps have a proximity reflector behind the filament grid (11, 12) so as to increase the light output in the direction of projection optics. Double-jacketed reflector lamps (13, 14) are all of the PAR type and are coated with aluminum or silver or with a dichroic mirror. Certain sources have also an additional dichroic filter which raises the light color temperature to 5000 K for daylight simulation.

Tungsten filaments are arranged in four main configurations which have specific photometric properties: linear axial (1–3, 8, 14) for an optical line source,

Table 8 Characteristics of tungsten-halogen lamps for stage and studio lighting

Lamp type	Power (W)	Voltage (V)	Flux (klm)	Efficacy (lm W ⁻¹)	Color temp (K)	Lifetime (h)
Linear double ended	60–10,000	120–230	0.81–255	13.5–34.0	2900–3400	15–3000
Single ended (2900–3000 K)	300–1200	120–240	5.2–27.6	17.3–25.0	2900–3000	50–2000
Single ended (3200 K)	300–20,000	77–240	7.3–580.0	24.3–29.2	3200	50–525
Single ended (3400 K)	150–1000	120–240	4.0–33.0	26.7–33.0	3400	15–75

monoplane (4, 11) for a square extended source, biplane with staggered filament position (12) for a homogeneous luminance with a high output flux, and compact barrel, box, or U shaped, and with double filaments (6, 7, 9 and 10 respectively) for maximum luminance. Certain high-wattage lamps are made with two individual filaments (5) so as to provide three levels of illumination with a constant light color temperature.

Overall, these lamps are classified into three color temperature levels, each suiting different applications: 3400 K for maximum efficacy, 3200 K for an optimum life-efficacy balance, and 2900–3000 K for maximum service life. The characteristics of these lamps are provided in Table 8.

Infrared Radiators

Introduction

Because of their radiative properties, incandescent lamps are best suited for the emission of infrared energy for scientific, industrial, medical, animal care, and cooking applications. Two important IR lamp design factors are the peak spectral emission, defined by the filament structure and temperature, and the irradiance, set by the linear power dissipation. Each application has specific requirements in terms of these parameters and lamps are designed accordingly.

Infrared sources are both of the standard and halogen types and share many design features with general lighting lamps. Most compact quartz sources feature a tungsten-bromine cycle so as to ensure a constant output through life. The filament temperature ranges typically from 1900 to 3000 K, with most radiators operating at 2400–2500 K. These lamps are designed for either an isotropic (Fig. 22) or a directed emission of light (Fig. 23).

Isotropic Infrared Sources

Isotropic infrared sources are made in both double- and single-ended configurations (Fig. 22(1–8) and (9, 10), respectively) and in most cases are intended for mains-voltage operation. With a power dissipation ranging from several hundreds of watts to a few tens of kilowatts, these lamps yield the highest infrared outputs.

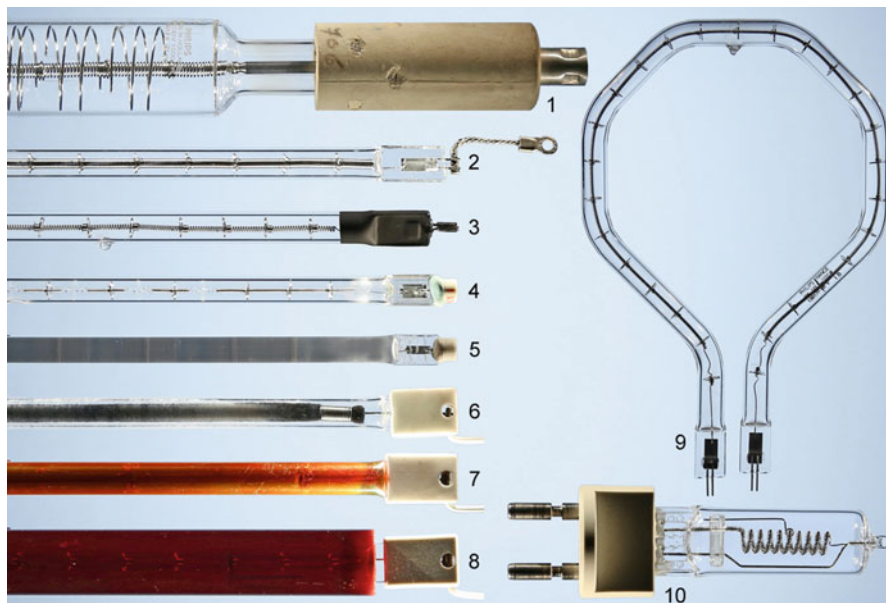


Fig. 22 Range of isotropic infrared sources: 10 kW–220 V 320 W cm⁻¹ (1) and 6 kW–480 V 250 W cm⁻¹ (2) with high power densities, 550 W–220 V 38 W cm⁻¹ (3) and 850 W–220 V 38 W cm⁻¹ (4) with low power densities, 950 W–220 V 26 W cm⁻¹ (5) with hairlined quartz jacket, 1500 W–220 V 28 W cm⁻¹ (6) with carbon felt radiator, 1500 W–240 V 50 W cm⁻¹ (7) with interference filter coating, 500 W–220 V 30 W cm⁻¹ (8) with copper-doped quartz sleeve, 1000 W–125 V (9) with octagonal-shaped tubular burner, and 2 kW–120 V (10) compact single-ended lamp



Fig. 23 Range of reflector infrared sources: 150 W–240 V (1) and 150 W–230 V (2) PAR lamps without and with an iron- or copper-based ceramic glaze, 75 W–120 V (3) and 150 W–15 V (4) with gold coating, and 500 W–220 V 29 W cm⁻¹ (5) with a ceramic half coat

Double-ended lamps are made in a wide range of configurations so as to fit different optical requirements. Very high irradiances are obtained from sources with a linear power dissipation above 100 W cm^{-1} (Fig. 22(1, 2)). Applications requiring lower power flux densities are catered for by lamps rated at below 50 W cm^{-1} , which are often made with a coiled coil filament operating at a lower temperature (3). In this case the quartz lamp has heat-conserving end coats which help maintain a 530 K cold spot temperature, needed for a proper halogen cycle operation. A shorter-wave irradiation at the same low power load is realized with segmented coil filaments (4) which provide an adequate load-temperature balance at higher radiator temperatures.

Incandescent sources optimized for a long-wave emission peaking in the IR-B (1400–3000 nm) or IR-C (3000–10,000 nm) domains often have a modified quartz jacket which acts as a secondary radiator. In this case the quartz infrared emissivity is enhanced by a multitude of hairline bubbles dispersed in the glass (Fig. 22(5)). Some long-wave sources have also a carbon felt radiator (6) operated at low temperature and which features near blackbody-like emission properties.

In applications where lamps are in direct view, visible light is filtered out with a dichroic filter (Fig. 22(7)) or with a copper-doped quartz jacket (8). The optical filters are designed to leak a sufficient amount of visible light so as to prevent staring into the infrared sources and avoid an overexposure at cataract-causing infrared wavelengths.

Single-ended isotropic sources are all halogen lamps made from shaped quartz tubing or with a short single-pinch construction. In the first case, a linear lamp is shaped into a circular form (Fig. 22(9)) so as to fit into cooking heaters. The second lamp type features a compact coiled coil filament (10) with a linear power dissipation reaching 650 W cm^{-1} for the emission of extreme optical power densities needed in high-temperature material processing.

Directed Infrared Sources

Directed infrared sources are of both standard and halogen designs (Fig. 23(1, 2) and (3–5), respectively) with a single- or double-ended construction. Single-ended reflector lamps are usually provided with standard R, PAR, and MR reflectors (1–4), while double-ended sources have a side coating of infrared-reflecting material (5). Reflecting materials include silver and gold in standard reflector lamps and an alumina-based glass enamel for quartz burners. Standard lamps are sometimes provided with a front optical filter so as to reduce the amount of projected visible light (2).

PAR, R, and linear lamps are designed for a mains-voltage operation, while MR sources are built with a 12 V halogen burner. The lamp power rating ranges typically from 150 to 500 W for single-ended sources and can reach several kilowatts in double-ended lamps.

Calibration and Reference Lamps

Incandescent lamps are particularly useful for the calibration of photometric and spectral measuring instruments because their optical output (flux, luminance,

spectrum) can be stable within less than 0.1 % when operated under well-controlled conditions and for short periods of time (i.e., several tens of hours) (Cohue 1966). Special lamps are thus designed to serve as references in photometric parameters such as the luminous flux, luminous intensity, color temperature, monochromatic luminance temperature, spectral distribution, and total optical output. There are two types of reference sources, each with its specific design and construction: flux standards and luminance/spectral standards.

Flux standard lamps have a thermal radiator consisting of a tungsten wire of high homogeneity, designed for the most stable electrical impedance over time. The filaments are always welded to their lead-in wires so as to prevent the risk of changing contact resistance and thermal conduction losses. The operating voltage never exceeds 120 V in order to limit the overall filament length, which is instrumental in ensuring an excellent output stability. These lamps are often made with a large outer jacket so as to limit the effect of filament evaporation on the optical output. Sources intended for in-line measurements on optical benches have a conical-shaped bulb which effectively prevents the reflection of any parasitic light in the line of sight.

There are two main types of flux standard sources: vacuum lamps for a use on optical benches and gas-filled lamps for the calibration of Ulbricht spheres. Vacuum lamps with long hairpin filaments (Fig. 24(1)) operate at a relatively low temperature of 2300 K so as to limit tungsten evaporation and ensure a very high optical stability over time. Lamps made with a shorter filament meander (2) are aimed at luminous flux and color temperature calibrations and operate at a higher temperature, resulting in a light color temperature of 2856 K matching that of the standard CIE illuminant A. Gas-filled lamps have a more conventional construction with a coiled filament arranged in a wreath configuration, as shown in Fig. 24(3).



Fig. 24 Range of optical calibration lamps: 40 W–2.86 A (1) and 72 W–4 A (2) with straight wire filaments, 40 W–100 V with welded single-coil filament (3), 165 W–16/17 A (4) and 96 W–8 A (5), both with a tungsten ribbon radiator

The filament is welded to its lead-in and support wires so as to provide maximum stability over time. All lamps are usually aged for around 100 h before actual use in order to reach proper material, electrical, and optical stabilities as required for the application.

Luminance/spectral standard sources are made with a thermal emitter consisting of a 40 μm -thick tungsten ribbon which provides a homogeneous radiating surface. For stability reasons, the metal ribbon is also welded to its lead wires and is brought to incandescence in the 2300–2500 K range via the circulation of a constant current. Since there is a slight temperature gradient along the ribbon's length due to end losses, a reference point indicator is usually provided inside the lamp in order to ensure consistent measurements.

These lamps are either gas filled or under vacuum depending on the model. Their geometry varies from tubular to complex conical shape (Fig. 24(4) and (5), respectively), the latter being designed to eliminate optical reflections in the direction of measurement. Some sources are also provided with an optically flat quartz window which ensures a very accurate imaging of the incandescent emitter and enables light transmission over a broader spectral range, typically between 250 and 2500 nm (borosilicate glass is suitable for a 300–800 nm transmission range only).

In a more recent development, the calibration/reference incandescent source was miniaturized by using a 1.15 mm-wide double-spiral planar filament structure micromachined from 50 μm -thick tungsten foil. This radiator is mounted in a compact leadless ceramic package, normally used for semiconductor devices, provided with a transparent window (Tuma et al. 2011). This configuration enables the miniature lamp to be soldered onto printed circuit boards for a seamless integration in electronic circuits. The compact source can also be coupled to an optical fiber for a simplified and efficient delivery of the emitted light.

Further Improvements of the Incandescent Lamp Technology

The trade-off between life and efficacy has always been a limiting factor in the design of incandescent lamps, but also a key driving factor in the research and development toward improved light sources. This section presents two types of high-performance lamps which resulted from recent developments: sources provided with means of infrared conservation and compact lamps with high-pressure xenon fill.

Infrared Conservation

Radiation Recycling and Impact on Efficacy

Incandescent lamps owe their low efficacy to a large fraction of their input power converted into infrared energy (see section “[Power Balance of Incandescent Lamps](#)”). Their performances can thus be improved significantly by returning part of the emitted infrared light back toward the filament using a selective mirror

coating (Brett et al. 1980; Köstlin and Frank 1983–1984; Vukceвич 1992). This conservation of optical energy has a significant impact on the overall energy balance of the lamp, and the lumen efficacy expressed in Eq. 9 of section “Overall Efficacy of Incandescent Sources” changes into the following relation (Kauer 1965):

$$\eta_{lm}(T) = \frac{\int_0^{\infty} B_{bb}(T, \lambda) \varepsilon_{em}(T, \lambda) \tau(\lambda) V(\lambda) d\lambda}{\int_0^{\infty} (1 - \rho(\lambda)) B_{bb}(T, \lambda) \varepsilon_{em}(T, \lambda) d\lambda + P_{cd} + P_{cv}} \quad (18)$$

where $\tau(\lambda)$ and $\rho(\lambda)$ are respectively the spectral transmittance and reflectance of the infrared-conserving (IRC) mirror. A suitable optical filter having a high visible transmittance and a high infrared reflectance results in a strongly reduced denominator value in Eq. 18 which leads to a significant increase of η_{lm} .

Vukceвич estimated that an ideal filter transmitting in the 360–830 nm range and reflecting all other wavelengths would result in a gas-filled lamp efficacy reaching 100 lm W⁻¹ at 2800 K (Vukceвич 1992). However, such performance is not realized in practice because of several limitations such as optical losses in the glass jacket, light scattering at the IRC mirror, mismatch between the mirror’s spectral reflectance and the filament’s spectral absorbance, filament misalignment with respect to its infrared image, optical imperfections of the filament, and infrared end losses in the bulb (Vukceвич 1992; Coaton and Marsden 1997). It thus results that IRC lamps require carefully optimized bulb and filament designs so as to ensure the most efficient operation (Levin 1966; Goldstein et al. 1986).

Commercial IRC Lamps

Commercial IRC lamps are all halogen sources provided with a dichroic mirror coated onto the outside bulb surface (Hoegler and McGowan 1984; Bergman 1991). Current IRC mirror structures include several superimposed multilayer stacks which are optimized for a high visible transmittance (90 % between 400 and 760 nm) combined with a broad infrared reflectance (above 75 % between 800 and 2200 nm) (Cottaar 2001). These infrared mirrors enable an efficacy gain up to 50 %, which brings halogen lamps into the B energy efficiency class, meeting the EU stage 6 requirements for nondirectional household lamps (Mekala and Van de Poel 2010).

Commercial IRC halogen lamps are made in four distinct configurations, as shown in Fig. 25(1–7): mains-voltage tubular double-ended with axial filament (1), mains-voltage compact single-ended (2–4), low-voltage compact single-ended (5–6), and mains-voltage integral ballasted with a low-voltage halogen burner (7). The characteristics of these lamps are given in Table 9.

The linear double-ended IRC halogen lamp (Fig. 25(1)) was the first successful commercial incandescent source using infrared conservation (Hoegler and McGowan 1984). This lamp features a krypton-based fill and replaces standard halogen lamps in their sockets while providing 25–40 % energy saving.

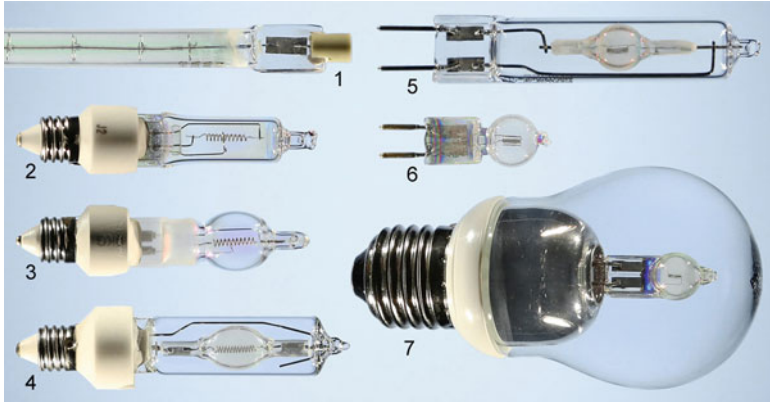


Fig. 25 Range of IRC tungsten-halogen lamps: 900 W–220 V double-ended design (1), 75 W–110 V with tubular burner (2) and 65 W–110 V with elliptical burner (3), both with a supported coiled coil filament, 65 W–110 V (4) and 45 W–12 V (5) with double jacket, 50 W–12 V with spherical burner (6), and an integral ballasted 20 W–220 V lamp with a low-voltage spherical halogen burner (7)

Table 9 Characteristics of tungsten-halogen lamps with infrared-conserving (IRC) coating

Lamp type	Power (W)	Voltage (V)	Flux (klm)	Efficacy (lm W ⁻¹)	Color temp (K)	Lifetime (kh)
Linear, mains voltage	225–900	100–277	5.95–32.0	26.4–35.6	2900–3000	2.0–3.0
Compact, mains voltage	50–90	100–130	1.09–2.40	21.8–26.7	2850–3000	2.0–3.0
Compact, low voltage	20–60	12	0.42–1.70	21.0–28.3	3000	4.0–5.0
Integral ballast, mains voltage	20–30	230	0.37–0.62	18.5–20.7	2900	3.0

Compact single-ended IRC halogen lamps are made for both mains-voltage (Fig. 25(2–4)) and low-voltage operation (5, 6). Mains-voltage lamps have a coiled coil filament and are made in three different configurations: single-ended tubular (2), single-ended elliptical (3), and double-ended elliptical (4). The last type is the most efficient of the three due to the nearly closed configuration of its IRC mirror cavity. Higher performances are realized with low-voltage compact sources (5, 6) due to their smaller filament. Single-ended burners (6) are often coupled with a dichroic MR reflector (see section “[Reflector Tungsten Halogen Lamps for General Lighting](#)”) for the precise projection of light with a low infrared output.

IRC lamps designed for the retrofit of standard GLS lamps are made with standard bases and bulbs. One lamp type combines an electronic transformer and a low-voltage IRC halogen burner into a single unit (Fig. 25(7)). This configuration

yields up to 50 % energy saving and is designed as a lumen retrofit providing similar illumination levels as the GLS lamps they replace.

High-Pressure Xenon Incandescent Lamps

As outlined in section “Gas Fill”, higher gas fill pressures and heavier gases significantly reduce the tungsten evaporation rate and enable higher filament temperatures without affecting the lamp service life. Although such design approach can yield better life-efficacy balances than with a halogen cycle, its implementation was limited by the traditional design of lamps. It is only in 2002 that a high-performance xenon-filled lamp with a 5 bar cold pressure was introduced following changes in lamp design and manufacturing processes (Brüggemann et al. 1998).

This xenon lamp is intended for automotive lighting application and features a compact heavy-walled glass bulb sealed to a sintered glass wafer base. This base, which serves as a reference plane, holds a metal exhaust tube and a high-precision filament-mounting frame. The filament is made in a more compact shape so as to increase the optical efficiency of light projection systems. To ensure the smallest variation possible in optical characteristics, the filament is laser welded onto the frame with maximum precision with regard to its position relative to the lamp base (Brüggemann et al. 1998).

The high-pressure xenon fill combined with an improved lamp processing results in an 18.8 lm W^{-1} lamp efficacy and a longer service life of 1500 h at 3 % failure, with 70 % flux maintenance at 1000 h (Lavaud et al. 2004). Such characteristics make these lamps compatible with sealed-for-life automotive lighting systems.

The xenon lamps are rated for 16, 19, and 24 W operation at 12 V and are made in three configurations, as shown in Fig. 26: omnidirectional color-coated (1) or clear (3) sources for a use in reflector luminaires (Haenen et al. 2002) or lamps provided with a reflector condenser bulb (2) for fiber optics systems (Lavaud et al. 2004).



Fig. 26 Range of 16 W–12 V high-pressure xenon incandescent lamps: omnidirectional source with orange lacquer for turn signals (1), reflector lamp for optical fiber systems (2), and clear omnidirectional source (3) for colored stoplights

Life Expectancy and Failure Mechanisms

The life expectancy of incandescent lamps depends on many parameters related to the lamp properties and applications. This section presents the most important aspects influencing and characterizing service life, such as cost considerations, filament failure mechanisms, survival rate, and anomalous end-of-life failures.

Balance Between Efficacy and Service Life

Since the lumen efficacy and the tungsten evaporation rate both increase with temperature, the design of incandescent lamps requires a compromise between life and efficacy which best matches the requirements set by the intended application. In most cases this balance is adjusted so as to result in the most economical use of the lamp. In a general manner, the most economical life expectancy L_e is calculated from the total cost C per lumen-hour (Valin and Magnien 1991):

$$C = \frac{1}{\eta_{lm}} \cdot \left(\frac{C_{LA}}{P_{LA} L} + \frac{C_{kWh}}{1000} \right) \quad (20)$$

where η_{lm} , C_{LA} , P_{LA} , and L are the lumen efficacy, the initial lamp cost price, the dissipated power, and the life expectancy, respectively, while C_{kWh} is the unit cost of electricity per kilowatt-hour. The optimum life expectancy is found by setting $\partial C/\partial L = 0$ and expressing the lamp efficacy in terms of life expectancy using the following relation: $1/\eta_{lm} = K \cdot L^{1/\delta}$ where K is a constant and δ a parameter relating the lamp life to its efficacy (Valin and Magnien 1991), giving the following expression for L_e :

$$L_e = \frac{1000(\delta - 1)C_{LA}}{P_{LA} \cdot C_{kWh}} \quad (21)$$

with the most accepted δ value for gas-filled lamps around 7 (Coaton and Marsden 1997). In the case of standard GLS lamps, L_e varies between 600 and 1700 h with an average around 1000 h (Valin and Magnien 1991). Although Eq. 21 provides the solution to the minimum cost per lumen-hour for a given lamp type, it does not take into account the labor cost of replacement. If this cost is high, then it becomes more economical to use lamps with a service life greater than 1000 h (Vukcevic 1992). Moreover, certain applications require maximum lamp reliability or very high efficacy, which result either in a very long or in a very short service life, respectively.

Normal Failure Mechanisms

Incandescent lamps fail primarily from filament rupture, which in coiled filament lamps occurs in 70 % of cases from hot spots on the tungsten wire, while the

remaining 30 % has other causes such as microbubble coalescence and recrystallization (Horacsek 1980).

The hot spot mechanism is the first theory devised by I. Langmuir (1915) to explain the finite life of incandescent lamps, and the concept was then developed further (Hörster et al. 1972; Covington 1973b; Harvey 1974). Hot spots emerge from local filament regions operating at a slightly higher temperature due to a defect in the wire or in the coil geometry (Brons 1987). Wire defects include variations in surface state or in diameter (Fischer et al. 1975) and the local increase in ohmic resistance caused by potassium bubble coalescence. It is found that these defects have a lesser impact on lamp life than coil deformations (Horacsek 1980).

The growth of hot spots is mainly driven by temperature gradients which cause the diffusion of tungsten through the Langmuir film and across the coil turns of filaments (Fischer et al. 1975). This process occurs also in halogen lamps because of the lack of operative tungsten cycle within the Langmuir film (see section “[The Halogen Cycle](#)”). In any cases, the higher tungsten evaporation rate at the hot spot location results in a local decrease in wire diameter, which further increases the hot spot temperature. This loop mechanism leads to a runaway process where the local temperature increases exponentially, as shown in Fig. 27, until the filament fails via a local melting of the wire. During this process the average filament temperature decreases over time due to a gradual reduction in mean wire diameter.

At the end of life, the filament has lost an even amount of material, and the total fraction of weight loss is constant for a given lamp type: 10–15 % for vacuum lamps, 3–5 % for single coil gas-filled lamps, and 2–3 % for coiled coil gas-filled lamps (Brons 1987). Since the tungsten vapor pressure at the filament surface defines the mass loss rate (Fischer et al. 1975), it results that lamp life is a direct

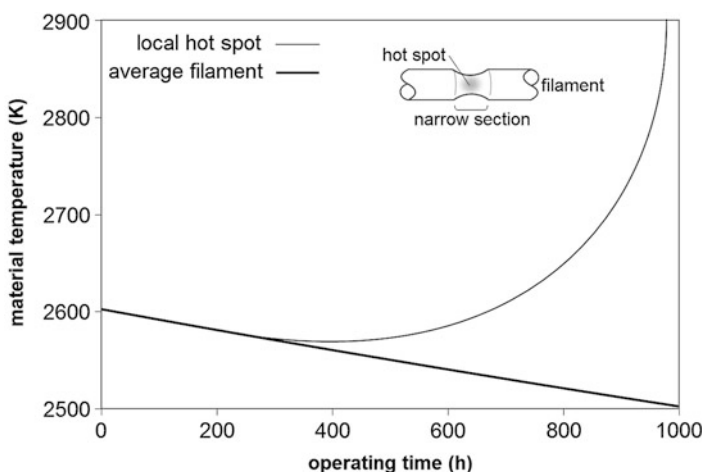


Fig. 27 Evolution over time of the average filament temperature and of the local hot spot temperature in a general lighting service (GLS) lamp (Data from Brons 1987). The insert is a magnified representation of a filament section presenting a hot spot due to a local wire constriction

function of the average filament evaporation rate. This relation leads to the law of fatal weight loss already expressed in Eq. 13, section “[Main Limiting Factor in the Operation of Incandescent Lamps](#)”, where T is the average filament temperature and K_L is also a function of the lamp atmosphere and of the filament quality (the lower the quality, the lower K_L) (Brons 1987).

As mentioned earlier, 30 % of coiled filament failures originate from causes other than hot spots. Some of these causes arise from the coalescence of potassium microbubbles which affects the mechanical properties of the tungsten material and results in wire sagging and in the formation of cracks (Horacek 1980). The other cause is the development of mechanical defects resulting from recrystallization, which is driven by surface and inter-coil tungsten diffusion (Fischer et al. 1975).

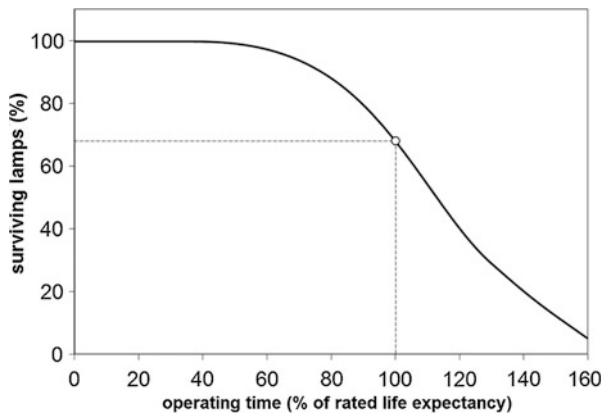
Because the different failure mechanisms interact with each other, the service life of lamps and the main mechanism leading to their failure both depend on the overall filament properties and quality, on the lamp type, and on the operating conditions. In any cases, filament rupture occurs most often at switch-on due to the strong inrush current which causes significant stress to the tungsten wire (Fax et al. 1971).

Survival Rate

The processes leading to filament failure do not cause all lamps to reach the end of life at exactly the specified service lifetime. Variations in gas fill purity and in filament properties (mostly the coil dimensions) affect the local filament temperature and the progress rate of hot spot formation and growth (Brons 1987). It thus results that the time to failure within a large lamp population follows a statistical distribution.

It is found that the percentage of lamps remaining in operation after a given time period follows the evolution shown in Fig. 28 for nominal operating conditions.

Fig. 28 Typical evolution over time of the percentage of GLS incandescent lamps remaining in operation under normal conditions at rated voltage (Data from ELA 1953)



The percentage of failure before half the rated life expectancy is usually negligible and then starts to rise beyond this point with an increasing rate until the 50 % failure point is reached (ELA 1953). Lamps failing at less than 70 % of the rated life are considered premature (Brons 1987). Failures after the 50 % lamp survival point become less frequent and lead to very few long-lived lamps remaining in operation.

Anomalous and Nonpassive Failure Mechanisms

When operating conditions or/and manufacturing quality deviate too much from nominal specifications, then anomalous failure mechanisms arise, and the effective lamp life drops below 70 % of the target life expectancy (Brons 1987). The nature of the main anomalous failure mechanism at play will determine the lamp failure rate and average point in time of occurrence. There are four main causes leading to premature end of life:

- Gaseous contamination from manufacturing defects, such as oil pollution and poor water vapor gettering, or from the development of air leaks, results in severe lamp blackening and in fast filament hot spot development.
- Bulb rupture, which in soft glass lamps, can originate from weather exposure or from residual glass stresses caused by a nonoptimal lamp sealing. In quartz lamps, the deposition of alkali compounds (e.g., from finger grease) or the local overheating can cause severe bulb damages due to cristobalite formation.
- Deviation from optimal voltage in halogen lamps can cause either the blackening of the bulb (undervoltage) or an overactive chemistry leading to metal erosion (overvoltage). In standard lamps the only danger comes from excessive filament evaporation at overvoltage.
- Vibrations can result in the deformation and tangling of the filament when it is hot or in its breakage when it is cold.

Although incandescent lamps are designed to fail in a safe way, a number of situations can result in nonpassive failure. One common end-of-life phenomenon that can cause severe damages is arcing, which occurs when an electrical discharge is pulled between the molten extremities of the parting filament sections after rupture. This plasma ignition then develops into a thermionic arc, as shown in Fig. 29, which significantly increases the lamp temperature and pressure. In the worst case, the arc attachments move toward the lead-in wires, and the plasma effectively short-circuits the filament, resulting in a significant rise in current and power followed by total lamp destruction. In order to prevent such disastrous development, most lamps are provided with an internal fuse so as to limit the drawn current.

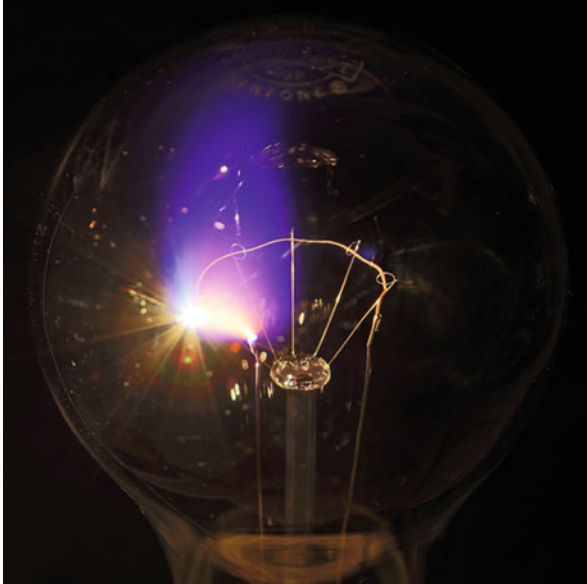


Fig. 29 Arcing in a gas-filled incandescent lamp initiated by filament rupture during operation

Alternative Developments and Conclusion

Incandescent lamps are made in a wide range of configurations and with a large variety of light technical properties so as to best fit in many lighting applications. This technology covers the whole general lighting field and addresses specific lighting needs in traffic signals, motor vehicles, on stages and studios, infrared processing, and other areas. However, the performances of incandescent lamps in visible light applications are limited by the intrinsic nature of the light generation process and by the filament properties at high temperature. The lamp design always involves a life-efficiency trade-off which is the least favorable of all light sources. The lamp performances are nevertheless improved with a tungsten-halogen cycle, with infrared conservation, and with a high-pressure gas fill, but in all cases the life-efficacy product never exceeds 150 klm h W^{-1} , while that of discharge and semiconductor lamps is a factor between 3 and 50 times higher.

The intrinsic limitations of incandescent lamps triggered and drove many research and development efforts aimed at improving the technology. Some of the investigated approaches included incandescence at melting point temperature and above in cluster lamps (Weber and Scholl 1993), a tungsten-fluorine regenerative cycle (Elenbaas 1972; Dettingmeijer et al. 1975), up-conversion of infrared energy into visible light (Hewes and Sarver 1969; Vukceovich 1992), luminescent

vapor atmospheres (Moore and Chamberlain 1968; Johnson 1970), candoluminescence (Thorington 1960; Ivey 1974), and the use of strongly selective thermal radiators. The investigated selective emitters included infrared-transparent materials (Kauer 1965; Kane and Sell 2001), infrared-reflecting ceramics (Kauer 1965; Milewski and Milewski 1989), photonic structures (Waymouth 1989; Sommerer et al. 2007), and carbon nanotubes (Mann et al. 2006; Shu et al. 2007).

Unfortunately, all these alternative developments failed to deliver commercially viable products, and the efficacy of incandescent lamps was never improved past 36 lm W^{-1} . As a result, incandescent lamps are being progressively replaced in a number of applications by other, more efficient light source technologies. However, there are some fields of application where thermal radiators will still prevail in the future, such as infrared processing, spectroscopy, and instrument calibration.

References

- Beesley EM, Makulec A, Schroeder HH (1963) Lamps with multilayered interference-film reflectors. *Illum Eng* 58(5):380–385
- Bergman RS (1991) Halogen-IR lamp development: a system approach. *J Illum Eng Soc* 20(2):10–16
- Brett J, Fontana RP, Walsh PJ, Spura SA, Parascandola LJ, Thouret WE, Thorington L (1980) Radiation-conserving incandescent lamps. *J Illum Eng Soc* 9(4):197–203
- Brett J, Fontana RP, Walsh PJ, Spura SA, Parascandola LJ, Thouret WE, Thorington L (1981) Development of high energy-conserving incandescent lamps. *J Illum Eng Soc* 10(4):214–218
- Bright AA (1949) *The electric-lamp industry: technological change and economic development from 1800 to 1947*. The MacMillan company, New York
- Brons H (1987) *Physics of incandescent lamps*. Philips Lighting, Weert
- Brüggemann U, Kohlmann WL, Geboers MJM, van Hees AJM (1998) Electric incandescent lamp. WO Patent 98/50942
- Carlson VH, Hurd DT (1965) Tungsten-rhenium alloy filament lamps for improved vibration service life. *Proc Nat Tech Conf Illum Eng Soc* (10):1–5, New York
- Clapp R (1950) Thermionic emission in gas-filled incandescent lamps. *Illum Eng* 45(6):357–362
- Clark CN (1966) Characteristics of incandescent lamps for theatre stages, television and film studio. *Illum Eng* 61(7):464–474
- Coaton JR, Marsden AM (1997) *Lamps and lighting*, 4th edn. Arnold, London
- Cohue M (1966) *Sources lumineuses*. Masson et Cie, Paris
- Coolidge WD (1913) Tungsten and method of making the same for use as filaments of incandescent electric lamps and for other purposes. US Patent 1,082,933
- Cottaar EJE (2001) Electric lamp and interference film. WO Patent 01/57913
- Covington EJ (1968) The Langmuir film model in incandescent lamps. *Illum Eng* 63(3):134–142
- Covington EJ (1973a) The life-voltage exponent for tungsten lamps. *J Illum Eng Soc* 2(2):83–91
- Covington EJ (1973b) Hot spot burnout of tungsten filaments. *J Illum Eng Soc* 2(4):372–380
- Déribéré M (1965) *Lampes à iode, lampes à iodures*. Dunod, Paris
- Dettingmeijer JH, Dittmer G, Klopfer A, Schröder J (1975) Regenerative chemical cycles in tungsten-halogen lamps. *Philips Tech Rev* 35(11/12):302–306
- Electric Lamp Association (ELA) (1953) *Electric lamps*. The British Empire Lighting Service Bureau, London
- Elenbaas W (1963) Rate of evaporation and heat dissipation of heated filament in a gaseous atmosphere. *Philips Res Rep* 18:147–160
- Elenbaas W (1972) *Light sources*. Crane, Russak & Company, New York

- Fax DH, Sell HG, Stickler R (1971) Transients in incandescent lamp filaments containing defects. *Illum Eng* 66(4):187–195
- Fischer E, Fitzgerald J, Lechner W, Lems W (1975) Transport and burn-out in incandescent lamps. *Philips Tech Rev* 35(11/12):296–302
- Gendre MF (1999) Microscopy analysis of HID lamp parts. University of Rhode Island, Kingston
- Goldstein IS, Fontana RP, Thorington L, Howson RP (1986) The design, construction and performance of an incandescent light source with a transparent heat mirror. *Light Res Tech* 18(2):93–97
- Haenen LJJ, Ansems J, Schuurmans J, de Montureux P (2002) New slim automotive taillight using HiPerVision lamps. *Proc SPIE* 4775:135–144, Seattle
- Harvey FJ (1974) Failure of incandescent tungsten filaments by hot spot growth. *J Illum Eng Soc* 3 (4):295–302
- Henderson ST, Marsden AM (1972) *Lamps and lighting*, 2nd edn. Edwards Arnold, London
- Hewes RA, Sarver JF (1969) Infrared excitation processes for the visible luminescence of Er³⁺, Ho³⁺, and Tm³⁺ in Y³⁺-sensitized rare-earth trifluoride. *Phys Rev* 182(2):427–436
- Hoegler LE, McGowan TK (1984) Practical high efficiency tungsten-halogen lamps using IR reflecting films. *J Illum Eng Soc* 14(1):165–174
- Horacek O (1980) Properties and failure modes of incandescent tungsten filaments. *IEE Proc A* 127(3):134–141
- Hörster H, Kauer E, Lechner W (1972) A concept for the burning-out mechanism of an incandescent tungsten wire. *J Illum Eng Soc* 1(4):309–317
- Hummel RE (1993) *Electronic properties of materials*, 2nd edn. Springer, New York
- Ivey HF (1974) Candoluminescence and radical-excited luminescence. *J Luminescence* 8:271–307
- Johnson PD (1970) Efficient incandescent light source including light-enhancing metallic iodide vapor. US Patent 3,497,754
- Kane R, Sell H (2001) *Revolution in lamps: a chronicle of 50 years of progress*, 2nd edn. The Fairmont Press, Lilburn
- Kauer E (1965) Generating light with selective thermal radiators. *Philips Tech Rev* 26(2):33–47
- Köstlin H, Frank G (1983–1984) Thin-film reflection filters. *Philips Tech Rev* 41(7/8):225–238
- Langmuir I (1915) The melting point of tungsten. *Phys Rev* 6(2):138–158
- Langmuir I (1916a) The characteristics of tungsten filaments as functions of temperature. *Phys Rev* 7(3):302–330
- Langmuir I (1916b) Incandescent electric lamp. US Patent 1,180,159
- Lavaud JF, Pelsma J, Haenen L (2004) HiPerVision reflex lamp: a new member of the HiPerVision family. In: *Proceedings of 10th international symposium on science and technology of light sources*, Toulouse, pp 493–494
- Law JR (1965) Multilayer filters. *Illum Eng* 60(10):603–608
- Levin RE (1966) An investigation of improving incandescent projection sources. *Proc Nat Tech Conf Illum Eng Soc* (6):1–10, Minneapolis
- Mandel L, Wolf E (1995) *Optical coherence and quantum optics*. Cambridge University Press, New York
- Mann D, Kato YK, Kinkhabwala A, Pop E, Cao J, Wang X, Zhang L, Wang Q, Guo J, Dai H (2006) Electrically driven thermal light emission from individual single-walled carbon nanotubes. *Nat Nanotechnol* 2(1):33–38
- Mekala SR, Van de Poel G (2010) Modeling performance limits for tungsten halogen energy saving lamps using infra-red reflecting films. In: *Proceedings of 12th international symposium on science and technology of light sources*, Eindhoven, pp 117–118
- Milewski JV, Milewski PD (1989) Single crystal whisker electric light filament. US Patent 4,864,186
- Moore JA (1958) History of the tungsten filament lamp. *GEC J* 25(4):174–188
- Moore JA, Chamberlain PFW (1968) Improvements in or relating to incandescent electric lamps and their operation. GB Patent 1,105,291

- Moore JA, Jolly CM (1962) The quartz-iodine tungsten lamp: mechanism, design and performance. *GEC J* 29(2):99–106
- Pacz A (1922) Metal and its manufacture. US Patent 1,410,499
- Pearson CW, Linsday EA, Dorsey RT (1956) Characteristics and applications of axial filament lamps. *Illum Eng* 51(12):782–790
- Schirmer H, Stober I, Friedrich J (1967) Über die Methode von Langmuir zur theoretischen Behandlung gasgefüllter Glühlampen. *Tech Abhan der Osram-Ges* 9:125–136
- Shu QK, Wei JQ, Wang KL, Li CG, Jia Y, Wu DH (2007) Low voltage energy-saving double-walled carbon nanotube electric lamps. *J App Phys* 101(8):084306
- Smithells CJ (1936) Tungsten – a treatise on its metallurgy, properties and applications, 2nd edn. Chapman & Hall, London
- Sommerer TJ, Meschter PJ, Midha V, Minnear WP, Bryan DJ (2007) Article incorporating a high temperature ceramic composite for selective emission. US Patent 2007/0228951
- Stoer GW (1986) History of light and lighting. Philips Lighting, Eindhoven
- Suits CG, Way HE (1960) The collected works of Irving Langmuir, volume 2: heat transfer – incandescent tungsten. Pergamon Press, Oxford
- T’Jampens GR, van de Weijer MHA (1966) Gas-filled incandescent lamps containing bromine and chlorine. *Philips Tech Rev* 27(7):173–179
- Thorington L (1960) Electric lamp. US Patent 2,920,222
- Thouret WE, Kaufman R, Orlando JW (1975) Energy and cost saving krypton filled incandescent lamps. *J Illum Eng Soc* 4(3):188–197
- Tuma ML, Collura JS, Helvajian H, Pocha MD, Meyer GA, McConaghy CF, Olsen BL, Hansen WW (2011) Ultraminiature broadband light source and method of manufacturing same. US Patent 2011/0006663
- Valin J, Magnien M (1991) *Les sources de lumières*, 2nd edn. Société d’Edition LUX, Paris
- van den Hoek W (2010) Notes on the history of incandescent lamps. In: Proceedings of 12th international symposium on science and technology of light sources, Eindhoven, pp 3–11
- van den Hoek WJ, Jack AG (1990) Lamps. *Ullmanns Encycl Ind Chem* A15:115–150
- Vukceovich MR (1992) The science of incandescence. Nela Press, Cleveland
- Waymouth JF (1987) Light sources. *Encycl Phys Sci Technol* 7:225–257
- Waymouth JF (1989) Where will the next generation of lamps come from. *J Light Vis Env* 13(2):51–68
- Weber B, Scholl R (1993) A new kind of light-generation mechanism: incandescent radiation from clusters. *J App Phys* 74(1):607–613
- Yuge Y (1995) Review of optical coatings for incandescent and other lamps. In: Proceedings of 5th international symposium on science and technology of light sources, New York, pp 221–230
- Zubler EG, Mosby FA (1959) An iodine incandescent lamp with virtually 100 per cent lumen maintenance. *Illum Eng* 54(11):734–740

Low-Pressure Gas Discharge Lamps

Graeme Lister and Yang Liu

Contents

Lamps with Electrodes	1066
The Physics of Low-Pressure Discharge Lamps	1066
Electroded Fluorescent Lamps	1069
Low-Pressure Sodium Lamps	1075
References	1077

Abstract

This chapter provides an overview of the current technology in low-pressure discharge lamps. Because of their dominance in the market place, and production of “white” light, the major part of the chapter is devoted to fluorescent lighting. Fluorescent lamps (FL), contain mercury, a highly efficient emitter of UV radiation, which is then converted to visible radiation by a phosphor coating on the lamp. Although low-pressure sodium (LPS) lamps, which use sodium as the principal radiation source, are more efficient emitters of visible radiation than FL, they are only suitable for limited outdoor applications, due to their predominant yellow colour and hence poor colour rendering. Low pressure discharges in rare gases such as neon have been used for specialist lighting applications, but they are currently being superseded by Light Emitting Diodes (LEDs).

Low-pressure discharge (*LPD*) light sources emit radiation as a result of the excitation of atoms and molecules into radiation-emitting states by electrons which are far from

G. Lister (✉)

Graeme Lister Consulting LLC, Wenham, MA, USA

e-mail: graeme_lister@graemelisterconsulting.com; graeme.lister@yahoo.com

Y. Liu

Fudan University, Shanghai, China

e-mail: ly@fudan.edu.cn

local thermal equilibrium (*LTE*) with the other species in the discharge. This chapter provides an overview of the current technology in low-pressure discharge lamps. The most commonly available *LPD* lamps are fluorescent lamps (*FL*), which use mercury as the principal source of radiation. Because of their dominance in the market place, and production of “white” light, the major part of the chapter is devoted to fluorescent lighting. Although low-pressure sodium (*LPS*) lamps, which use sodium as the principal radiation source, are highly efficient light sources, they are only suitable for limited outdoor applications, due to their predominant yellow color.

FL are filled with rare gas, typically argon at around 3 Torr (400 Pa) pressure, with a few mTorr (0.5–5 Pa) of mercury vapor. At room temperature, mercury has the highest vapor pressure of any of the elements suitable for producing radiation. Under optimum conditions, 70–75 % of electrical power in these discharges is converted to *UV* radiation by mercury atoms. The *UV* is then converted to visible light by means of a phosphor; the energy difference between the incoming *UV* photon and the outgoing visible photon (the Stokes shift) results in a total conversion efficiency of electrical power to visible light of about 25 %.

In a conventional fluorescent lamp, the energy required to supply the light source is provided by an alternating electric current, maintained between electrodes within the lamp. The presence of electrodes places severe restrictions on lamp design and is a major cause of failure, limiting lamp life. Since the early 1990s, a number of “electrodeless” fluorescent lamps have been introduced into the market place, and they will be discussed in chapter “► [Electrodeless Lamps and UV Sources.](#)”

LPS lamps operate on a similar principle to fluorescent lamps, with sodium replacing mercury and containing neon at around 8 Torr (1 kPa) pressure as buffer gas. These lamps produce monochromatic yellow light and do not require a phosphor. *LPS* lamps are therefore much more efficient in converting electrical power to visible light, but they can only be used for applications where color discrimination is not considered important, such as some street and outdoor area lighting. Further, the optimum vapor pressure of sodium is obtained at 260 °C, which means that the lamp requires good thermal insulation and glass which is resistant to sodium at the operating temperature.

Rare gas atoms are emitted in the *VUV* and the visible spectrum, but the efficiency of converting electrical energy to radiation is significantly less than in *FL* and *LPS* lamps. There have been a limited number of lighting applications for rare gas discharges, such as the red line in neon for automobile tail lights, but in general, these applications have been superseded by *LEDs*.

Lamps with Electrodes

The Physics of Low-Pressure Discharge Lamps

Introduction

The discharge in a conventional *LPD* lamp is maintained between two electrodes. Electrons emerging from the cathode are accelerated through the cathode fall into a region of relatively weak electric field, the negative glow. There is an overproduction

of ions in the negative glow, which is compensated by a dark region of low ionization, the Faraday dark space. Following the Faraday dark space is a region of constant electric field, the positive column, which produces almost all the light. A bright anode glow is separated from the positive column by an anode dark space. All regions in the neighborhood of electrodes contribute to inefficient use of power, and this effect is minimized by ensuring that the positive column is much longer than the other regions of the discharge. “Electrodeless” lamps in which the discharge is maintained by external radio frequency (*rf*) electric fields are discussed in chapter “► [Electrodeless Lamps and UV Sources](#)”.

Collisions of electrons with atoms and molecules are the dominant mechanism of energy coupling in low-pressure discharge lamps. Elastic collisions of electrons with atoms and ions couple the electric power to the discharge through the electrical conductivity and provide gas heating, while electron–electron collisions redistribute the electron energy and strongly influence the electron energy distribution function (*EEDF*). Collisional processes between atoms in excited states and other atoms and molecules in the discharge (such as chemi-ionization) can also play an important role, quenching radiative states and providing an extra channel for ionization, which influences the current–voltage characteristics.

Radiation Mechanism

Atoms with a radiative transition to the ground state (resonance transition) from an energy level which is close to half the ionization energy can provide efficient radiation in low-pressure discharges. The radiative levels (and neighboring metastable levels) are more readily populated by electron impact excitation than higher levels, providing efficient channels for both radiation and two-step ionization to maintain the discharge. Mercury is an efficient radiator because the energy level of the first excited state for resonance radiation is 4.89 eV, and the ionization energy is 10.4 eV. Sodium (2.1 and 5.1 eV) and barium (2.2 and 5.2 eV) have similar properties but require higher temperatures to maintain the required vapor pressure. Rare gas atoms have radiative states which are much closer to the ionization level and therefore are less efficient radiators. For example, neon has levels for *VUV* resonance radiation at 16.7 and 16.9 eV and an ionization level of 21.6 eV.

Spectral lines emitted by atoms are broadened and shifted by a number of perturbing influences (Lister et al. 2004; Molisch and Oehry 1998). These considerations are quantified through the radiation transport equation (Eq. 1). For simplicity, we consider the variation in spectral radiance $\Gamma(\nu, s)$ of a ray of frequency ν along a “line of sight” s in the medium, such that all the plasma parameters depend only on the position s along the line:

$$\frac{d\Gamma(\nu, s)}{ds} = \varepsilon(\nu, s) - \alpha'(\nu, s)\Gamma(\nu, s) \quad (1)$$

where $\varepsilon(\nu, s)$ is the spectral emission coefficient, the spontaneous transition rate per unit volume, per hertz of bandwidth at frequency ν , and $\alpha'(\nu, s)$ is the effective absorption coefficient, which includes all the important collision processes in the

discharge. The *optical depth* $\tau = \alpha'(\nu, s)d$, where d is the discharge diameter, is a dimensionless parameter describing the average number of collisions experienced by a photon prior to it escaping from the discharge.

The most important absorption processes in *LPD* lamps are Doppler broadening, due to thermal motion of the radiating atom, and collisional resonance broadening resulting from perturbation of an atom radiating to the ground level by an identical atom in the ground level. Radiation emitted at one point in the discharge may be absorbed and reemitted many times before reaching the walls, broadening the radial density profiles of radiating states and consequently influencing the spectral output and power balance of the lamp. Radiation from the wings of each spectral line is less trapped than that from the center of the line, and for strongly absorbed lines, the major contribution to the radiation emitted from the discharge is from the line wings. Radiation transport also influences other plasma properties, such as the maintenance electric field and the *EEDF*.

Energy Balance

The total electrical power W_{elec} in a gas discharge is dissipated through radiation (W_{rad}), heat conduction (W_{heat}), diffusion of particles from the discharge (W_{diff}), and acceleration of ions in the sheaths at the walls and electrodes (W_{sheath}), i.e.,

$$W_{\text{elec}} = W_{\text{rad}} + W_{\text{heat}} + W_{\text{diff}} + W_{\text{sheath}}. \quad (2)$$

Gas discharges for lighting are designed to maximize the fraction of electrical power emitted as visible radiation or *UV* radiation that can be converted to visible radiation by a phosphor.

Elastic collisions of low-energy electrons with atoms and ions couple the electric power to the discharge through the electrical conductivity and provide gas heating, while electron–electron collisions redistribute the electron energy. The efficiency of converting electrical power to visible or *UV* radiation depends on the fraction of high-energy electrons in the *EEDF* available to excite mercury atoms.

Charged particle diffusion plays an important role in low-pressure discharge lamps by influencing the spatial distribution of atoms and molecules. Ion diffusion to the walls also represents an important loss mechanism in the power balance in the positive column. Electrons are much more mobile than ions, and the ambipolar space-charge field (Lister et al. 2004; Waymouth 1971) is established to maintain an equal radial flow of ions and electrons to the walls and preserve charge neutrality. Ions are thus accelerated away from the center of the discharge, while the electron motion is retarded.

The rare buffer gas in a fluorescent lamp serves a number of functions (Lister et al. 2004; Waymouth 1971):

- (i) It controls the ambipolar diffusion rate, which in turn establishes the required balance between ionization of mercury atoms to maintain the discharge and excitation of mercury to provide *UV* radiation.

- (ii) It determines the lamp voltage, since the electrical conductivity depends principally on the elastic electron collisions with the buffer gas atoms.
- (iii) It acts as a buffer to reduce ion bombardment on electrodes, and where applicable, the phosphor coating, and also reflects and redistributes sputtered emitter material back on the cathode, prolonging lamp life.
- (iv) It helps in lamp starting by reducing the voltage required to initiate the discharge.

In a discharge containing a minority species with an ionization energy less than that of the atoms of the buffer gas (such as mercury or sodium), ambipolar diffusion leads to a depletion of the minority species at the center of the discharge (*cataphoresis*). Radial cataphoresis can play a role in fluorescent lamps, particularly for high current densities (Curry et al. 2002), and is a dominant process in low-pressure sodium lamps (Jack and Vrenken 1980). Axial cataphoresis also occurs along the electric field of direct current discharges, so lamps must operate with alternating current to prevent the minority (radiating) species from accumulating at the cathode.

Electroded Fluorescent Lamps

Introduction

Fluorescent lamps are filled with a noble gas (usually argon, krypton or neon, or a combination thereof) at a pressure of a few Torr, with a minority of mercury (typically a few mTorr). They operate at gas temperatures of 300–700 K, with electron temperatures $\sim 11,000$ K (~ 1 eV). 60–75 % of the electrical power supplied to the discharge is converted to *UV* radiation, which is subsequently converted to visible light by a phosphor applied to the inner wall of the lamp. The overall efficiency of converting electrical power in the discharge to visible light is ~ 25 %.

FL are typically long and thin; very short *FL* are inefficient because the electrode energy loss is significant compared to the energy converted to light. A standard 40 W T8 lamp has a diameter of 24 mm (the notation T n refers to a tubular diameter of $n/8$ in.) and a length of 120 cm. Compact fluorescent lamps (*CFL*) are made by folding the discharge tube, often with electronics that allow them to fit into a regular incandescent lamp socket. “Electrodeless” *FL* (see chapter “► [Electrodeless Lamps and UV Sources](#)”) have been developed over the last 20 years, with considerable increase in lamp life.

FL operate at wall loadings of 0.04 – 0.2 W/cm² – two orders of magnitude less than those of *HID* lamps. This is another reason why *FL* discharges are long, in order to provide the same lumen output as *HID* lamps. However, since *FL* require no outer jacket, overall they produce 2,000–4,000 lm/m of light output, which is only a factor of 10 below that of *HID* lamps. Due to the wide variety of available phosphors, this family of lamps can provide a very wide range of *CCT* and *CRI*, with no significant modifications to the discharge itself.

Radiation Mechanism

Discharge Plasma

Typically, 60 % of the electrical power in the positive column is dissipated in resonance radiation from the $6p^3P_1$ level of mercury, with a wavelength of 254 nm. A second resonance state (from $6p^1P_1$) emits *UV* radiation at 185 nm. Resonance radiation in mercury is a special case, due to the five isotopes in natural mercury, resulting in a complex lumped line profile which reduces the radiation trapping of these lines (Maya and Lagushenko 1990). There are also a number of visible emissions (mainly in the blue and green) from higher excited states radiating to lower levels. These visible lines are particularly important at high current densities and influence the choice of phosphor to optimize efficacy and color temperature.

The radiation transport in *FL* is described by Eq. 1. The center of the 254 nm line in the discharge has an optical depth τ of 50–100, while for some of the stronger visible and near *UV* lines $\tau \approx 1$, and for other emitted lines $\tau \ll 1$. There is generally little or no radiation from noble gas atoms. The large optical depth of the 254 nm resonance line is due to the fact that the lower level of the transition is the ground state, which is occupied by 99 % of the mercury atoms present. This has the effect of extending the effective radiation lifetime of the photons, as they collide and reemit with ground-state atoms, thus increasing the chance for the dissipation of excitation energy by non-radiative processes. As the density of mercury atoms increases, the total emitted radiation at each point in the discharge also increases. However, increased mercury density also results in an increase in imprisonment time of radiation and therefore an increase in energy dissipation in non-radiative processes. The optimum density (or mercury vapor pressure) is achieved by balancing these two competing mechanisms. In a conventional T8 *FL*, the maximum efficiency of producing 254 nm radiation occurs when the saturated mercury vapor pressure is ~ 6 mTorr, corresponding to a “cool spot” temperature on the wall of ~ 40 °C.

Phosphors

UV radiation reaching the walls of the discharge tube is absorbed and converted to visible light by coatings of luminescent phosphor powders applied to them. These phosphors are composed of inorganic crystalline materials, which have been synthesized to incorporate specific luminescent centers to absorb the desired wavelengths of *UV* light and emit visible rays. This process results in the emission of a photon of lower energy than the absorbed photon (the Stokes shift) with the difference in energy between the two photons dissipated as heat in the phosphor.

FL became commercially viable in 1941, with the discovery of calcium halophosphate phosphors, which emit over a broad band of visible wavelengths. Although these phosphors yield a relatively poor *CRI* of 50–75, they are inexpensive and therefore still used in some *FL* products today. The discovery of rare-earth phosphors in the mid 1970s, with narrow emission bands at 610 nm in the red, 545 nm in the green, and 450 nm in the blue, led to the development of three band

phosphors with CRI between 80 and 85. Multiband phosphors, with CRI above 90, have been developed since that time, but the improved CRI is at the expense of lower efficacy.

The quantum efficiency of converting UV to visible light in most phosphors is close to 1; hence, the efficacy of FL is limited by the Stokes shift. There has been some research activity to develop phosphors with quantum efficacies greater than one, but the only success to date has been for VUV radiation, which has an even larger Stokes shift and is therefore impractical.

Energy Balance

Discharge Plasma

Under optimum operating conditions, the positive column of a T8 lamp converts about 70 % of electrical energy to UV radiation, the remainder being dissipated as gas heating (20 %) and losses due to particles diffusing to the wall (~10 %). The radiation efficiency of these discharges decreases with increasing radius, primarily because the electron temperature also decreases (Ingold 1991). Gains in efficacy can be obtained by using oval or rectangular shapes, since the electron temperature is maintained as the cross section is increased. However, for aspect ratios greater than 3:1, the discharge constricts along the major axis and does not fill the tube. This precludes the use of fluorescent lamps in flat-panel applications.

FL are arc discharges, because the electric current required to maintain the discharge is provided by thermionic emission at the cathode. The regions of the discharge in the vicinity of the electrodes dissipate energy which is not converted to light and therefore represent a reduction in lamp efficacy.

The negative glow extends for about one tube radius on either side of the cathode, and the Faraday dark space extends for a length slightly smaller than the tube diameter (Waymouth 1971). No cathode dark space is visible, because the cathode sheath is extremely thin (~0.1 mm) and electrons from the cathode (*beam* electrons) enter the negative glow with the full energy of the cathode fall.

Since the anode does not emit ions, all the currents crossing the anode surface are carried by the electrons. If the required current to the anode I is more than the random electron current at the anode surface I_T , the anode charges positively with respect to the main discharge in order to extract the necessary current. Conversely, if $I_T > I$, the anode charges negatively (Waymouth 1971). The anode potential with respect to the plasma (the anode fall) depends on the physical properties of the anode.

Lamp-Electric-Circuit System

During steady-state operation, fluorescent lamps, in common with most gas discharges, have a “negative static differential resistance” – as *rms* discharge current increases, the *rms* voltage required to maintain the discharge decreases. This situation is inherently unstable when driven with a constant voltage source and would allow the current to grow unimpeded. The electric circuit device that supplies the necessary impedance to restore stability is called a *ballast*.

Most fluorescent lamps today operate on electronic ballasts, supplying current to the discharge at a frequency of typically 25–60 kHz. Note that at these frequencies ($f \gg \tau_D^{-1}$), the instant differential resistance is positive, but the static differential resistance remains negative and a ballast is still required. The cathode fall is reduced and the anode fall eliminated at this frequency; hence, power losses near the electrodes are reduced compared to 50/60 Hz operation. High-frequency operation also enhances the high-energy tail of the *EEDF* leading to increased radiation efficiency in the positive column. In practice, the gain of efficacy with increasing frequency appears to saturate at 15 kHz (Guest and Mascarenhas 1997).

The cathodes in a fluorescent lamp are multi-coiled helices of tungsten, the interstices of which are impregnated with alkaline-earth oxides to enhance electron emission. During normal operation, they are heated by the passage of current through the tungsten wire and by ion bombardment from the plasma. The presence of excess barium dissolved in the mixed oxide crystals and at the surface makes the oxides semiconducting at typical operating temperatures and reduces the work function of the cathode, allowing them to supply the current to the discharge at an operating temperature of 1,200–1,400 K. The current at the cathode contracts to a diameter of less than 1 mm (the “spot” mode), and the position of the spot varies during the life of the lamp as the emissive material is locally evaporated and sputtered.

In common with all discharge lamps, *FL* require a higher applied voltage to ignite the lamp than during steady-state operation. This can be achieved in three ways. The simplest ignition method is the switch–start system, used in most countries besides the USA. The glow starter switch is the most common of these methods: a small glass bulb containing two bimetal contacts and filled with a low-pressure gas is mounted, together with a capacitor, in a container with two contact pins. Before switching on, the bimetal contacts are separated by a small gap, but on ignition, a glow discharge forms and the contacts heat and bend toward each other until they touch. This extinguishes the glow discharge and provides a relatively large current through the ballast and electrodes, which warms the cathode coils. Since the circuit is inductive, a high-voltage pulse (600–1,500 V) is produced between the lamp electrodes, ionizing the gas and allowing current to flow in the lamp. During steady-state operation, the lamp voltage is too low to cause a discharge in the starter switch. The glow bottle can be fitted into a socket and is inexpensive, and it can be easily replaced.

In the USA, either rapid start or instant-start systems are used. In rapid start ballast systems, individual filament transformers wound on the ballast core continuously supplying 3.75 V to the cathodes, and the resulting current heats the cathode sufficiently for electrons to be emitted and the discharge ignited. This represents a considerable system loss – 2.5 W in a typical T8 *FL*. The second, and more common method, is to use an instant-start electronic ballast, which supplies a high voltage across the lamp terminals when the lamp is switched on and the electrodes are cold. The high voltage causes increased sputtering of materials from the electrodes, with a consequent reduction in lamp life but an improved efficacy compared to rapid start ballasts.

Material Limitations

FL have rated lives of 40,000 to 100,000 h on rapid start ballasts. *Average rated life* is the number of hours at which half of a large sample of lamps has failed, which is the median life of the group. The standard operating cycle for this test, as defined by the Illuminating Engineering Society, is 3 h on, 20 min off. However, for a typical 8–9 h workday in which lamps are operated continuously, median lamp life may double the rating reported in the lamp catalogs. On the other hand, reductions in rated life by as much as 25 % may occur when the lamps are operated using instant-start ballasts (National Lighting Product Information Program 2006).

Lamp life is primarily determined by the erosion of electron emission material from the cathodes. During ignition, there is a momentarily enhanced erosion of the cathodes, and frequent cycling of the lamps is deleterious to lamp performance. Lamps that are intended to operate with many ignitions per hour are operated on ballasts supplying continuous filament heat, even when the lamp is not operating, to minimize damage to the electrodes when starting.

A major material limitation in *FL* operation is the degradation of phosphors. During lamp operation, phosphors deteriorate as a result of photolytic decomposition and color center formation, ion bombardment and chemical reactions with mercury, and glass and impurity gases. Phosphor deterioration increases with wall loading and wall temperature and depends on the type of phosphor used. The phosphor is also discolored due to sputtering of electrode material during starting. Standard cool white halo-phosphate *FL* lose ~20 % of their output over 8,000 h of operation, and this is reduced to ~10 % by the use of triphosphate phosphors. For this reason, most compact fluorescent lamps (*CFL*) intended as replacements for incandescent lamps use triphosphate phosphors, although the cost is several times that of conventional phosphors.

Fabrication Technology

FL fabrication is highly automated, with almost no manual handling. The principal manufacturing processes require the application of a uniform coating of phosphor to the inside walls of the tubular bulb, the processing of electrode materials, and the evacuation of very large lamp volumes through exhaust tubes of small diameter (Lister et al. 2002).

The phosphor is manufactured by blending raw materials of controlled particle size and firing at high temperature to produce the required compounds. The phosphor is applied to the bulb in the form of a paint with the phosphor as a pigment, flow coating of the inside surface of the bulb to a thickness controlled by viscosity and drying. The remaining binder is then removed by pyrolysis, passing the bulbs through furnaces at close to the melting temperature of glass, with axial air flow to remove the carbon and hydrogen constituents.

Since the alkaline-earth oxides used for electron emission are extremely reactive to moisture; these materials are applied to the electrode coils in the form of alkaline-earth carbonates, and the carbonates are reduced to oxides by heating as part of the evacuation process. After the coating and binder burnout, the stems with lead in

wires, electrodes, and exhaust tubes are flame sealed to the ends of the bulbs, and the bulbs are then exhausted by a process of flushing inert gas in one end and out the other. Thus, several liter-atmospheres of air can be removed and replaced in a fraction of a minute, and the entire process of basing, aging, and testing can be achieved at production speeds of 3,000–6,000 lamps per hour.

Mercury is introduced into the lamp into linear *FL* in the form of a droplet. Sophisticated techniques for performing this function have been developed, to minimize the amount of mercury used and reduce the environmental impact; today, lamps operate with 2.5 mg of mercury, compared to 50 mg in the early *FL*.

During the fabrication process, care must be taken to ensure that no impurities are introduced into the lamps. Water vapor dissociates in the discharge, liberating hydrogen, which attacks the phosphor and liberates oxygen. Oxygen in turn reacts with the cathode coating and may “poison” the lamp. Baking of the lamp must also be controlled, to ensure that residual hydrocarbons from the phosphor do not result in the deposition of carbon on the phosphor or the liberation of hydrogen to damage the phosphor.

CFL are manufactured using very narrow (T4 or T5) glass tubes, which are bent and/or fused and joined to form a compact unit, intended to replace an incandescent bulb. This process is automated, with production rates comparable to linear *FL*.

Commercially Available Products

FL are used for commercial, industrial, institutional, and retail applications. Today, the majority of these lamps use the more expensive rare-earth phosphors, to provide a *CRI* of 70–90, compared to *CRI* ~ 60 for halo-phosphates. The most common *FL* in use today is the T8 lamp, which has replaced the less efficient T12 lamp, although there has been considerable penetration of linear T5 into the market, due to the improved efficiency for smaller diameters. *FL* are available in five standard colors, with *CCT* = 3,000, 3,500, 4,100, 5,000, and 6,500 K.

Many T8 lamps in both Europe and the USA are filled with argon and typically ~25 % krypton, to compensate for the increased ion–electron losses compared to T12 lamps; for a given rated power, the T8 lamp can run with the same rated power as the equivalent T12 lamp, and the same ballast may be used.

T5 and T8 lamps may be bent into a U shape, with a two-pin base at the end of each tube connecting the electrode to the ballast, thus reducing the overall length of the lamps.

Compact fluorescent lamps (*CFL*) have been designed such that the volume occupied by the discharge is similar to that of an incandescent lamp. They have found wide application for residential lighting, since they consume one quarter of the electrical power required for the equivalent incandescent lamp. There are three basic types of *CFL*: an integral (or self-ballasted) lamp contains the lamp and all the electronics required to run it in a single unit, with a base that can be fitted to an incandescent socket; the second type has the electronics separated from the tube, which is contained in a purpose-made fixture such as a table lamp – the advantage is that when the lamp fails, only the tube needs to be replaced, whereas in the integral version, the entire lamp must be replaced; the third type of *CFL* also has a separate

tube, but the electronics may be purchased separately, again allowing for substitution of the tube alone after failure.

Tubes may be constructed in one plane, circular, and coiled or with two parallel tubes. These tubes may be placed in a reflector or covered by a diffuser, to give them a similar appearance to incandescent bulbs. The factors influencing the life of a *CFL* lamp are similar to those for a linear *FL*. The typical life of a *CFL* is 6,000–15,000 h, *CRI* ~ 80, efficacy 55–65 lm/W, and *CCT* 2,700–5,000 K. Many *CFL* contain an amalgam in place of liquid mercury, which greatly extends the temperature range over which they can be operated efficiently. *CFL* containing mercury are more efficient when they are operated base up; when *CFL* are operated base down, the mercury collects near the electrodes, where the temperature is higher, resulting in a higher vapor pressure and loss of efficacy.

Low-Pressure Sodium Lamps

Introduction

The *LPS* lamp was introduced in the 1930s and is still the most efficient light source available. The physics and operation of *LPS* lamps is discussed in Waymouth (1971) and Denneman (1981), and *LPS* lamp technology is reviewed in Kirby (1997).

LPS lamps contain a minority of sodium vapor in a buffer gas which is typically neon. The optimum vapor pressure of sodium in these lamps is 3 mTorr (0.4 pa), which is attained at a temperature of 260 °C. In order to maintain the lamp at the required temperature, additional means of heating the lamp are necessary. In early *LPS* lamps, this was achieved by including a neon buffer gas at a pressure of 20 Torr (3 kPa), to provide gas heating by elastic scattering from electrons, and by enclosing the arc tube in a vacuum-sealed outer jacket to reduce convection losses (Waymouth 1971). In current *LPS* lamps, a coating on the inside of the outer jacket is applied to reflect infrared radiation back to the arc tube, which reduces the heat losses, enabling the lamps to operate with neon pressures around 8 Torr (1 kPa).

Radiation Mechanism

LPS lamps emit radiation principally in two resonance lines, 589.0 and 589.6 nm (*D* lines), which are near the peak of the eye sensitivity curve; the absence of a phosphor means that an *LPS* lamp with the same energy efficiency in converting electrical power to resonance radiation as a fluorescent lamp would have more than three times the efficacy, i.e., about 300 lpw (Waymouth 1971). In the 1930s, *LPS* lamps achieved 50 lpw, but significant improvements since that time have led to efficacies as high as 200 lpw (Kirby 1997). However, the radiation produced is principally yellow, which gives very poor color rendition, and the use of these lamps is mainly restricted to outdoor applications, such as highways and parking areas, where brightness is more important than color definition or in the vicinity of astronomy observatories, where the narrow spectrum can be filtered out during observations.

A further problem affecting the efficacy of *LPS* lamps is due to the fact that radiation trapping of the *D* lines is more effective than trapping of the 254 nm line in mercury discharges, and most of the radiation from the lamp is emitted near the surface. In a U-shaped lamp, some of the radiation emitted in one leg may be trapped in the other (Kirby 1997). Efficacy can be increased by replacing the circular cross section with one with an increased ratio of surface area to volume (such as a cross or crescent shape (Waymouth 1971)).

Energy Balance

Discharge Plasma

LPS lamps were first investigated experimentally in 1933 (Druvesteyn 1933) using Langmuir probes to measure the electron temperature and absorption spectroscopy to measure the relative densities of excited sodium atoms and sodium ions. Druvesteyn was the first to observe sodium depletion at the center of the discharge, finding that for his conditions, 83 % of sodium atoms were ionized in this region. This cathoporesis is due to the fact that *LPS* lamps operate at much higher current and electron densities than fluorescent lamps.

The main difficulty in attaining the theoretical efficacy limit of the *LPS* is to achieve both the optimum current density and wall temperature simultaneously (Waymouth 1971). The discharge power may be increased for constant current by raising the discharge voltage or lengthening the lamp. The latter solution has the further advantage, in common with the fluorescent lamp, of reducing the fraction of input power which is lost due to the electrodes. The discharge in *LPS* lamps is lengthened by bending the tube into a U shape (Waymouth 1971; Kirby 1997); in a 90 W *LPS* lamp, 30 % of input power is outputted as visible radiation and 5 % as infrared radiation, 22 % of losses are due to electrodes, and the remaining 43 % are due to other losses, such as gas heating (de Groot et al. 1984).

During the starting phase before the lamp has heated sufficiently to vaporize the sodium, the discharge is essentially a rare gas plasma. This phase lasts typically 10–15 min. A small amount (0.5–1 %) of argon is added to the neon, forming a so-called Penning (1926) mixture, to reduce the starting voltage. Penning ionization occurs when ground-state argon atoms collide with neon metastable atoms to produce argon ions. During the long starting phase, however, argon ions diffuse to the wall and may become embedded in the glass arc tube, leading to the removal of argon from the gas mixture and a rise in starting voltage. The glass used in *LPS* lamps must therefore be both nonabsorbing for argon and resistive to chemical interaction with sodium.

Lamp-Electric-Circuit System

Conventional *LPS* lamps operate on a 50/60 Hz *ac* ballast. The effect of sodium depletion can be reduced by use of a square wave ballast, to reduce the peak value of the current, but this has limitations due to the control gear requirements. Increases in efficacy of 10–20 % have been reported in discharges at operating frequencies of 100–400 kHz (de Groot et al. 1984). This is well above the frequency associated

with sodium ion diffusion (1 kHz) and of order of the decay of sodium atoms (100 kHz) in a typical *LPS* discharge. In contrast to fluorescent lamps, efficacy *decreases* as the operating frequency is increased for 1–100 kHz compared to 50/60 Hz operation, while a monotonic increase in efficacy is found as the operating frequency is increased from 100 to 400 kHz (de Groot et al. 1984).

References

- Curry JJ, Lister GG, Lawler JE (2002) Experimental and numerical study of a low-pressure Hg-Ar discharge at high current densities. *J Phys D Appl Phys* 35:2945–2953
- de Groot JJ, Jack AG, Coenen H (1984) *J Illum Eng Soc* 14:188–209
- Denneman JW (1981) *IEE Proc* 128A:397–414
- Druvesteyn MJ (1933) *Physica* 1:14–27
- Guest RA, Mascarenhas EJP (1997) In: Coaton JR, Marsden AM (eds) *Lamps and lighting*. Arnold, London, pp 292–335
- Ingold JH (1991) Ambipolar diffusion theory of the rectangular positive column with quadratic ionization. *J Appl Phys* 69:6910–6917
- Jack AG, Vrenken LE (1980) *IEE Proc* 127A:149–157
- Kirby MW (1997) Low pressure sodium lamps. In: Coaton JR, Marsden AM (eds) *Lamps and lighting*. Arnold, London, pp 227–234
- Lister GG, Waymouth JF (2002) *Light Sources*. In: *Encyclopedia of physical and technology*, 3rd ed, vol 8. Academic, pp 577–594
- Lister GG, Lawler JE, Lapatovich WP, Godyak VA (2004) The physics of discharge lamps. *Rev Mod Phys* 76:541–598
- Maya J, Lagushenko R (1990) *Adv At Mol Opt Phys* 26:321–373
- Molisch AF, Oehry BP (1998) *Radiation trapping in atomic vapours*. Clarendon, Oxford
- National lighting Product Information Program (2006) *Lighting answers*, vol 9, issue 1
- Penning FM (1926) *Z Phys* 46:335–348
- Waymouth JF (1971) *Electric discharge lamps*. The MIT Press, Cambridge, MA

Mercury-Vapor Lamps

Heinz Schöpp and Steffen Franke

Contents

Introduction	1080
Historical Overview	1080
Basic Working Principle	1080
Types of Hg Lamps and Their Applications	1080
Technological Aspects of Mercury-Vapor Lamps	1081
Electrode Feed-Through	1081
Electrodes	1081
Fill	1083
Quartz Technology	1084
Outer Bulb	1084
Physical Principles of Mercury-Vapor Lamps	1085
Ignition and Breakdown	1085
Warm-Up	1085
Stationary Operation	1085
Radiation Properties	1087
Optically Thin Plasma Radiation	1089
Optically Thick Plasma Radiation	1089
Energy Balance	1091
UHP Lamps	1092
Further Reading	1093
Directions for Future Research	1093
References	1093

Abstract

High-pressure mercury-vapor lamps are nowadays available for more than 80 years and will keep importance for the future at least for disinfection applications or lacquer curing. Although the lighting market changed dramatically, Hg

H. Schöpp • S. Franke (✉)
INP Greifswald, Greifswald, Germany
e-mail: schoepp@inp-greifswald.de; steffen.franke@inp-greifswald.de

lamps are still present; moreover they formed the basis for most of the advanced technologies. Mercury is not only an ingredient in metal-halide lamps but also in many sodium-vapor lamps or in xenon lamps, because of its outstanding material properties.

Introduction

Historical Overview

First significant types of high-pressure mercury-vapor lamps were made in the beginning of the 1930s in the twentieth century with the upcoming of borosilicate glass technology (Elenbaas 1951). This kind of glass allowed an operation temperature of around 450 °C, tungsten electrodes could be vacuum-tight connected, and it was resistant for mercury. So the important process of mercury evaporation could be initiated. Some years later the quartz technology improved this lamp type and up to now quartz technology is the state of the art for this kind of lamps (Elenbaas 1972). In last decades improvements concerning the electrodes, filling, and geometry of the arc tube were made. So a stable operation of many thousand hours is reached. The continuity in developments and efforts are reported during LS symposia (LS = International Symposium on the Science and Technology of Lighting) and international journals.

Basic Working Principle

The basic working principle can be explained as follows. A quartz vessel with two tungsten electrodes is filled with some milligram (mg) of mercury and some tens of millibar (mbar) of a starting gas like argon. For operation of the lamp a power supply is required including a component like an inductor which limits the current. The arc can be ignited, e.g., by an auxiliary voltage peak. In the beginning the discharge runs in argon. Ohmic heating of the arc causes a temperature increase of the tube wall which results in an evaporation of mercury. With increasing mercury vapor pressure, the arc is dominated by mercury, where the partial pressure is controlled by the amount of mercury, the power input into the lamp, and the tube wall temperature. In conjunction with the arc temperature, the plasma radiation is determined. This kind of lamp type was the basis for later developments like metal-halide lamps (MH lamps) with additives as well as high- and ultra-high-pressure lamps (UHP).

Types of Hg Lamps and Their Applications

A broad variety of mercury-vapor lamps is nowadays available. The input power varies between some tens of watt up to several kW. Hg lamps are still widespread in outdoor illumination. The lamp luminous efficacy varies between 30 and 55 lm/W

and a color rendering between 30 and 65. The luminous flux varies with the input power from 2,000 lm (50 W) up to 57,000 lm for a 1,000 W lamp. The lifetime can reach up to 4 years by operating 11 h every day. As metal-halide lamps provide much higher luminous efficacies and color rendering properties, mercury-vapor lamps are phased out in general lighting applications. Legal regulations accelerate this process like in the European Union, where Hg lamps are banned up to the year 2020 in general lighting (European Parliament 2005, 2009a, b). However, there will be still applications in lithography and disinfection which make use of the UV radiation of mercury-vapor lamps. Of particular interest are mercury spectral lines at 185 and 254 nm. Making use of 185 nm spectral line requires special quartz types and moderate wall temperatures because quartz absorbs radiation below 200 nm at higher temperatures (Franke et al. 2006). Compared to low-pressure lamps, very high power densities can be achieved with high-pressure mercury lamps. In Table 1 a rough list of the different types of mercury-based lamps is given (see, e.g., product catalogs of several lamp manufactures, like OSRAM, Philips, General Electric, Ushio, etc.).

Technological Aspects of Mercury-Vapor Lamps

Electrode Feed-Through

Into the cylindrical quartz vessel, two electrodes of tungsten are vacuum-tight sealed with a distance between them of some mm (projection lamps) up to some 10 mm (general lighting). Such vacuum-tight connection is realized by Mo foils of some millimeters width and some microns thickness. Special technologies improved the material properties (Ginesin et al. 2001). The Mo foil must carry the current of some ampere. It has a rounded form to decrease mechanical stress in the quartz bulb during warm-up and cooling down and is welded with the tungsten electrode.

Electrodes

Electrodes are optimized in diameter and length to the nominal input power. If electrodes are too thick, cooling by the tungsten rod is too strong and sputtering of electrode material occurs. On the other hand, if electrodes are too thin, the electrode tip melts and material is evaporated.

The electron emission of the cathode is mainly determined by the work function of the material and the temperature. The Richardson equation describes the current density in case of thermionic emission of electrons:

$$j = A T^2 e^{-\frac{W}{k_B T}} \quad (1)$$

Here j denotes the current density, T the temperature of the electrode, k_B the Boltzmann constant, W the work function of the material, and A the theoretical

Table 1 Different types of mercury-based lamps

Type	Lamp input power (W)	Luminous flux (lm)	Luminous efficacy (lm/W)	General color rendering index	Operating pressure (bar)	Main application
HQL, HWL, HPL	50–1,000	2,000–57,000	40–57	50–43	3–10	General lighting
HBO	50–16,000	2,000–640,000	Not specified	Not specified	20–50	Short arc, UV lithography
UV	Up to 60,000	Not specified	Not specified	Not specified	ca. 2	Printing, disinfection
UHP	100–300	6,000–	<70	100	ca. 200	Projection

Richardson constant of $1.20173E6 \text{ A}/(\text{m}^2 \text{ K}^2)$. The Richardson constant differs for different materials significantly. The Schottky effect describes the dependence of the work function on the field strength at high electric fields. Any decrease in the work function reduces the required power to extract the current from the electrodes. This can be achieved by emitter materials. For a long time thorium was used, which is an emitter of beta radiation and is embedded in the tungsten structure. It acts as a tool for pre-ionization of the plasma and reduces the work function of the electrodes. At the same time, mixed oxides were developed, e.g., BaCaWO_5 . They are deposited in the interspace of a coil which is arranged around the tungsten rod and reduces the work function of around 2 V. The coil itself affects the heat balance of the electrode. Modern mercury-vapor lamps own simple rod tungsten electrodes avoiding the radioactive thorium and do not need any emitter materials due to optimized diameters. As electrodes are optimized for nominal power input, there is increased electrode erosion during warm-up and in dimmed operation, reducing the lifetime.

The theoretical estimation of the work function requires a physical simulation of the plasma boundary layer (Benilov 2008; Lichtenberg et al. 2005).

Although cathode effects have been the focus of research, there are also a number of papers about anode effects (Almeida et al. 2009; Heberlein et al. 2010; Redwitz et al. 2006). In contrast to the cathode, which is cooled by the electron current, the anode is heated by the electron current. The dimension of anode has to be chosen appropriately to guarantee an optimum heat management between anode and cathode phase of the electrodes.

In all cases the mercury-based light sources need a special power supply. In simple cases for general lighting, an inductive ballast is used, whereas modern lamps are driven with an electronic device. The ballast type is also responsible for the starting process of the lamps. While the ignition of cold lamps requires 3–6 kV, hot re-strike (re-ignition) needs about 10 kV which requires special sockets and cables to ensure safe operation.

For some minor cases of lamp types for better ignition, an additional electrode near the main ones (distance is only around 1 mm) is inserted.

Fill

The vessel is filled with some mg of mercury (depending on the power class) and of a starting gas (mostly argon, around 50 mbar) for the ignition. Although Xe has a lower ionization energy appropriate for easy ignition, Ar is preferred for cost reasons. It is possible to use a mixture of Ar and the radioactive isotope Kr85 to improve ignition behavior. Ne has an increased heat conductivity but a high ionization energy and tends to permeate through the hot quartz tube. From temperature-dependent mercury vapor pressure, the partial pressure during operation can be estimated if mercury is not completely evaporated and the cold spot temperature is known. The cold spot temperature is the temperature of the coldest region of the quartz tube wall. For cold spot temperatures of up to 1,300 °C, the vapor pressure is of around 70 bar (Hansen et al. 1998) in case of saturated vapor pressure (incomplete

evaporation of mercury). As in lamps for general lighting vapor pressures are below 10 bar and wall temperatures are lower than this 1,300 °C and in most cases around 1,000 °C, in these lamps consequently the mercury fill is completely evaporated. The mercury vapor pressure is unsaturated. In this case the partial pressure does not follow from the temperature-dependent vapor pressure.

Quartz Technology

There are a number of quartz and glass materials which can be applied for high-pressure lamps and special applications. One can distinguish between pure quartz materials with varying oxygen and water content shifting the VUV absorption edge. Allowing emission between 180 and 200 nm requires quartz of best available purity. On the contrary it is possible to dope quartz material to shift the absorption edge toward higher wavelengths. Hence, materials are available with defined absorption of UV below 400 nm. The temperature dependence of VUV transmission of fused silica was investigated, for instance, by Franke et al. (2006). It has to be noted that reabsorption of plasma radiation in the tube wall over the whole spectral range affects the temperature of the cold spot and herewith the evaporation of the fill. This is of particular importance for metal-halide lamps (see chapter ► [“Metal-Halide Lamps”](#)).

Outer Bulb

However, doped quartz materials usually do not resist the required wall temperatures. Hence, mercury-vapor lamps for general lighting have an outer bulb with borosilicate glass to absorb UV radiation, which would be hazardous for humans. The evacuated outer jacket contains a getter material ensuring vacuum and improves thermal stable operation of the burner. Figure 1 shows a picture of a lamp without a socket. In the past some types of lamps were built with a phosphor on the inner side of the outer bulb, converting UV radiation into visible light (Kohmoto 1999). However, resistance of the phosphor against higher temperatures around 200 °C limited life of these lamp types.

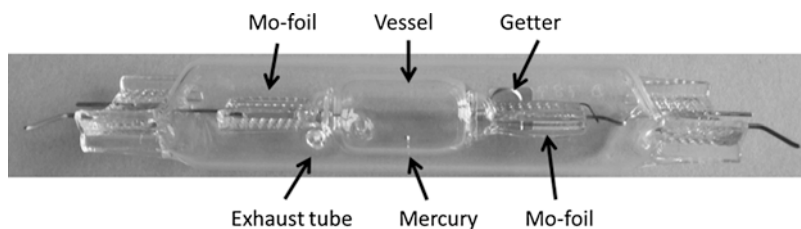


Fig. 1 Picture of a mercury-filled lamp with some details

Physical Principles of Mercury-Vapor Lamps

From a physical point of view, three phases of arc operation can be distinguished.

Ignition and Breakdown

First there is the ignition phase. A number of papers deal with topics related to ignition like Beckers et al. (2008), Czichy et al. (2008), Sobota et al. (2008, 2009), and Wendt et al. (2009). With respect to the timescale of the development of a stationary discharge, which may take up to several minutes, ignition occurs in the first tens of nanoseconds including field emission of electrodes, cathodic as well as anodic streamer propagation, and finally a breakdown with a diffuse attachment of the arc at the electrode tip.

Warm-Up

Second, there is the warm-up phase, which is described in papers like Araoud (et al. 2010), Stambouli et al. (1999), and Zalach et al. (2011). During this phase, thermo-emission of the electrodes is established and a transition from diffuse to spot mode of the arc attachments takes place. It takes only seconds that the argon-dominated discharge is taken over by mercury. This is obvious from the disappearance of argon spectral lines and the occurrence of mercury emission (see Fig. 2). After the breakdown mostly lines are emitted from the starting gas. The electron transport and the radiation lead to the warm-up of the gas, the electrodes, and the tube (see Fig. 3). With the increasing temperature of this system, the evaporation of mercury and additives also increases. The evaporation increases the particle densities of different species and the physical properties of these species determine the radiated output.

Furthermore, a thermalization of the whole discharge vessel is reached where all physical processes reach a stationary state like thermal energy balance of electrodes, material convection inside the burner as well as Ohmic heating of the arc which is balanced by radiation emission, and cooling at the tube walls.

Stationary Operation

This leads to the third state of arc operation, which is called the stationary operation phase. Depending on the driving current, the stationary phase can be considered as a quasi-DC discharge, if a square wave is applied to the lamp, or it has to be described in terms of a time-dependent plasma, if a sine-like current is supplied. In any case mercury almost completely dominates the plasma properties. This includes the electric conductivity, the thermal conductivity, as well as the radiation properties of the arc. With respect to electric conductivity, mercury is a unique material as it

Fig. 2 Warm-up of Ar-Hg lamp (Zalach et al. 2012). Voltage and voltage gradient with distinct points (characteristic structures of $U(t)$ better visible in a gradient view): (A) Hg near tube endings vaporized, (B) Hg layer on inner wall vaporized, (C) simultaneous evaporation of three cold areas, and (D) complete evaporation. *Bottom:* line intensity evolution shows the change of the dominant species

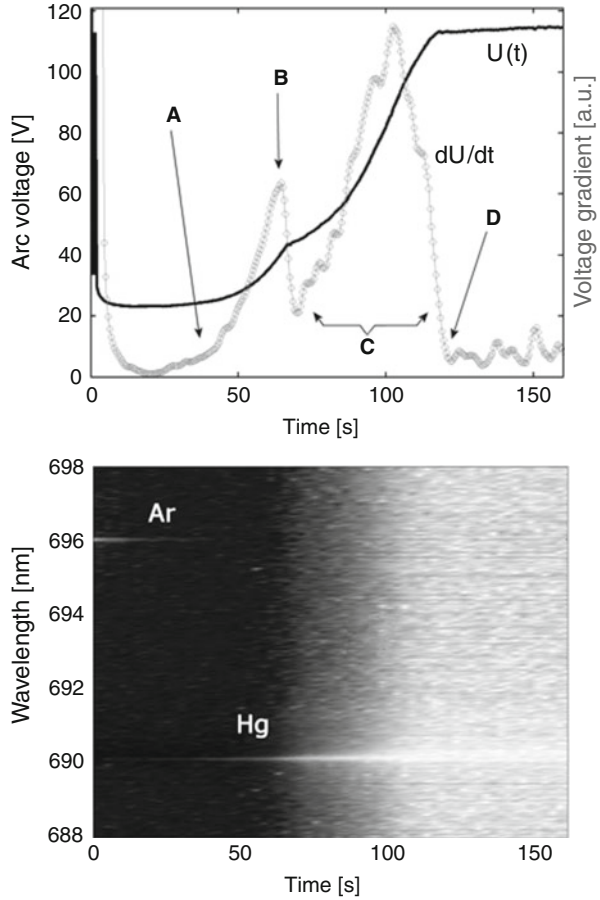
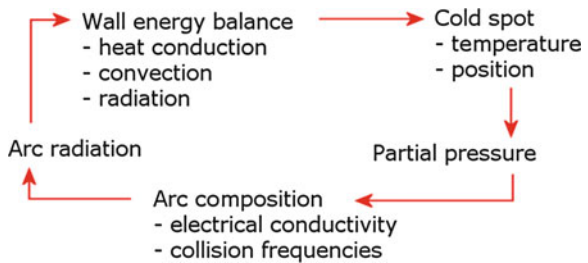


Fig. 3 Interaction scheme of cold spot and arc



combines a high electron momentum transfer cross section with low vapor pressure. Both properties ensure a significant evaporation of mercury even at quite low temperature and finally a low electric conductivity of the arc. The low electric conductivity is advantageous as it allows a high power input at low electric currents, which avoids excessive stress to the electrodes. Combined with the low thermal

conductivity compared to argon, an acceptable thermal load of the tube wall can be achieved and plasma temperatures allow an effective emission of radiation. The atomic structure of mercury with its beneficial distribution of energy levels favors radiation in the visible spectral range with prominent atomic spectral lines in the blue, green, and red. For years these spectral lines have been the standard primary colors for RGB displays.

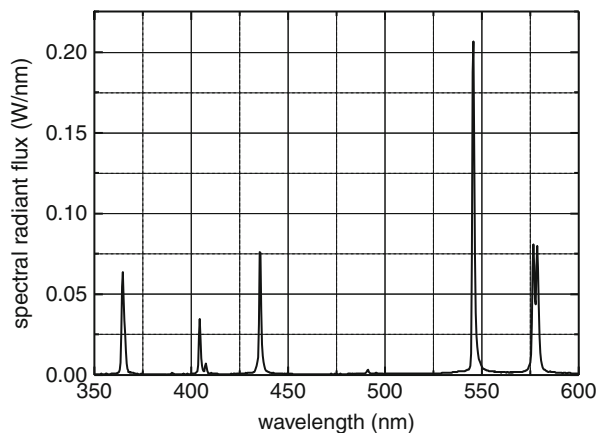
Before going into the details of plasma radiation, two states of the discharge should be named, which are the afterglow and re-ignition. It is a special feature of high-pressure lamps that lamps have to cool down before they can be re-ignited. This is due to the fact that after switching off the lamp, an afterglow can be observed, where radiation decreases and recombination of electrons occurs at a timescale of submilliseconds. The breakdown voltage of the hot but electrically neutral mercury vapor is very high and needs high re-ignition voltages of some 10 kV. However, even in lamps with sine current operation, re-ignition peaks can be observed after current zero, due to a temporal cooling down of the plasma.

Radiation Properties

As described above in stationary state, the emission spectrum of a mercury high-pressure lamp is dominated by atomic spectral lines in the visible spectral range as can be seen in Fig. 4. There are also important UV lines around 185, 254, and 365 nm. However, for UV applications usually medium-pressure and low-pressure mercury vapor lamps are used, because the ratio of UV radiation to the total radiative output is advantageous. One reason is the increasing reabsorption of atomic spectral lines with increasing mercury partial pressure. Continuum emission is of minor importance in high-pressure mercury vapor lamps but is reported in [Burm \(2005\)](#) and [Käning et al. \(2007\)](#).

Spectral emission can be described following [Griem \(1997\)](#) and [Lochte-Holtgreven \(1995\)](#) by the local spectral emission coefficient:

Fig. 4 Typical spectrum of a high-pressure mercury vapor lamp at 250 W nominal power input with quartz tube burner



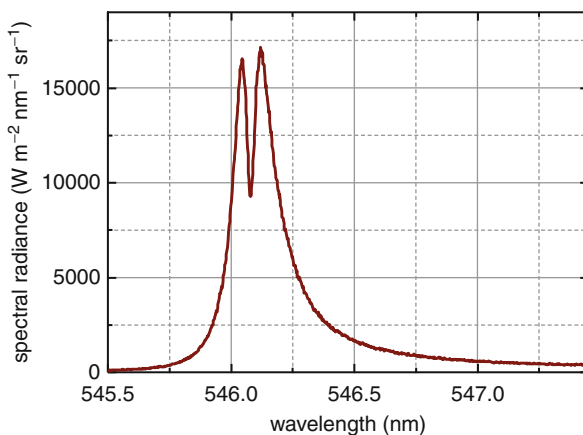
$$\varepsilon_\lambda = \frac{1}{4\pi} \frac{hc}{\lambda_0} \frac{g_u}{Z(T)} A_{ul} \frac{p_0}{k_B T} \exp\left(-\frac{E_u}{k_B T}\right) P_\lambda, \quad (2)$$

which is given in units of $\text{W m}^{-3} \text{nm}^{-1} \text{sr}^{-1}$. The symbols have the following meanings: h Planck constant, c vacuum speed of light, λ_0 wavelength of the spectral line, n_u population density of the excited upper level, g_u statistical weight of the upper level, $Z(T)$ partition function of mercury, A_{ul} transition probability, p_0 mercury partial pressure, k_B Boltzmann constant, T plasma temperature, E_u energy of the excited level, and P_λ spectral line profile (area under curve normalized to 1). This equation assumes local thermodynamic equilibrium with a Boltzmann distribution of excited levels and the validity of the ideal gas law. As the main broadening mechanism, the pressure broadening (also called van der Waals broadening) is assumed (Griem 1974) which leads to a Lorentz-like spectral line profile:

$$P_\lambda(T, p_0) = \frac{1}{\pi} \frac{\lambda_L(T, p_0)/2}{(\lambda - \lambda_0)^2 - \lambda_L^2(T, p_0)/4}, \quad (3)$$

with λ_L – line profile width (full width at half maximum = FWHM). The profile width depends on the plasma temperature and partial pressure. It has to be mentioned that the Lorentz line profile is obtained for impact approximation of the van der Waals broadening, if the time of collision is small compared to the time between two collisions. For quasi-static van der Waals broadening, the line profile becomes asymmetric (see Fig. 5). The according line profile is described in Stormberg and Schäfer (1983) as a convolution of a Lorentz profile and a profile for quasi-static broadening. A more exact treatment of van der Waals broadening is given in Al-Saqabi and Peach (1987) which can also be applied to mercury spectral line profiles (Wendt and Franke 2008). This approach also includes resonance and quadratic Stark broadening which both form a Lorentzian line shape (Griem 1997; Lochte-Holtgreven 1995). Doppler broadening usually has no significant effect on spectral lines in high-pressure discharge lamps. In principle plasma parameters (like temperature or particle densities)

Fig. 5 Example of a self-reversed spectral line at 546 nm



can be determined by comparison of experimental spectra with radiative transfer calculations (Wendt and Franke 2008). However, there exist methods to evaluate isolated spectral lines allowing the estimation of different plasma properties.

Optically Thin Plasma Radiation

Assuming an optically thin plasma, the spectroscopically observable spectral radiance can be obtained from an integration of emission coefficients along a line of sight. The reverse process is called Abel inversion (Andanson et al. 1978; Dribinski et al. 2002) and delivers again the local emission coefficient at different radial positions in case of a cylindrically symmetric discharge. If the mercury partial pressure is known, the plasma temperature can be determined from this implicit Eq. 2, e.g., via a lookup table.

The mercury partial pressure p_0 can be roughly estimated by Kenty formula (Kenty 1942):

$$p_0 = 0.75 \times W^{1/4} \times M^{0.9} / (2R)^{2.1}, \quad (4)$$

with p_0 in atm; W , the input power in Watt per unit arc length in cm (W/cm); M , the mercury mass in mg per unit arc length in cm (mg/cm); and R , the arc radius in cm. Alternatively it can be estimated iteratively from the effective plasma temperature as exemplified in Zalach and Franke (2013) for a Xe arc.

The temperature-dependent partition function $Z(T)$ be calculated from

$$Z(T) = \sum_j g_j \exp\left(-\frac{E_j}{k_B T}\right) + g_j \exp\left(-\frac{E_{ion}}{k_B T}\right) \times \left(\frac{E_j - E_{ion}}{k_B T} - 1\right) \quad (5)$$

where the first term is the partition function according to Drawin and Felenbok (1965). However, this sum diverges for $j \rightarrow \infty$. Therefore it needs to be limited either by a defined cutoff energy or by a smooth convergence given with the second term according to Planck-Larkin (Ebeling et al. 1976); see also Wendt (2011). In the case of mercury, the partition function is around $1+/-0.03$ for plasma temperatures between 1,000 and 10,000 K.

Typically in high-pressure mercury vapor lamps, only the Hg lines at 577/579 nm are suitable for an optically thin evaluation, whereas the other prominent lines underlie self-absorption.

Optically Thick Plasma Radiation

Optical thickness of a plasma leads to a reabsorption of radiation and increases with increasing particle densities, i.e., with increasing optical path length along a line of sight (Zwicker 1995). Furthermore, it depends on the atomic properties of the transition. Reabsorption leads to a redistribution of radiation energy. Photons emitted

in the line center are absorbed at outer radial positions, whereas photons emitted in the line wings escape from the discharge. In spatially inhomogeneous plasmas, self-reversed line contours are formed as demonstrated in Fig. 5 for the mercury line at 546 nm. Self-reversed spectral lines are characterized by a self-reversal minimum at maximum optical depth (>5) and self-reversal maxima at optical depths around 2–3. Toward the line wings the optical depth is further decreasing.

The process of reabsorption is described by the radiative transfer equation

$$L_{\lambda}(y) = \int_{-x_0}^{+x_0} \varepsilon_{\lambda}(x) \exp \left\{ - \int_{x'}^{x_0} \kappa(\lambda, x') dx' \right\} dx, \quad (6)$$

with L_{λ} spectral radiance in $\text{W m}^{-2} \text{nm}^{-1} \text{sr}^{-1}$ at the radial position y and κ the absorption coefficient in $1/\text{m}$. The integration is performed along a line of sight from one border ($-x_0$) to another border ($+x_0$).

Like optically thin spectral lines, also self-reversed lines are suitable for plasma temperature determination. Two methods are established for this purpose. One is the Bartels method (Bartels 1949a, b); the other is the Karabourniotis method (Karabourniotis 1983, 1984, 2005, 2006) based on Cowan and Dieke (1948).

Bartels' analysis (Bartels 1949a, b) gives the most complete insight into the self-reversal mechanism, but the practical use is often limited to LTE conditions. The only required experimental information is the absolute spectral radiances at the reversal maxima at different side-on positions. For a constant partial pressure of the radiating species, the data analysis becomes pressure independent and delivers the temperature distribution directly. In the general case, the inclusion of computed plasma compositions and line broadening details is recommended to increase the accuracy of the method and to extend the applicability to plasmas not in LTE (Schneidenbach and Franke 2008). Bartels equations in its simplest approximation are given below:

$$\begin{aligned} T(\lambda_0) &= \frac{\Delta E/k_{\text{B}}}{\ln(2hc_0^2/\lambda_0^5) + \ln(M Y) - \ln(I_{\text{max}})} \\ M &= \sqrt{E_l/E_u} \\ Y &= 0.736 + 0.264p^2 \\ p &= \frac{6}{\pi} \arctan \left(\frac{M^2}{\sqrt{1 + 2M^2}} \right) \end{aligned} \quad (7)$$

This approximation works well for high-pressure mercury lamps. The striking feature is an explicit expression for the axis plasma temperature as a function of the self-reversal maximum radiance at a given side-on position. Furthermore, no transition probability is needed as well as features of line broadening. In the given simplest approximation, Bartels method is limited to LTE conditions.

The Karabourniotis method (Karabourniotis 1983) based on Cowan and Dieke (1948) is not limited to LTE conditions. Position-independent line profiles are assumed. The key problem is the determination of the inhomogeneity parameter which requires the additional measurement of absolute radiances at the reversal minimum and the relative distances from the considered reversal maximum. The data analysis is based on available general functions. The procedure inherits uncertainties resulting from the basic assumptions which increase with decreasing lower-level energy and can exceed 20 %.

Both methods deliver excellent results for diagnostic lines with high lower-level energies E_i . Examples are the mercury lines at 365, 404, 436, and 546 nm. The inaccuracies caused by the methods itself are markedly smaller than 5 %. For lines with low E_i , especially for resonance lines, uncertainties arise which strongly depend on the considered plasma conditions. In the simplest approximation of Bartels method, resonance lines are excluded. The reliability of the results has to be proven by additional measurements of different lines, application of alternative methods like absorption measurements, or detailed theoretical analysis.

Energy Balance

As mentioned above in optically thick plasmas, the radiative transfer affects the local energy balance as energy from the arc core is deposited at outer radial positions by reabsorption. Over the years the theoretical description of the mercury arc became more and more detailed. Relevant atomic data for calculation of the radiation and the energy balance are estimated and checked, e.g., by Elenbaas (1951), Stormberg and Schäfer (1983), and Hartel et al. (1999).

It reveals that the radiative output of high-pressure mercury vapor lamps is dominated by UV radiation. In modern lamps for general lighting, UV lines are absorbed in the outer bulb, whereas the Hg I 185 nm resonance line is absorbed in the inner quartz tube. Depending on the quartz material and technology, a temperature shift of the transmission is observed (Franke et al. 2006).

An important feature of high-pressure mercury vapor lamps is their ability to determine the electrode sheath voltages (ESV) as given in Hartel et al. (1999). The average ESV was assumed to be around 15 V (Elenbaas 1951; Stormberg and Schäfer 1983), whereas in Hartel et al. (1999) a pronounced time-dependent behavior was observed for the first time.

A total energy balance of a 400 W Hg lamp is estimated by Jack and Koedam (1974). Only 15 % of total input power is emitted in the visible spectral range leading to a luminous efficacy of around 50 lm/W. Around 18 % of the power input are dissipated by UV radiation and 15 % by infrared radiation. Electrode losses account for less than 8 %. Therefore the remaining 44 % of input power are released into the tube wall by heat conduction and radiation absorption, where the energy mainly is emitted by heat radiation.

UHP Lamps

A very special and important mercury lamp is the ultra-high-pressure (UHP) lamp. In projection systems this kind of lamp becomes important in the last decade. UHP lamps are short arc discharges where the advantage of a point source in an imaging system is used. Filled with Ar and mercury a very high pressure of more than 200 bar arises and leads to a strong continuum which dominates the total radiation. Their system efficacy of 10 lm/W is highest of the display techniques today (Derra et al. 2005). So very good color rendering can be reached after image generation of the display. Some key developments were necessary:

- The burner must have a geometry which withstand the high pressure at high temperatures.
- The wall temperature requires around 1,200 K at coldest spot inside the lamp to generate the high mercury pressure, but should not exceed 1,400 K before deformation or recrystallization of the quartz occurs.
- The electrodes need an optimum temperature distribution for small burn-back with a sufficient cooling by radiation and a hot tip for stable arc attachment which requires an optimized geometry.
- A halogen cycle has to guarantee that no electrode material is deposited on the burner wall because around 77 % of the electrical energy input leave the lamp as radiation (Derra et al. 2005) and a small blackening of the quartz wall has dramatical consequences on the thermal balance of the whole system.
- The electrode distance should be very small for the projection requirements; on the other hand the lamp voltage is the sum of electrode fall voltages (anode and cathode) and arc length; for stable operation the arc length should not change during lifetime by the deformation of the electrode geometry because a length change of around 0.1 mm results in a lamp voltage change which causes a lamp power change by 15 % and affect the power balance dramatically.

The lamp power reaches up to 300 W and a further increase is becoming more and more problematic because the burner design is limited by the material properties of quartz. A special power supply generates the stabilized high frequent current waveform with a dip on the rectangular current reversal to influence the electrode behavior. Another point is the ignition of the UHP lamp. To generate primary electrons in modern lamps, a so-called UV enhancer is used which consists of a small capillary discharge in the electrode seal of the burner. This discharge is capacity driven in an Ar-Hg mixture and produces UV radiation near the main discharge part. So lamp can be ignited at voltages much below 5 kV.

An integrated reflector increases the radiation of the continuum by 60 %, whereas the line radiation is only increased by around 35 %. That means that in the vicinity of strong atomic transitions, the radiation is reabsorbed by the plasma and the reabsorbed radiation serves as an additional heating mechanism which has to be taken into account for the whole power balance. An investigation of the molecular continuum of UHP lamps can be found in Wharmby (2008).

The UHP concept is very successful because the very high pressure in combination with the halogen cycle realizes high luminance, long lamp life, and stable lamp parameters.

Further Reading

Further descriptions of high-pressure mercury vapor lamps and other high-intensity discharge lamps can be found in Flesch (2006), Kirby (1997), Lister et al. (2004), and Waymouth (1971). An extensive collection of images showing some of the historical developments of HID lamps is held on the Lamptech website (see www.lamptech.co.uk).

Directions for Future Research

High-pressure mercury vapor lamps will be more and more replaced by metal-halide lamps due to the efficacy and color rendering properties for general lighting applications. However, they will stay in use for applications like ozone generation, disinfection, and lacquer curing. The mercury vapor lamp was the “white mouse” of high-pressure discharge lamp physics and may keep some interest as a test system for the study of basic physical phenomena.

Acknowledgments The authors kindly acknowledge valuable contributions to this chapter by Stuart Mucklejohn, Hartmut Schneidenbach and Manfred Kettlitz.

References

- Almeida NA, Benilov MS, Hechtfisher U, Naidis GV (2009) Investigating near-anode plasma layers of very high-pressure arc discharges. *J Phys D Appl Phys* 42(4)
- Al-Saqabi BNI, Peach G (1987) Unified theories of the pressure broadening and shift of spectral lines. 2. Van der waals interactions. *J Phys B At Mol Opt Phys* 20(6):1175–1191
- Andanson P, Cheminat B, Halbique AM (1978) Numerical-solution of Abel integral-equation – application to plasma spectroscopy. *J Phys D Appl Phys* 11(3):209–215
- Araoud Z, Ahmed RB, Hamida MBB, Franke S, Stambouli M, Charrada K, Zissis G (2010) A two-dimensional modeling of the warm-up phase of a high-pressure mercury discharge lamp. *Phys Plasmas* 17(6):063505–063512
- Bartels H (1949a) Über Linienemission aus inhomogener Schicht. I. Teil. *Z Phys* 125:597–614
- Bartels H (1949b) Über Linienemission aus inhomogener Schicht. II. Teil. *Z Phys* 126:108–140
- Beckers J, Manders F, Aben PCH, Stoffels WW, Haverlag M (2008) Pulse, dc and ac breakdown in high pressure gas discharge lamps. *J Phys D Appl Phys* 41(14):144028
- Benilov MS (2008) Understanding and modelling plasma–electrode interaction in high-pressure arc discharges: a review. *J Phys D Appl Phys* 41(14):144001
- Burm KTAL (2005) Continuum radiation spectroscopy in a high-pressure argon-mercury lamp. *J Quant Spectrosc Radiat Transf* 95(1):93–100
- Cowan RD, Dieke GH (1948) Self-absorption of spectrum lines. *Rev Mod Phys* 20(2):418–455

- Czichy M, Hartmann T, Mentel J, Awakowicz P (2008) Ignition of mercury-free high intensity discharge lamps. *J Phys D Appl Phys* 41(14):144027
- Derra G, Moench H, Fischer E, Giese H, Hechtfisher U, Hensler G, Koerber A, Niemann U, Noertemann FC, Pekariski P et al (2005) UHP lamp systems for projection applications. *J Phys D Appl Phys* 38(17):2995–3010
- Drawin HW, Felenbok P (1965) Data for plasmas in local thermodynamic equilibrium. Gauthier-Villars, Paris
- Dribinski V, Ossaditchi A, Mandelshtam V, Reisler H (2002) Reconstruction of Abel-transformable images: the Gaussian basis-set expansion Abel transform method. *Rev Sci Instrum* 73(7):9
- Ebeling W, Kraeft WD, Kremp D (1976) Theory of bound states. In: Rompe R, Steenbeck M (eds) *Ergebnisse der Plasmaphysik und der Gaselektronik*. Akademie, Berlin
- Elenbaas W (1951) The high pressure mercury vapour discharge. North-Holland, Amsterdam
- Elenbaas W (1972) *Light sources*. The Macmillan, London
- European Parliament T (2005) Directive 2005/32/EC of the European Parliament and of the Council of 6 July 2005 on energy-using products. *Off J Eur Union* L191:29–58
- European Parliament T (2009a) Commission Regulation (EC) No 244/2009 of March 2009 on ecodesign requirements for non-directional household lamps. *Off J Eur Union* L76:3–16
- European Parliament T (2009b) Commission Regulation (EC) No 245/2009 of 18 March 2009 on ecodesign requirements for fluorescent lamps and high intensity discharges. *Off J Eur Union* L076:17–44
- Flesch P (2006) *Light and light sources: high-intensity discharge lamps*. Springer, Berlin/Heidelberg
- Franke S, Lange H, Schoepp H, Witzke HD (2006) Temperature dependence of VUV transmission of synthetic fused silica. *J Phys D Appl Phys* 39(14):3042–3046
- Ginesin BA, Karpov MI, Glebovsky VG, Karelin BA (2001) High-purity solid solution as a new type of molybdenum alloy. *J Adv Mater* 33(3):3–9
- Griem HR (1974) Spectral line broadening by plasmas. Academic, New York/London
- Griem HR (1997) Principles of plasma spectroscopy. In: Haines MG, Hopcraft KI, Hutchinson IH, Surko CM, Schindler K (eds) *Cambridge monographs on plasma physics*. Cambridge University Press, Cambridge
- Hansen S, Getchius J, Brumleve TR (1998) Vapor pressure of metal bromides and iodides. *APL Engineered Materials*, Urbana
- Hartel G, Schöpp H, Hess H, Hitzschke L (1999) Radiation from an alternating current high-pressure mercury discharge: a comparison between experiments and model calculations. *J Appl Phys* 85(10):7076–7088
- Heberlein J, Mentel J, Pfender E (2010) The anode region of electric arcs: a survey. *J Phys D Appl Phys* 43(2):023001
- Jack AG, Koedam M (1974) Energy balances for some high pressure gas discharge lamps. In: *Journal of IES, Annual IES conference, Light division*. N.V. Philips' Gloeilampenfabrieken, Eindhoven, pp 323–329
- Känning M, Schalk B, Schneidenbach H (2007) Experimental determination of parameters for molecular continuum radiation of rare-earth iodides. *J Phys D Appl Phys* 40(13):3815–3822
- Karabourniotis D (1983) Plasma temperature determination from the maximum intensity of a symmetric self-reversed line. *J Phys D Appl Phys* 16:1267–1281
- Karabourniotis D (1984) Correction: plasma temperature determination from the maximum intensity of a symmetric self-reversed line. *J Phys D Appl Phys* 17(6):1325
- Karabourniotis D (2005) Validity of plasma temperature determination from line self-reversal. In: *Proceedings of the 27th ICPIG, no. 08–160*. ICPIG, Eindhoven
- Karabourniotis D (2006) Effect of the one-parameter model on the spectral intensity of a self-absorbed line. *High Temp Mater Process US* 10(3):479–490
- Kenty C (1942) Pressures and temperatures in high-pressure mercury lamps. *Phys Rev* 61:545
- Kirby MW (1997) Mercury lamps. In: Coaton JR, Marsden AM (eds) *Lamps and lighting*, 4th edn. Routledge, New York, pp 254–262

- Kohmoto K (1999) Phosphors for lamps. In: Shionoya SH, Yen W (eds) High-pressure mercury lamps. CRC Press, Boca Raton, pp 375–379, 2000 Corporate Blvd NW, Boca Raton, FL 33431
- Lichtenberg S, Dabringhausen L, Langenscheidt O, Mentel J (2005) The plasma boundary layer of HID-cathodes: modelling and numerical results. *J Phys D Appl Phys* 38(17):3112–3127
- Lister GG, Lawler JE, Lapatovich WP, Godyak VA (2004) The physics of discharge lamps. *Rev Mod Phys* 76(2):541–598
- Lochte-Holtgreven W (1995) Plasma diagnostics. American Institute of Physics, New York
- Redwitz M, Dabringhausen L, Lichtenberg S, Langenscheidt O, Heberlein J, Mentel J (2006) Arc attachment at HID anodes: measurements and interpretation. *J Phys D Appl Phys* 39(10):2160–2179
- Schneidenbach H, Franke S (2008) Basic concepts of temperature determination from self-reversed spectral lines. *J Phys D Appl Phys* 41(14):144016
- Sobota A, van Veldhuizen EM, Stoffels WW (2008) Discharge ignition near a dielectric. *IEEE Trans Plasma Sci* 36(4):912–913
- Sobota A, Lebouvier A, Kramer NJ, van Veldhuizen EM, Stoffels WW, Manders F, Haverlag M (2009) Speed of streamers in argon over a flat surface of a dielectric. *J Phys D Appl Phys* 42(1):015211
- Sambouli M, Charrada K, Costache C, Damelincourt JJ (1999) Modeling the warm-up phase of a high-pressure-lamps lighting network. *IEEE Trans Plasma Sci* 27(3):646–654
- Stormberg H-P, Schäfer R (1983) Time-dependent behavior of high-pressure mercury discharges. *J Appl Phys* 54(8):4338–4347
- Waymouth JF (1971) Electric discharge lamps. M.I.T. Press, Cambridge, MA
- Wendt M (2011) Net emission coefficients of argon iron plasmas with electron Stark widths scaled to experiments. *J Phys D Appl Phys* 44(12):125201
- Wendt M, Franke S (2008) Broadening constants of mercury lines as determined from experimental side-on spectra. *J Phys D Appl Phys* 41(14):144018
- Wendt M, Peters S, Loffhagen D, Kloss A, Kettlitz M (2009) Breakdown characteristics of high pressure xenon lamps. *J Phys D Appl Phys* 42(18)
- Wharmby DO (2008) Estimates of molecular absorption cross-sections in mercury plasmas at very high pressures using self-reversed line diagnostics. *J Phys D Appl Phys* 41(14):144017
- Zalach J, Franke S (2013) Iterative Boltzmann plot method for temperature and pressure determination in a xenon high pressure discharge lamp. *J Appl Phys* 113(4):043303–043307
- Zalach J, Araoud Z, Charrada K, Franke S, Schoepp H, Zissis G (2011) Experimental and theoretical investigations on the warm-up of a high-pressure mercury discharge lamp. *Phys Plasmas* 18:033511
- Zalach J, Franke S, Schöpp H (2012) Experimental characterization of the warm-up of mercury lamps. In: Devonshire R, Zissis G (eds) Light sources. FAST-LS Ltd, Troy, pp 157–158, Belmayne House, 99 Clarkhouse Rd., Sheffield, S10 2LN, UK
- Zwicker H (ed) (1995) Evaluation of plasma parameters in optically thick plasmas. American Institute of Physics, New York, pp 214–249

High-Pressure Sodium-Vapor Lamps

Heinz Schöpp and Steffen Franke

Contents

Introduction	1098
Historical Overview	1098
Lamp Types and Applications	1098
Basic Working Principle	1099
Technological Aspects of High-Pressure Sodium-Vapor Lamps	1099
Physical Principles of High-Pressure Sodium-Vapor Lamps	1099
Radiation Properties	1099
Electrical and Thermal Conductivities	1100
Power Balance	1101
Electrode Phenomena	1101
Pulsed High-Pressure Sodium-Vapor Lamps	1101
HPS Lamps with Improved Color Rendering	1102
Note About Historical Street Lighting	1102
Directions for Future Research	1103
References	1103

Abstract

The striking feature of sodium-vapor lamps is their radiation dominated by the emission of two resonance lines at 589/590 nm. This emission is quite close to the maximum luminous efficiency function at 555 nm and hence contributes effectively to the luminous efficacy. However, due to the limited spectral range of the emission, the spectrum is yellowish and suffers from a poor color rendering.

H. Schöpp • S. Franke (✉)
INP Greifswald, Greifswald, Germany
e-mail: schoepp@inp-greifswald.de; steffen.franke@inp-greifswald.de

Introduction

Historical Overview

Due to the limited luminous efficacy and color rendering of high-pressure mercury-vapor lamps, novel lamp types were needed to increase the energy efficiency as well as color rendition. One option to increase energy efficiency was to introduce Na into the lamps, where the sodium resonance lines at 589/590 nm significantly contribute to the spectrum. However, the use of sodium required a modified arc tube design and materials resistant to metallic sodium. This development took place in the 1960s of the twentieth century.

Lamp Types and Applications

Nowadays, high-pressure sodium (HPS) lamps are widespread in outdoor illumination, where they replaced most of the high-pressure mercury-vapor lamps in Europe, due to their high-energy efficiency. A typical 70 W HPS lamp (see Fig. 1) has an energy efficacy of around 90 lm/W. Highest efficacies possible go up to 150 lm/W. However, the color rendition is still less than that of mercury-vapor lamps with a general color rendering index below 25 for the HPS lamps. HPS lamps are available for powers of 50 W up to 1,000 W. Although white LED

Fig. 1 Photography of a sodium-vapor lamp (OSRAM NAV-T 70 W)



lamps gain an increasing impact on outdoor illumination, there are parties that favor sodium-vapor lamps, because they emit almost no radiation in the blue spectral range. However, even if there is an effect of blue light of outdoor illumination on the human circadian system, no conclusions can be drawn to the optimum lighting scenario in terms of the ecological impact of outdoor illumination. For this purpose, issues like the visual perception of mesopic vision, energy efficiency, action on human nocturnal melatonin suppression, and optimal spectra for minimized hazard of outdoor illumination on ecosystems have to be investigated.

Basic Working Principle

Most of high-pressure sodium-vapor lamps also contain mercury to improve the start-up behavior and to increase the lamp voltage as well as the spectral line broadening. It is possible to use xenon instead of argon to provide a higher radiative output immediately after breakdown of the arc. However, the working principle is mainly the same as for the high-pressure mercury-vapor lamp (see chapter “Mercury-Vapor Lamp”). The arc is ignited in a noble gas and is just seconds later overtaken by mercury, when the tube walls heat up. Sodium comes later into play, due to the lower vapor pressure of this element. In stationary operation, when a saturated vapor pressure of sodium is established, the radiative output is dominated by sodium resonance lines, whereas the lamp voltage is still dominated by the mercury content.

Technological Aspects of High-Pressure Sodium-Vapor Lamps

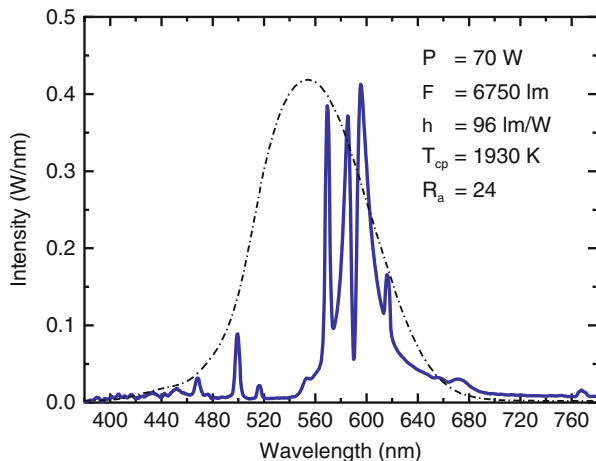
Alkali metals are known to be chemically very aggressive in both the liquid and gas phases, and quartz is not sufficiently inert to withstand these conditions. Therefore, a novel tube material was needed that should be chemically resistant, accepts high thermal loads, and is transparent or at least translucent. Two materials that came into focus were yttrium oxide and aluminum oxide. As yttrium oxide turned out to be much too expensive, sintered polycrystalline alumina became favored, providing all the required properties. Particularly wall temperatures up to 1,800 C are possible for this material. However, special techniques are necessary for the mass production of those lamps (de Groot and van Vliet [1986](#)).

Physical Principles of High-Pressure Sodium-Vapor Lamps

Radiation Properties

As mentioned above, the spectrum is dominated by the sodium resonance lines, which appear to cover a wide range in the visible spectral region due to their broadening. This enormous broadening is caused by two mechanisms. One is

Fig. 2 Spectrum of high-pressure sodium-vapor lamp (OSRAM NAV-T 70 W)



resonance broadening caused by the reabsorption of radiation. For this, one has to keep in mind that sodium pressure reaches 0.07–0.4 bar, and the energy of the upper excited levels is at 2.1 eV, which is even lower than the lower level of the Hg (I) 546 nm transition at 5.46 eV. This may motivate the dramatic self-reversed line contour of the sodium resonance lines (see Fig. 2). The second broadening mechanism is caused by foreign pressure broadening. HPS lamp containing mercury provides a mercury partial pressure of around 1 bar and a xenon partial pressure of around 0.2 bar. This causes a further broadening of the sodium lines by a quasistatic van der Waals broadening, which can be observed directly from the asymmetry of the line profile in Fig. 2. Furthermore, both gases act as buffer gases, which influence the electrical (Hg) and the heat conductivities but not the radiative properties by Hg or Xe atomic spectral lines.

Electrical and Thermal Conductivities

Despite the lower vapor pressure of sodium compared to mercury, sodium delivers most of the electrons to the plasma, because its ionization energy of 5.14 eV is lower than that of Hg at 10.4 eV. Mercury has a considerable influence on the electron mobility and can increase the effective electrical field strength. Xe affects the mobility of electrons slightly, but the increased thermal inertia results in lower temperature variations, and so the electrical reignition peak in every sinusoidal half cycle is reduced. The electric conductivity increases proportional to the electron density but decreases proportional to the electron momentum transfer cross section on the other side, which is dominated by the mercury vapor pressure. To keep the lamp voltage and compensate the increased electron density, arc tube length is increased. This leads to the characteristically high aspect ratio of sodium-vapor arc tubes. Heat losses are reduced by the lower thermal conductivity of the composition

which increases the light output. Electron and gas temperature results in a small difference, but complete local thermodynamic equilibrium does not exist, especially in regions near the wall (Waszink 1973). The balance between Na, Xe, and Hg determines the electrical (ignition, reignition, field strength, operation voltage) and radiative (spectral distribution, luminous efficacy) properties. All together have to be in relation to the length and diameter of the discharge tube. Modern high-pressure sodium lamps without mercury are operated with electronic ballasts.

Power Balance

Looking at the power balance of HPS lamps, it becomes obvious that the increased luminous efficacy is due to decreased UV radiation losses and an increased power fraction that is emitted in the visible spectral range. In Jack and Koedam (1974), this is quantified, where 30 % of the input power is emitted into the visible spectral range, and only 0.5 % is emitted into the UV. Furthermore, it must be noted that the sodium resonance lines are placed nearby the maximum of the photopic luminous efficacy of the human eye at 555 nm.

Electrode Phenomena

Electrode phenomena were under investigation for the improvement of lamp lifetime (Saito and Murakami 2009). In Elenbaas (1972) it reported about electrode construction, and additionally special effects at electrodes improve the lamp behavior (Hartmann et al. 2010).

Pulsed High-Pressure Sodium-Vapor Lamps

A pulsed current waveform enables mercury-free HPS lamps (e.g., COLORSTAR by OSRAM) with an adjustable color temperature (Günther et al. 1990). With such an operation regime, additionally stimulated transitions cause an improved color rendition. An example for a pulsed current waveform is given in Fig. 3, where the inner diagram shows the current versus time and the spectra of the occurrence of the changed line intensities and broadening.

In contrast to pulsed high-pressure sodium-vapor lamps, a pulsed operation of metal-halide lamps usually does not improve the spectral distribution of radiation. In sodium lamps, the spectrum is dominated by resonance lines, and a pulsed operation leads to a redistribution of population densities with increased excitation of higher levels. This leads to an enriched spectrum. In metal-halide lamps, there is almost a rich spectrum with numerous lines all over the visible spectral range, and a pulsed operation would increase the population densities of excited ionized species leading to a higher UV output.

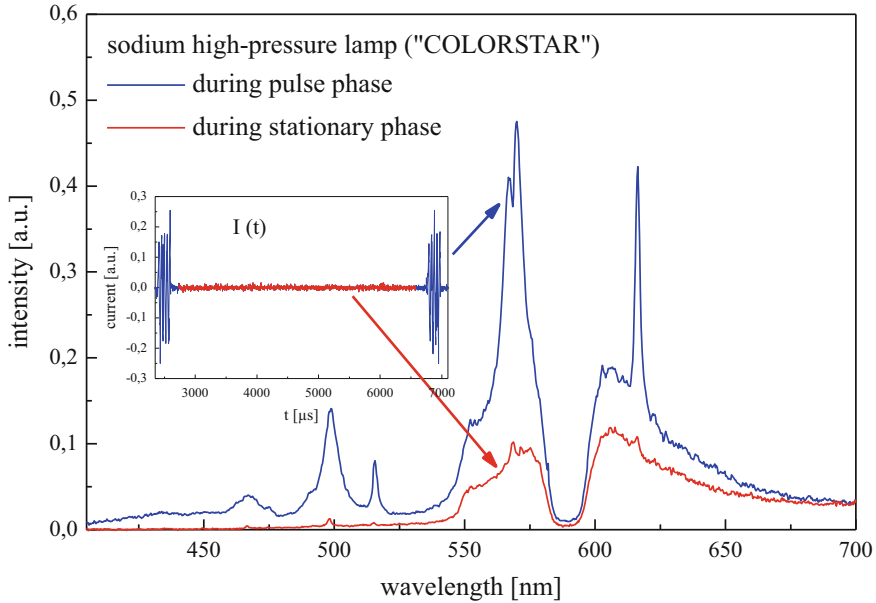


Fig. 3 Spectra of a pulsed high-pressure sodium-vapor lamp (COLORSTAR)

HPS Lamps with Improved Color Rendering

It is possible to increase the color rendering index of HPS lamps by attaching heat shields to the ends of the arc tube which produces higher cold spot temperatures and thus higher sodium pressures. This type of design results in a CRI of around 60. However, this is accompanied by a lower lamp efficacy when compared to a standard HPS lamp of the same power (Hollo 1997). The higher arc tube temperatures also lead to a reduced lifetime. For example, while a standard 250 W HPS lamp (CRI 25) might have a rated efficacy of 130 lm/W, the corresponding high color rendering version (CRI 60) might have a rated efficacy of 93 lm/W. The corresponding rated lamp lives are 40,000 and 24,000 h, respectively. The design and construction of HPS lamps are considered in detail in Geens and Wyner (1993) and Hollo (1997).

Note About Historical Street Lighting

There is a further plasma light source used for street lighting up to the present, which is gas lighting. Up to now, thousands of gas lanterns illuminate numerous streets in Berlin and can be found also in other cities of Germany or the United States. In the so-called Welsbach mantle (Auer-Glühstrumpf), the gas is transformed by a chemical process into visible light with a yellowish spectrum like the HPS lamp. As any combustion flame contains ionized particles and free electrons, it can be considered as plasma even if it is maintained by a chemical process and not by an electric current (Fig. 4).

Fig. 4 Photography of a gas lantern in Berlin which is still used for street lighting



Nowadays, gas lanterns did not only survive graphite arc lamps, which are another kind of historical street lighting. Graphite arc lamps have been replaced by more efficient high-pressure mercury-vapor lamps, which again have almost been replaced by more efficient high-pressure sodium-vapor lamps. In the last decade, a number of street lights were replaced by efficient white-light metal-halide lamps. However, at the moment, LED street lighting is more and more penetrating the outdoor-lighting market.

Directions for Future Research

It is not expected that high-pressure sodium-vapor lamps are in the focus of future research, as LEDs penetrate the outdoor illumination market.

Acknowledgments The authors kindly acknowledge valuable contributions to this chapter by Stuart Mucklejohn.

References

- de Groot JJ, van Vliet J (1986) The high-pressure sodium lamp. Kluwer technische Boeken B.V, Deventer
- Elenbaas W (1972) Light sources. Macmillan, London
- Geens R, Wyner E (1993) Progress in high pressure sodium lamp technology. IEE Proc A 140 (6):450–464
- Günther K, Kloss H-G, Lehmann T, Radtke R, Serick F (1990) Pulsed operation of high-pressure-sodium discharge lamps. Contrib Plasma Phys 30(6):715–724
- Hartmann T, Guenther K, Mentel J (2010) The gas phase emitter effect at the anode in a high pressure sodium vapour discharge. J Phys D Appl Phys 43(2):025201
- Hollo S (1997) High-pressure sodium discharge lamps. In: Coaton JR, Marsden AM (eds) Lamps and lighting, 4th edn. Routledge, New York, pp 235–248

-
- Jack AG, Koedam M (1974) Energy balances for some high pressure gas discharge lamps. J IES, Annual IES conference. Light Division, N.V. Philips' Gloeilampenfabrieken, Eindhoven, pp 323–329
- Saito N, Murakami K (2009) Electrode constructions of high pressure sodium lamp and their life characteristics. J Light Vis Environ 33(3):131–136
- Waszink JH (1973) Nonequilibrium calculation on an optically thick sodium discharge. J Phys D Appl Phys 6(8):1000–1006

High-Pressure Xenon Lamps

Heinz Schöpp and Steffen Franke

Contents

Introduction	1105
Historical Overview	1105
Types of Xe Lamps, Applications	1106
Basic Working Principle (Difference to Hg Lamps)	1106
Technological Aspects	1106
Physical Principles	1107
Directions for Future Research	1108
References	1109

Abstract

In the family of high-pressure discharge lamps, up to now high-pressure xenon lamps provide one of the most homogeneous spectral distributions over a broad spectral range of light. This makes them still to preferred light sources in cinema projection, to give one example. Although the discharge plasma consists only of single element, scientific research is addressed to this kind of lamps up to the presence.

Introduction

Historical Overview

First high-pressure xenon lamps were developed in the 1940s (Schulz 1947). It was found that they emit a broad spectrum like sunlight, in contrast to the spectrum of high-pressure mercury-vapor lamps which is characterized by a limited number of

H. Schöpp • S. Franke (✉)
INP Greifswald, Greifswald, Germany
e-mail: schoepp@inp-greifswald.de; steffen.franke@inp-greifswald.de

spectral lines. Although different noble gases have been investigated, xenon revealed to provide the most efficient light output. Comparing the design of the very first xenon lamps with present designs, only small differences are found, whereas the design of mercury lamps and metal-halide lamps underwent significant changes throughout the last decades. This already indicates that the physics tightly defines the requirements on the discharge vessels.

Types of Xe Lamps, Applications

There are different types of high-pressure xenon lamps. Xenon projection lamps are available for lamp powers ranging from 50 W up to 15 kW and currents between 2.5 A and 1 kA. Xenon is also widespread in automotive headlight with or without additions of mercury. Here the lamp power usually is around 35 W. Medium pressure xenon lamps are also used as flash lamps in photo cameras, because they provide intense white light within microseconds. But they are also applied as aircraft warning lights. Medium pressure xenon lamps have a different design and operating regime compared to high-pressure xenon lamps.

Basic Working Principle (Difference to Hg Lamps)

The main difference to high-pressure mercury-vapor lamps is that xenon itself has to guarantee a sufficient lamp voltage by its electron momentum transfer cross section. Therefore xenon is filled at high pressures into the discharge vessels. On the counterpart this has the consequence that a high ignition voltage of several tens of kV is necessary. An advantage of xenon lamps is that they provide almost white light just after ignition. This is of particular importance for automotive headlight lamps.

Technological Aspects

The technological challenges of high-pressure xenon lamps depend on the lamp type. Xenon projection lamps are characterized by a very short electrode distance and comparably high electrode diameters. They are operated with a DC current requiring a power source that is able to supply up to 1 kA of current.

A principle scheme of this type of lamps is given in Fig. 1. The design is similar to high-pressure mercury-filled lamps and consists of a thick-walled quartz bulb with an elliptically formed tube and feed-through of the electrodes. The cathode is light, slender, and precisely designed because the electrode emission mechanism and especially the work function lead to a cooling of this electrode, whereas the anode is thick and massive to hold the energy balance. The quartz tube material is much thicker than in usual mercury lamps to withstand the much higher pressure of some ten bars. The material must be very pure to avoid reabsorption of radiation in the wall. The plasma of the arc is very constricted and this point-shaped form can be well

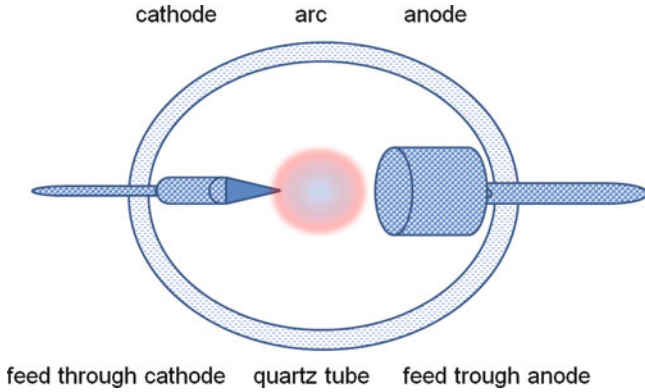
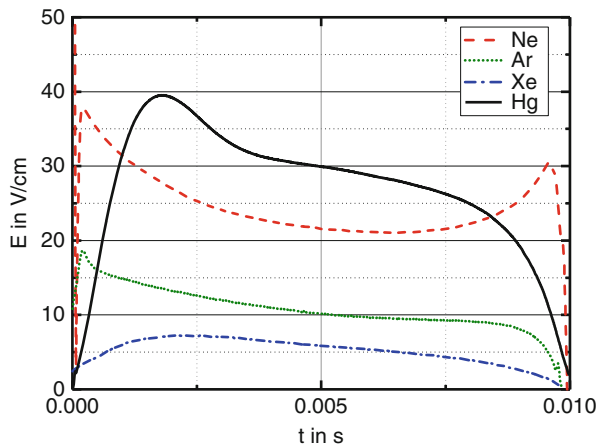


Fig. 1 Principal scheme of a high-pressure xenon lamp

Fig. 2 Field strength of noble gas discharges driven with conventional ballasts compared to mercury



imaged in an optical system. The exhaust tube is arranged so that it does not disturb the image. During the production process, this exhaust tube is used to evacuate the bulb, flush, and fill to appropriate gas pressure. After that procedure is completed the exhaust tube is sealed close to the arc tube wall.

Physical Principles

Noble gas arc discharge lamps are the most simple plasma lamps one can imagine, because there is only one-component plasma which does not even need evaporation, like in mercury lamps. However, despite the high ionization energy of noble gases (compared to mercury), the electric field strength of the plasma is low due to the small electron momentum transfer cross section of noble gases (compared to mercury). In Fig. 2 the electric field strength of discharges in noble gases is compared to

a discharge in mercury. The noble gas with the highest ionization energy has the highest field strength, which is still lower than for a comparable mercury lamp. This corresponds to the breakdown voltages given in (Hess 1976). Moreover deviations from local thermal equilibrium (LTE) can be expected with decreasing mass of noble gas atoms, which means that electrons may have higher temperatures than the heavy particles. This can be explained as follows. The electrons gain energy in the electric field, which can be dissipated between electrons by Coulomb interaction of charged particles. However, if the electron momentum transfer cross section and buffer gas density are not high enough, this energy cannot be dissipated effectively to the heavy particles, which remain at lower temperatures. Zalach et al. proofed that in xenon discharges at pressures around 8 bar and 3 A, already LTE can be assumed (Zalach and Franke 2013), where also a procedure is given to estimate the operating pressure and plasma temperature. Another approach to estimate the operating pressure by an empirical formula is given by (Motomura et al. 2010).

Finally, the reduced field strengths lead to increased currents at same power levels and hence to an increased thermal load of the electrodes, potentially shortening the lifetime of the lamp. Furthermore the luminous efficacy and the color rendering of noble gas discharges are poor, as the distribution of spectral lines over the visible spectral range is not optimal for lamps and excitation energies are higher than, e.g., for important mercury spectral lines. The more lightweight noble gases like helium and neon show an increasing permeation through the hot quartz wall with increasing temperature (Norton 1957). The fact that despite all the disadvantages given above there is a high-pressure noble gas lamp on the market is in the recombination continuum of xenon (D'Yachkov et al. 1998; Hofsaess 1978; Radtke and Kettlitz 1992). Xenon is the heaviest nonradioactive noble gas and has the lowest ionization energy. Hence there are high electron densities at comparably low plasma temperatures supporting a pronounced recombination continuum which occurs, if free electrons recombine with xenon ions. The xenon recombination continuum is very broad ranging from the ultraviolet (UV) up to the near infrared (NIR). Even in the visible spectral range, the spectral distribution is very homogeneous leading to color rendering properties which are even better than for mercury ultra-high pressure (UHP) lamps. One of the first scientific publications dealing with the investigation of high-pressure xenon discharges is that one of Neumann (1956). Instabilities of xenon arcs caused by plasma-anode interaction are investigated in Benilov and Hechtfisher (2012). Worth noting is also the high-pressure xenon flash lamp which is driven in pulsed operation and makes use of the xenon continuum to provide almost black-body radiation with a temperature of about 12,000 K (Günther and Radtke 1975).

Directions for Future Research

No significant technological progress is expected for high-pressure xenon lamps. However, they may keep some interest as a subject to basic research, because the plasma consists only of one element.

Acknowledgments The authors kindly acknowledge valuable contributions to this chapter by Hartmut Schneidenbach.

References

- Benilov MS, Hechtfischer U (2012) Stability of very-high pressure arc discharges against perturbations of the electron temperature. *J Appl Phys* 111(7):073305–073308
- D’Yachkov LG, Kurilenkov YK, Vitel Y (1998) Radiative continua of noble gas plasmas. *J Quant Spectrosc Radiat Transf* 59(1–2):53–64
- Günther K, Radtke R (1975) A proposed radiation standard for the visible and UV region. *J Phys E Sci Instrum* 8:371–376
- Hess H (1976) *Der elektrische Durchschlag in Gasen*. Vieweg, Braunschweig
- Hofsaess D (1978) Emission continua of rare-gas plasmas. *J Quant Spectrosc Radiat Transf* 19(3):339–352
- Motomura H, Enoki K, Jinno M (2010) Non-destructive measurement of Xe filling pressure in mercury-free metal halide lamp. *J Phys D Appl Phys* 43(23):234003
- Neumann W (1956) Spektralphotometrische Messungen an einer wandstabilisierten Xenon-Hochdruckentladung. *Ann Phys* 452(2–3):146–154
- Norton FJ (1957) Permeation of gases through solids. *J Appl Phys* 28(1):34–39
- Radtke R, Kettlitz M (1992) On the energy balance of isothermal wall-stabilized xenon arcs. *Plasma Sources Sci Technol* 1(4):274–279
- Schulz P (1947) Elektrische Entladungen in Edelgasen bei hohen Drücken. I. Bogenformen und spektrale Eigenschaften der Edelgashochdruckentladungen. *Ann Phys* 1(6):95–106
- Zalach J, Franke S (2013) Iterative Boltzmann plot method for temperature and pressure determination in a xenon high pressure discharge lamp. *J Appl Phys* 113(4):043303–043307

Metal-Halide Lamps

Steffen Franke and Heinz Schöpp

Contents

Introduction	1112
Historical Overview	1112
Basic Working Principle (Difference to Hg Lamps)	1112
Types of MH Lamps, Applications	1112
Technological Aspects of Metal-Halide Lamps	1115
Physical Principles of Metal-Halide Lamps	1116
Radiation Properties	1116
Power Balance	1117
Electrical Conductivity	1117
Electrode Phenomena	1117
Electronic Ballasts, Acoustic Resonances, and Pulsed Operation	1118
High-Temperature Material Chemistry, Demixing, and Segregation	1118
Re-ignition and Hot Relight	1119
Mercury-Free High-Pressure Lamps	1120
Directions for Future Research	1120
References	1120

Abstract

Silica metal-halide lamps (MH lamps) are responsible for the breakthrough of high-pressure discharge lamps in the general lighting market as they combine high-luminous efficacy and good color rendering properties. This was not achieved so far by technologies like high-pressure mercury lamps or sodium-vapor lamps. The progress was possible by the addition of compounds, which provide numerous spectral lines in the visible spectral range.

S. Franke (✉) • H. Schöpp
INP Greifswald, Greifswald, Germany
e-mail: steffen.franke@inp-greifswald.de; schoepp@inp-greifswald.de

Introduction

Historical Overview

Simultaneous to the development of high-pressure sodium-vapor lamps in the 1960s of the twentieth century, metal-halide lamps were developed to improve the radiative properties of mercury lamps (efficacy and color rendering). Mainly the rare earth elements but also Sc, Na, Tl and In were in focus as part of the lamp fill in form of metal halides. The additional elements contribute a multitude of spectral lines and therefore improve the color rendering index of the lamp. In contrast to high-pressure sodium-vapor lamps, first metal-halide lamps were made of cost-effective quartz burners. Although cylindrical discharge vessels made of polycrystalline alumina (PCA) were available for many years, a significant progress in lamp technology was achieved with the development of ellipsoidal PCA burners with arc tube seals that were chemically resistant over the lamps' lifetime in the first decade of the twenty-first century (see chapter “► [Ceramic Metal Halide Lamps](#)” by Stuart Mucklejohn).

Basic Working Principle (Difference to Hg Lamps)

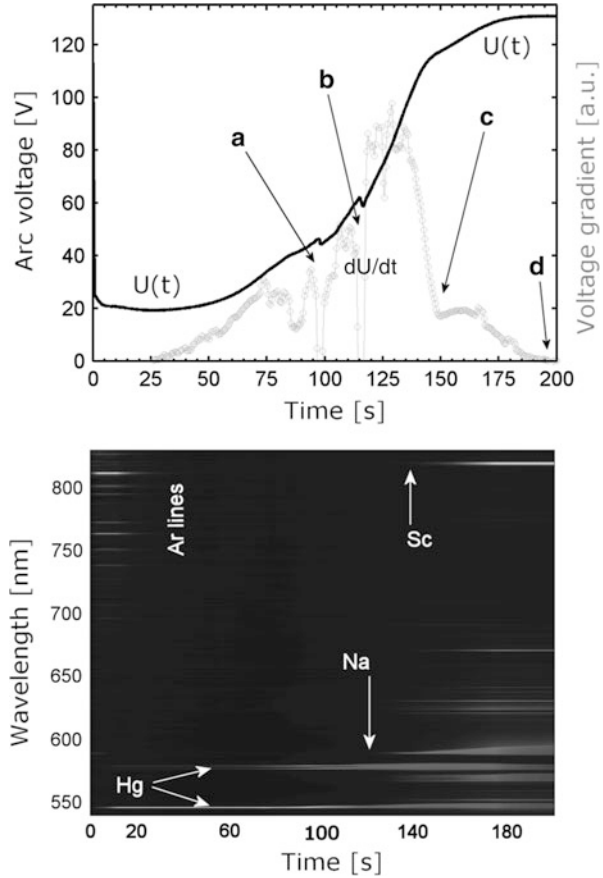
The basic working principle is generally the same as for the high-pressure sodium-vapor lamps. The main difference is that there is a dynamic transition during the warm-up phase from a noble gas arc to a mercury-dominated arc to an arc with different additives entering the discharge region when the cold spot temperature increases. For an Ar/Hg/NaI/ScI₃-fill system, the start-up behavior is given in Fig. 1. The additives modify the warm-up dynamics leading to additional structures in the arc voltage evolution (compared to an Hg lamp). These are related to reservoir evaporation, which begins above material-specific threshold temperatures. Finally a change of line intensities and profiles can be observed. Every additional species introduces a radiative loss channel (Zalach et al. 2012). In Fig. 1, the line intensity evolution shows the change of the dominant species. The changes are well observable in the gradient view of the voltage characteristic. A decrease of the voltage gradient over time indicates the almost complete evaporation of a reservoir of liquid fill material and the beginning evaporation of a material with lower vapor pressure (see points a and b in Fig. 1).

Stationary operation can be achieved within few minutes by electronic control gears which accelerate the warm-up phase by increased currents within the first seconds after breakdown (Lin et al. 2011; Zalach et al. 2011). Due to the low-vapor pressure of many metal halides, a saturated vapor is found for these compounds.

Types of MH Lamps, Applications

A striking feature of metal-halide lamps is the opportunity to adjust the color temperature of the spectrum by the composition of the fill materials. Elements

Fig. 1 Characteristic structures of $U(t)$ better visible in a gradient view (*upper part*). (a) Begin of sodium evaporation, (b) begin of scandium evaporation, (c) complete evaporation of sodium and mercury, (d) complete evaporation of scandium. Line intensity evolution shows the change of the dominant species (*lower part*)



with strong blue spectral lines like In can be combined with Tl, which emits particularly in the green, or with Na, that is characterized by its yellow resonance lines. Iodides of the lanthanides contribute a multitude of spectral lines distributed over the whole visible spectral range with typical accents in different spectral regions. Lamps are available with warm daylight (3,000 K), neutral daylight (4,200 K), and daylight (5,200 K) with nominal lamp powers ranging from 35 W to some thousands of watts. Applications are shop lighting, flood lighting, and exterior lighting, say everywhere where one lamp shall illuminate a large area at a high illuminance level (see Table 1). The general color rendering index (CRI) typically is between 80 and 90 for general lighting applications. For some special applications, the CRI might be lower. For general lighting silica, metal-halide lamps are almost phased out by some manufacturers and replaced by ceramic metal-halide lamps with ellipsoidal burners. A special application of silica metal-halide lamps is in automotive headlights. Here, the transparent quartz bulb is inevitable for a proper projection of light. An overview of metal-halide lamps with silica arc tubes up to the mid-1990s is provided in Preston and Odell (1997).

Table 1 Types of MH lamps with photometric properties and applications

Type	Lamp input power (W)	Luminous flux (lm)	Luminous efficacy (lm/W)	General color rendering index	Operating pressure (bar)	Main application
OSRAM	70–2,000	6,000–215,000	ca. 100	70–90	5–10	General lighting, spot light
Philips	145–1,000	15,000–107,000	ca. 100	65–90	5–10	General lighting, indoor – outdoor
GE	50–2,000	3,200–170,000	87–100	65–70	5–10	General lighting, sportlight
D2	35	3,200	90	Not specified	ca. 50	Automotive headlights

Technological Aspects of Metal-Halide Lamps

The discharge vessels of metal-halide lamps are similar to that ones used for high-pressure mercury-vapor lamps (Fig. 2). The burner is made of quartz and the electrode is made of tungsten. The feedthrough is realized by molybdenum foil. In contrast to the mercury-vapor lamp, the outer bulb is essential for lamp operation to avoid axial segregation of plasma components that would be amplified by convective cooling. Hence, the outer bulb is needed for the thermal management. Furthermore, it is doped with elements that shift the UV absorption edge toward 400 nm. As a burner material, quartz of highest available purity is advantageous as it can withstand temperatures up to 1,400 °C. This is required to ensure sufficient evaporation of metal halides. The electrode regions of the quartz tube are covered with an infrared-reflecting material like ZrO or corundum particles (Al_2O_3). This reduces thermal losses in these regions and heats up the cold spot. A further contribution to the heating of the cold spot is the reabsorption of plasma radiation in the tube wall (see also chapter “► [Mercury-Vapor Lamps](#)”). A critical process in metal-halide lamps is the corrosion of the molybdenum foil by the reactive chemicals introduced into the lamp. Particularly during switching on and switching off the lamp, there is an increased risk of lamp damage and corrosion due to the different thermal expansion coefficients of molybdenum and quartz. Furthermore, the chemicals react on the tungsten electrode itself. Therefore, destruction of electrodes and wall blackening are further reasons of limited lifetime. Finally, a corrosion of the quartz tube takes place, driven by temperature gradients along the tube wall and the transport of material in the liquid and gaseous phase. This effect can be recognized by a decreasing transparency of the quartz tube, which is going to become translucent by recrystallization processes.

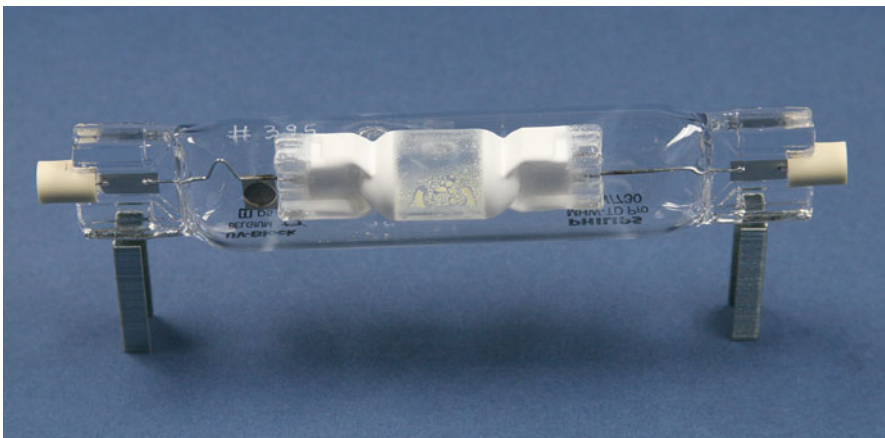


Fig. 2 Photography of a silica metal-halide lamp (Philips MHW-TD Pro 150W/730)

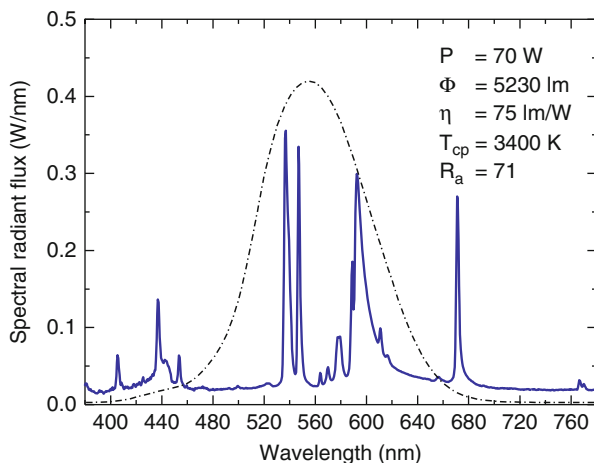
The chemical properties of the fillings influence considerably the corrosion of both electrodes and quartz tube. These processes determine the lifetime of the lamps by blackening the wall and destruction of the electrode.

Physical Principles of Metal-Halide Lamps

Radiation Properties

The radial temperature profile of a pure mercury-vapor arc can be approximated by a parabola. Adding a small amount of InI, TlI, or NaI, this temperature profile becomes flattened. This is due to the redistribution of energy by emission of atomic radiation in the arc core that is reabsorbed at outer radial positions. Further increase of the amount of fill substances will then lead to a constriction of the temperature profile, as a new energy loss channel comes up, that is, molecular radiation. Molecules usually are stable at a plasma temperature below 4,000 K but are excited for radiation at a temperature above 3,000 K. Hence, the arc is cooled at medium radial positions. Besides the molecular radiation, it seems that iodine itself affects the radial temperature profile of the arc. As outlined in Franke et al. (2007), the electronegativity of iodine may reduce the electrical conductivity, thus reducing ohmic heating, which causes a decrease of the plasma temperature and a constriction of the radial profile. One of the most important facts is to accept that in metal-halide lamps, molecular radiation, e.g., of the TmI molecule, contributes significantly to the total radiative output. As proofed in Käning et al. (2007), around 50 % of radiative power in the visible spectral range is caused by a molecular quasi-continuum, which is dominant at medium radial positions. For TmI, DyI, and HoI, two molecular emission bands have been identified see (Fig. 3) for a typical spectrum of a MH lamp.

Fig. 3 Spectrum of an OSRAM Powerstar HQI-T 70 W/830. The *dashed line* gives the relative spectral luminous efficiency



Power Balance

The total input power of metal-halide lamps distributes into the different loss channels as follows. Into the visible spectral range, 30 % of the input power is emitted, leading to a luminous efficacy around 100 lm/W at a general color rendering index (CRI) of about 80–90. The same percentage of energy is dissipated into the UV and NIR. Non-radiative losses (heat conduction to the tube wall and thermal radiation of the discharge vessel) account again for 30 %. The remaining 10 % of input power is consumed by the electrode regions; see, e.g., (Jack and Koedam 1974; Nelson et al. 2001).

Electrical Conductivity

Another contribution of the metal halides is their function to deliver electrons. Compared to mercury, almost all additives show lower ionization energies. Therefore, the effect of mercury on the electric conductivity now is mainly due to its large electron momentum-transfer cross section. Hence, it can be concluded that the metal halides do not act as lighting additives only, because they affect the discharge physics far beyond their emitted radiation by their contribution to the electrical conductivity and the energy redistribution inside the arc by radiative transfer.

Electrode Phenomena

Most research is concentrated on cathode effects, where the work function is of particular interest. Different cathode phenomena can be observed in high-pressure discharge lamps. Basically one has to distinguish between field emission of electrons during the ignition and warm-up of a lamp and the thermionic emission of electrons which takes over when electrodes are hot enough. Hence, in stationary operation, thermionic emission is expected (see also chapter “► [Mercury-Vapor Lamps](#)” for further aspects of electrode phenomena). The arc attachment at electrodes can be distinguished between a diffuse and a spot mode, where the cathode prefers the spot mode and the anode, the diffuse mode. Loss of electrode material can be driven by sputtering effects at low temperatures and high sheath voltages or by evaporation of material at high electrode temperatures. Electrodes have an optimum operation regime with minimum sputtering and evaporation.

There are additional effects relevant in metal-halide lamps compared to mercury-vapor lamps. During the last years, the gas-phase emitter effect attracted some attention in research. It was found that some metal iodides generate monolayer coatings on the electrodes and lower the work function (Luijks et al. 2005; Schmidt et al. 2013; Westermeier et al. 2013a, b). This reduces the electrode sheath voltage and hence sputtering effects, leading to an increased lifetime of electrodes. On the contrary, some materials tend to form an alloy with the tungsten on top of the electrodes, which have a reduced melting point and hence contribute to the material

loss at the electrode. However, for many common fill compositions, the gas-phase emitter process seems to be dominant. Especially Dy, Ho, and Tm in combination with Na and/or Tl are investigated and show a reduction of the electrode tip temperature (Westermeier et al. 2013a, b). Important contributions to the theoretical understanding of electrode phenomena in HID lamps are reported in Benilov et al. (2004), Bötticher and Kettlitz (2006), Dabringhausen et al. (2002), Kettlitz et al. (2005), and Luhmann et al. (2002). Further reading: Dabringhausen et al. (2005), Flesch (2006), Kettlitz and Grossjohann (2002), Kloss et al. (2000), Lichtenberg et al. (2005), Pursch et al. (2002), and Redwitz et al. (2005).

Electronic Ballasts, Acoustic Resonances, and Pulsed Operation

Conventional ballasts made of copper coils suffer from several disadvantages. They are heavy, made of cost-intensive copper, and dimensioned for a given power line frequency and voltage. Furthermore, they consume around 15 % of the lamp power. Therefore, one advantage of electronic ballasts, also called electronic control gears, is their reduced power consumption, which accounts to approximately 10 % of the lamp power. At this point, it is important to introduce the term “system efficiency,” which relates the luminous flux to the sum of the lamp power and the power consumption by the ballast. Electronic ballasts operate at 120 Hz square wave including an ignition circuit applying high-frequency pulses within the first microseconds.

An additional advantage of electronic ballasts is the opportunity to provide higher frequencies in stationary operation to excite or suppress acoustic resonances. This can be utilized for stabilization and straightening of the arc. But it can also be used to act against axial segregation and radial demixing of the plasma components (Dreeben 2008; García-García et al. 2006; Hamady et al. 2013; Olsen and Dreeben 2011; Schneidenbach et al. 2008; Stockwald et al. 2008). Pulsed operation is of minor importance for metal-halide lamps. Baksht reported over decades on pulsed cesium discharge lamps with a detailed experimental and theoretical description of the plasma radiation (Baksht and Moizhes 1965; Baksht and Lapshin 2013).

High-Temperature Material Chemistry, Demixing, and Segregation

In metal-halide lamps, high-temperature chemistry plays an important role. Usually it is distinguished between the chemistry in the liquid phase and the chemistry in the gas phase. High-temperature chemistry in the liquid phase is mainly attributed to corrosion effects at the plasma boundaries, even though the material transport via the gas phase is also important for these processes. The high-temperature chemistry in the gas phase includes all aspects of the temperature-dependent and spatial-dependent plasma composition. This topic covers the formation of complexes in the liquid phase, which may lead to a change of vapor pressure of chemical compounds and hence affecting the plasma composition in the gas phase. For instance, it is known that complexes of NaI and DyI_3 , i.e., $NaDyI_3$ and Na_2DyI_5 ,

cause a reduction of the sodium-vapor pressure but an increase of the dysprosium vapor pressure; see also (Curry et al. 2013; Hilpert and Niemann 1997). In the gaseous phase, one has to consider the dissociation of molecules which depends on the temperature. Hence, one can find maxima in the density of components at different radial positions. Demixing effects describe deviation from stoichiometric ratio of the fill substances, for instance, caused by much higher diffusion coefficient of the atoms compared to the diffusion coefficients of molecules (Schalk et al. 2003). Another topic that is related to the demixing is the axial segregation that means a demixing of plasma components in vertical operation. This effect is driven by convection of material and leads, e.g., to an accumulation of dysprosium at the lower part of the discharge due to a previous radial demixing (Flikweert et al. 2006, 2008). In Murphy and Hiraoka (2000), different effects leading to demixing are discussed: “The most important are demixing due to partial pressure gradients, which concentrates the chemical elements with the higher ionization energies in the high-temperature regions of the plasma; demixing due to frictional forces and due to thermal diffusion, both of which concentrate the lighter chemical elements in the high-temperature regions; and cataphoresis or demixing due to electric fields, which concentrates the chemical elements with the higher ionization energies near the anode.” See also (Murphy et al. 2008).

The determination of the composition in such discharges is problematic by emission spectroscopy. So other methods have to be used. Especially, absorption methods are suitable. Proved and tested for determination of ground-state and excited-state densities, e.g., in Hg, are the Hook-method (Kindel et al. 1998) and x-ray-based density measurements (Curry et al. 2004, 2011; Denisova et al. 2008).

Further diagnostic methods that gained importance for the experimental investigation of high-pressure lamps are self-absorption experiments with an additional mirror behind the lamp and absorption measurements with an auxiliary light source or more recently with tunable laser diodes. Hook method provided ground-state densities even at outer radial positions. X-ray absorption diagnostics proved to be a valuable but expensive diagnostic method.

Re-ignition and Hot Relight

The advantage of mercury is that it evaporates at comparable low temperatures. However, that means it condenses very slowly after switching off the lamp. But even a small mercury partial vapor pressure hinders re-ignition of the discharge. Typically it takes several minutes before a high-pressure metal-halide lamp can be switched on again. Hot relight (or re-strike) is usually realized by increased ignition voltage peaks (Czichy et al. 2008; Lapatovich et al. 2007; van den Bos et al. 2013), if necessary with some auxiliary electrode. Automotive headlights require a hot-relight feature. Further lamps for special applications, e.g., studio lighting and sports stadia, are also designed for hot re-strike. However, these use high-voltage pulses of ca. 50 kV. The lamps, sockets, and luminaires are carefully designed to withstand these very high voltages.

Mercury-Free High-Pressure Lamps

As mercury is known to be potentially hazardous for the environment, there has been significant effort in research to omit mercury in metal-halide lamps. To give some examples of mercury-free lamps, it is pointed to microwave-driven sulfur lamps (Turner et al. 1997), dielectric-barrier discharges in xenon (Müller and Zahn 1996; Stockwald and Neiger 1995; Vollkommer and Hitzschke 1998), high-pressure xenon lamps for automotive lighting (Guenther et al. 2004), and the cluster lamp (Weber and Scholl 1993). In metal-halide lamps, Philips followed an approach to replace mercury by zinc, which is similar to Hg in some properties (ionization energy, excitation energies, electron momentum-transfer cross section) but suffers from a low-vapor pressure (Born 2001). This could be overcome by the introduction of ZnI providing sufficient high-lamp voltages. However, this compound is chemically much more aggressive to the tube walls than mercury. OSRAM followed an approach to replace mercury by a combination of xenon and AlI₃ leading to an enhanced molecular radiation (Käning et al. 2011; Franke et al. 2007). The voltage drop due to the absence of mercury is in parts compensated by a constricted temperature profile of the arc and by xenon at working pressures around 3 bar. More than 50 % of the radiation in the visible spectral range could be attributed to molecular emission. However, the chemical activity of the mercury-free fill cannot be compared to the almost inert behavior of mercury.

Directions for Future Research

Before LEDs gained increasing interest, research on metal-halide lamps was devoted to features like small wattage and hot relight. This topic receded into the background like new approaches to electrodeless metal-halide lamps. (See the section in this work by Graeme Lister for more information “► [Electrodeless Lamps and UV Sources](#).”) One challenge still might be the replacement of mercury by more environment-friendly solutions as well as optimized operating regimes utilizing acoustic resonances by enhanced electronic control gears.

Acknowledgments The authors kindly acknowledge valuable contributions to this chapter by Stuart Mucklejohn, Hartmut Schneidenbach, and Manfred Kettlitz.

References

- Baksh FG, Lapshin VF (2013) On the efficiency of emission from a plasma column in high-pressure pulse-periodic discharge in cesium. *Tech Phys Lett* 39(10):847–850
- Baksh FG, Moizhes BY (1965) Theory of low voltage arc in cesium. *Sov Phys Tech Phys USSR* 10 (2):214
- Benilov MS, Cunha MD, Naidis GV (2004) Modelling interaction of metal halide plasmas with a thermionic cathode. In: Zissis G (ed), *Light Sources 2004*. Inst Phys Conf Ser (182):537–538

- Born M (2001) Investigations on the replacement of mercury in high-pressure discharge lamps by metallic zinc. *J Phys D Appl Phys* 34:909–924
- Bötticher R, Kettlitz M (2006) Dynamic mode changes of cathodic arc attachment in vertical mercury discharges. *J Phys D Appl Phys* 39(13):2715–2723
- Curry JJ, Adler HG, MacPhee A, Narayanan S, Wang J (2004) X-ray absorption imaging of Hg vapour in a ceramic metal-halide lamp using synchrotron radiation. *Plasma Sources Sci Technol* 13(3):403–408
- Curry JJ, Lapatovich WP, Henins A (2011) X-ray methods in high-intensity discharges and metal-halide lamps: X-ray induced fluorescence. *Adv At Mol Opt Phys* 60:65–117
- Curry JJ, Estupinan EG, Henins A, Lapatovich WP, Shastri SD, Hardis JE (2013) Enhancement of lanthanide evaporation by complexation: dysprosium tri-iodide mixed with indium iodide and thulium tri-iodide mixed with thallium iodide. *J Chem Phys* 139(12):124310
- Czichy M, Hartmann T, Mentel J, Awakowicz P (2008) Ignition of mercury-free high intensity discharge lamps. *J Phys D Appl Phys* 41(14):144027
- Dabringhausen L, Nandelstädt D, Luhmann J, Mentel J (2002) Determination of HID electrode falls in a model lamp I: pyrometric measurements. *J Phys D Appl Phys* 35:1621–1630
- Dabringhausen L, Langenscheidt O, Lichtenberg S, Redwitz M, Mentel J (2005) Different modes of arc attachment at HID cathodes: simulation and comparison with measurements. *J Phys D Appl Phys* 38(17):3128–3142
- Denisova N, Haverlag M, Ridderhof EJ, Nimalasuriya T, Mullen JJAM (2008) X-ray absorption tomography of a high-pressure metal-halide lamp with a bent arc due to Lorentz-forces. *J Phys D Appl Phys* 41(14):144021
- Dreeben TD (2008) Modelling of fluid-mechanical arc instability in pure-mercury HID lamps. *J Phys D Appl Phys* 41(14):144023
- Flesch P (2006) *Light and light sources: high-intensity discharge lamps*. Springer, Berlin/Heidelberg
- Flikweert AJ, van Kemenade M, Nimalasuriya T, Haverlag M, Kroesen GMW, Stoffels WW (2006) Axial segregation in metal-halide lamps under varying gravity conditions during parabolic flights. *J Phys D Appl Phys* 39:1599–1605
- Flikweert AJ, Beks ML, Nimalasuriya T, Kroesen GMW, Haverlag M, Mullen JJAM, Stoffels WW (2008) Semi-empirical model for axial segregation in metal-halide lamps. *J Phys D Appl Phys* 41(19):195201
- Franke S, Methling R, Hess H, Schneidenbach H, Schöpp H, Hitzschke L, Käning M, Schalk B (2007) Mercury-free high-intensity discharge with high luminous efficacy and good colour rendering index. *J Phys D Appl Phys* 40(13):3836–3841
- García-García J, Cardesin J, Ribas J, Calleja AJ, Corominas EL, Rico-Secades M, Alonso JM (2006) New control strategy in a square-wave inverter for low wattage metal halide lamp supply to avoid acoustic resonances. *IEEE Trans Power Electron* 21(1):243–253
- Guenther K, Hartmann T, Sarroukh H (2004) Hg free ceramic automotive headlight lamps. In: Zissis G (ed) *Institute of Physics Conference Series*. Iop Publishing Ltd., Bristol, pp 219–220
- Hamady M, Lister GG, Stafford L (2013) Emission spectra from direct current and microwave powered Hg lamps at very high pressure. *J Phys D Appl Phys* 46(45):455201
- Hilpert K, Niemann U (1997) High temperature chemistry in metal halide lamps. *Thermochim Acta* 299:49–57
- Jack AG, Koedam M (1974) Energy balances for some high pressure gas discharge lamps. *Journal of IES, Annual IES conference, Light Division*. N.V. Philips' Gloeilampenfabrieken, Eindhoven, pp 323–329
- Käning M, Schalk B, Schneidenbach H (2007) Experimental determination of parameters for molecular continuum radiation of rare-earth iodides. *J Phys D Appl Phys* 40(13):3815–3822
- Käning M, Hitzschke L, Schalk B, Berger T, Franke St, Methling R (2011) Mercury-free high pressure discharge lamps dominated by molecular radiation. *J Phys D Appl Phys* 44(22):224005
- Kettlitz M, Grossjohann R (2002) On the plasma constriction close to the electrodes of high-pressure mercury and sodium lamps. *J Phys D Appl Phys* 35(14):1702–1706

- Kettlitz M, Sieg M, Schneidenbach H, Hess H (2005) Lowering of the cathode fall voltage by laser exposure of the cathode in a high-pressure mercury discharge. *J Phys D Appl Phys* 38 (17):3175–3181
- Kindel E, Kettlitz M, Schimke C, Schöpp H (1998) Application of the hook method and emission spectroscopy for the determination of radial density and temperature profiles in high-pressure mercury discharges. *J Phys D Appl Phys* 31(11):1352–1361
- Kloss A, Schneidenbach H, Schöpp H, Hess H, Hitzschke L, Schalk B (2000) Electrode-sheath voltages in high-pressure mercury arcs. *J Appl Phys* 88(3):1271–1275
- Lapatovich WP, Budinger AB, Pereyra R, Lister GG (2007) Molecular influence on hot-relight in HID lamps. In: Liu M-Q, Devonshire R (eds) *Light Sources 2007. FAST-LS*, Shanghai, China, pp 111–112
- Lichtenberg S, Dabringhausen L, Langenscheidt O, Mentel J (2005) The plasma boundary layer of HID-cathodes: modelling and numerical results. *J Phys D Appl Phys* 38(17):3112–3127
- Lin D, Yan W, Hui SYR (2011) Modelling the warm-up phase of the starting processes of high-intensity discharge lamps. *IET Sci Meas Technol* 5(6):199–205
- Luhmann J, Lichtenberg S, Langenscheidt O, Benilov MS, Mentel J (2002) Determination of HID electrode falls in a model lamp II: Langmuir-probe measurements. *J Phys D Appl Phys* 35:1631–1638
- Luijks GMJF, Nijdam S, Von Esveld H (2005) Electrode diagnostics and modelling for ceramic metal halide lamps. *J Phys D Appl Phys* 38(17):3163–3169
- Müller S, Zahn RJ (1996) On various kinds of dielectric barrier discharges. *Contrib Plasma Phys* 36 (6):697–709
- Murphy AB, Hiraoka K (2000) A comparison of measurements and calculations of demixing in free-burning arcs. *J Phys D Appl Phys* 33(17):2183–2188
- Murphy AB, Boulos MI, Colombo V, Fauchais P, Ghedini E, Gleizes A, Mostaghimi J, Proulx P, Schram DC (2008) Advanced thermal plasma modelling. *High Temp Mater Processes US* 12 (3–4):255–336
- Nelson GJ, Gibson RG, Jackson AD (2001) An efficacy analysis of HID lamps. *J Illum Eng Soc* 30 (1):68–75
- Olsen J, Dreeben TD (2011) Experimental and simulated straightening of metal halide arcs using power modulation. *IEEE Trans Ind Appl* 47(1):368–375
- Preston B, Odell EC (1997) Metal halide lamps. In: Coaton JR, Marsden AM (eds) *Lamps and lighting*, 4th edn. Routledge, New York, pp 263–280
- Pursch H, Schoepp H, Kettlitz M, Hess H (2002) Arc attachment and fall voltage on the cathode of an ac high-pressure mercury discharge. *J Phys D Appl Phys* 35(14):1757–1760
- Redwitz M, Langenscheidt O, Mentel J (2005) Spectroscopic investigation of the plasma boundary layer in front of HID-electrodes. *J Phys D Appl Phys* 38(17):3143–3154
- Schalk B, Hitzschke L, Hartel G, Käning M (2003) Radial demixing in plasma composition calculations. In: *ICPIG, Proceedings of the international conference on physics of ionized gases*, Greifswald, pp 93–94
- Schmidt M, Schneidenbach H, Kettlitz M (2013) Pyrometric cathode temperature measurements in metal halide lamps. *J Phys D Appl Phys* 46(43):435202
- Schneidenbach H, Franke S, Wendt M, Baeva M (2008) Analysis of spectral line shapes in HID plasmas. In: 61th GEC, American Physical Society, Dallas. *Bull Am Phys Soc* 53(10): BAPS.2008.GEC.FTP1.83
- Stockwald K, Neiger M (1995) Some properties of a novel far UV xenon excimer barrier discharge light-source. *Contrib Plasma Phys* 35(1):15–22
- Stockwald K, Kaestle H, Weiss H (2008) Significant efficacy enhancement of low wattage metal halide hid lamp systems by acoustically induced convection configuration. *Proc. 35th International Conference on Plasma Science (ICOPS, Karlsruhe, Germany, 2008)*, p. 6C7. doi:10.1109/PLASMA.2008.4591142
- Turner BP, Ury MG, Leng Y, Love WG (1997) Sulfur lamps - Progress in their development. *J Illum Eng Soc* 26(1):10–16

- van den Bos R, Sobota A, Manders F, Kroesen GMW (2013) Note: measuring breakdown characteristics during the hot re-ignition of high intensity discharge lamps using high frequency alternating current voltage. *Rev Sci Instrum* 84(4):046103
- Vollkommer F, Hitzschke L (1998) Dielectric barrier discharge. In: Proceedings of the 8th international symposium on the science and technology of light sources. ICPIG, Greifswald, pp 51–60
- Weber B, Scholl R (1993) A new kind of light-generation mechanism: incandescent radiation from clusters. *J Appl Phys* 74(1):607
- Westermeier M, Ruhmann C, Bergner A, Denissen C, Suijker J, Awakowicz P, Mentel J (2013a) A study of electrode temperature lowering in Dy-containing ceramic metal halide lamps: II. An investigation of the converse effect of Tl and/or Na additives. *J Phys D Appl Phys* 46(18):185202
- Westermeier M, Ruhmann C, Bergner A, Denissen C, Suijker J, Awakowicz P, Mentel J (2013b) A study of electrode temperature lowering in Dy-containing ceramic metal halide lamps: I. The effect of mixtures of Dy, Tl and Na compared with pure Dy. *J Phys D Appl Phys* 46(18):185201
- Zalach J, Araoud Z, Charrada K, Franke S, Schoepp H, Zissis G (2011) Experimental and theoretical investigations on the warm-up of a high-pressure mercury discharge lamp. *Phys Plasmas* 18(3):033511–033519
- Zalach J, Franke S, Schöpp H (2012) Experimental characterization of the warm-up of mercury lamps. In: Devonshire R, Zissis G (eds) *Light Sources 2012*. FAST-LS, Troy, USA, pp 157–158

Ceramic Metal Halide Lamps

Stuart A. Mucklejohn

Contents

Introduction	1125
Arctube Design	1127
Arctube Seals	1128
The Evolution of Ceramic Metal Halide Lamps	1132
Lamp Performance	1132
Dimming	1136
End of Life Mechanisms	1137
The Next Generation of Ceramic Metal Halide Lamps	1137
Summary	1138
References	1139

Abstract

The evolution of metal halide lamps with ceramic arctubes, known as ceramic metal halide (CMH) lamps, is outlined from concept through prototypes to successful product realization and thence to a firmly established and highly reliable light source technology with a vast array of high-quality products.

Introduction

The potential advantages of ceramic arctubes for metal halide discharge lamps have been recognized for many years. In particular, it was evident from an early stage in the development of high-pressure sodium (HPS) lamps that the combination of metal halide dose chemistry and HPS ceramic arctube technology had the potential to produce light sources with much improved performance characteristics which would

S.A. Mucklejohn (✉)
Ceravision Limited, Sherbourne Drive, Tilbrook, UK
e-mail: stuart.mucklejohn@ceravision.com

appeal to end users. Ceramic arctubes offer three key advantages over fused silica arctubes:

Higher operating temperature giving increased luminous efficacy and better color rendering properties

Greatly reduced sodium loss from the plasma with respect to fused silica arctubes thus giving stable color through the life of the lamp

Improved control over arctube dimensions which gives rise to low color spread and which also helps with the control of lamp voltages during production

The major problem which confronted lamp technologists was how to make long life seals between the ends of the ceramic arctube and the metallic conductors supporting the electrodes.

It is only since the mid-1990s that metal halide lamps with ceramic arctubes have been commercially available (Carleton et al. 1997; Seinen 1995). The development of such lamps has a history dating back to the 1970s and was continually hampered by difficulties with corrosion and material transport. The first detailed publication describing a metal halide lamp having a ceramic arctube with commercial potential was published in 1982 (Brown et al. 1982). This lamp, illustrated in Fig. 1, contained a mixture of sodium and tin halides in a polycrystalline alumina (PCA) arctube with cermet closures and leadthroughs.

The evolution of low wattage metal halide lamps with ceramic arctubes to 1998 has been outlined previously (Mucklejohn 1998). A more general review of low wattage high-intensity discharge lamps was published in 2000 (Mucklejohn and Preston 2000). An overview of the development of ceramic metal halide lamps has also been included in a wider review of recent advances in lighting quality and energy efficiency (Mucklejohn 2012).

The kinetics of chemical interactions between polycrystalline alumina and some metal halide systems are such that arctubes can operate at temperatures up to $\sim 1,150$ °C compared to a maximum of ~ 950 °C for fused silica. The high operating

Fig. 1 150 W ceramic metal halide lamp described in reference (Brown et al. 1982). Image from www.lampstech.co.uk

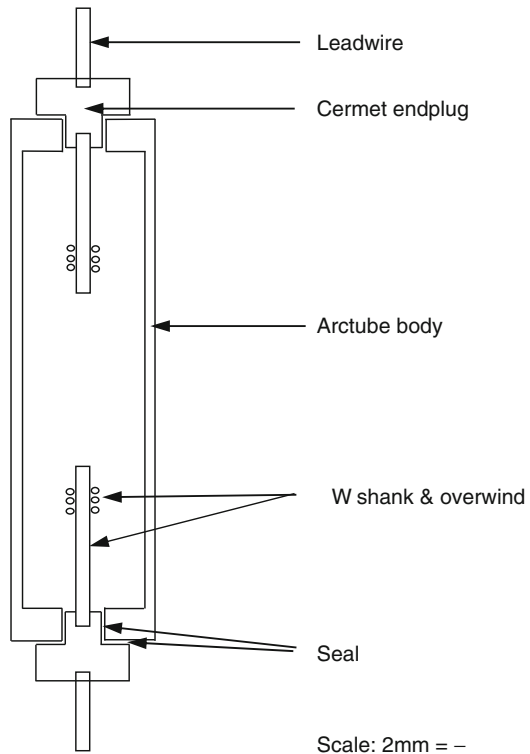


temperatures of lamps with ceramic arctubes, however, do give rise to many potential materials problems. Thus, the early development of metal halide lamps with PCA arctubes was beset by difficulties with corrosion which offset the performance advantage over fused silica arctubes. The first generation of commercially available ceramic metal halide lamps, covering the range 35–150 W, was an immediate success as the materials problems that had, in the past, limited life and lumen maintenance to levels below that acceptable to the market had been addressed in an elegant and reliable design.

Arctube Design

Brown et al. (1982) describe a 150 W lamp in a tubular outer jacket. The PCA arctube body is a right cylinder with electrically conducting cermet end plugs; see Fig. 2. The leadwire and electrode assembly are sintered separately into the cermet component. The end plugs are sealed to the arctube body with a high melting temperature glass which is designed to be resistant to chemical attack by the dose and stable at high temperatures during lamp operation. A notable feature of this design is the disparity between the arc gap (~15 mm) and the internal length of the arc chamber (~33 mm). The corresponding difference is much smaller in the current

Fig. 2 Schematic diagram of the arctube in the ceramic metal halide lamp described in Brown et al. (1982)



designs, as described below. This large difference was required to maintain seals at reasonable temperatures while still providing cool spot temperatures sufficiently high to give enhanced performance, but in spite of this, the poor seal reliability was one of the factors that prevented commercialization of these designs.

The problem of avoiding rapid chemical degradation of seals by metal halides, while permitting a high-dose temperature, has been solved elegantly by Geven et al. 1993 as shown in Fig. 3. It was the developments in white light HPS lamps that finally pointed the way conceptually and led to the successful introduction of metal halide lamps into the market with ceramic arctubes. The construction is similar to the protruded plug leadthrough design employed in white light HPS arctubes. The dose in the arctube body is maintained at a substantially higher temperature by the long separation between the seal and the area occupied by the discharge. Chemical interactions between the seal glass and the metal halide dose are minimized by the relatively low temperatures at the interface, while at low temperatures, the stability of the seal material is assured. It is this invention that has enabled the design of metal halide lamps to make full use of the advantages offered by ceramic arctubes.

However, the leadthrough assembly used in HPS arctubes is not suitable for metal halide lamps as both the niobium component and the sealing glass would react rapidly with the liquid and gaseous phases of the metal halide dose. The niobium component is therefore replaced by a two-part assembly consisting of a molybdenum leadthrough welded to a niobium leadwire. It is essential that the seal glass completely covers the niobium component, the join between the niobium and molybdenum, and part of the molybdenum component; see Fig. 4. The formation of cracks in the PCA end plug due to the difference in coefficients of thermal expansion between the alumina leg and the molybdenum leadthrough is avoided by employing a coiled molybdenum overwind on the leadthrough which absorbs the thermal stresses.

The sealing glass is a high melting temperature mixture of Al_2O_3 , SiO_2 , and Ln_2O_3 (where Ln = lanthanide element, typically Dy; Datta 1977). Figure 5 is a schematic diagram of a 5-part ceramic arctube assembly showing details of the niobium and molybdenum components of the leadthroughs.

An alternative approach is to use a molybdenum-alumina cermet, similar to the material employed by Brown et al. (1982) that has a coefficient of thermal expansion close to that of alumina.

Arctube Seals

A robust and reliable seal is needed between the niobium leadwire and the arctube leg. The properties required for these seals are demanding:

The seals must be hermetic;

Able to cope with many cycles of thermal expansion and contraction;

Be sufficiently robust to survive vibrations while the lamp is in service;

Resistant to the dose components within the arctube at high temperatures over the rated life of the lamp.

Fig. 3 Ceramic discharge vessels after Geven et al. (1993)

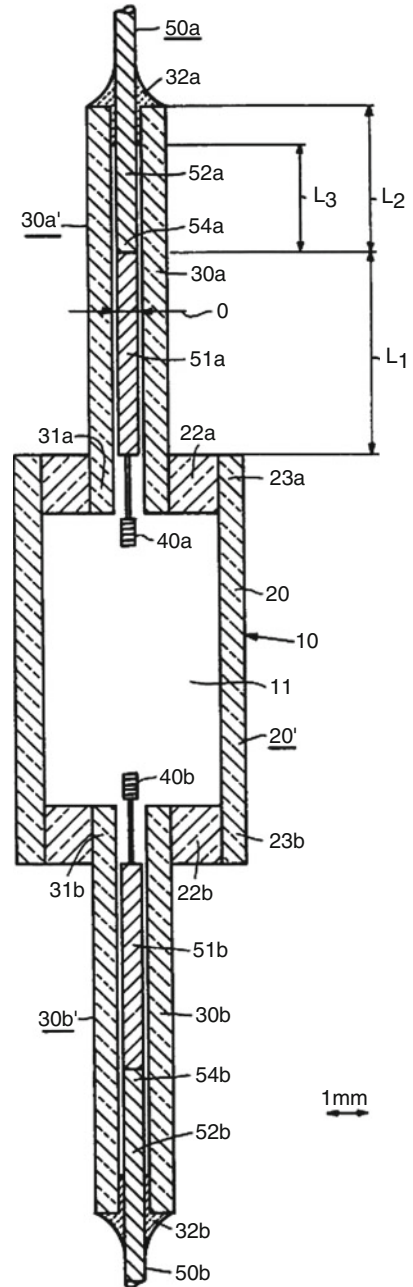


Fig. 4 Image from EP 0 587 238 B1 showing the seal between arctube leg and the Nb + Mo leadthrough assembly

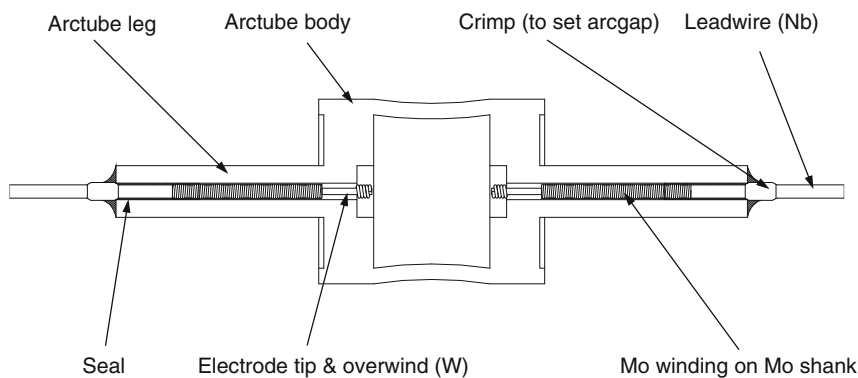
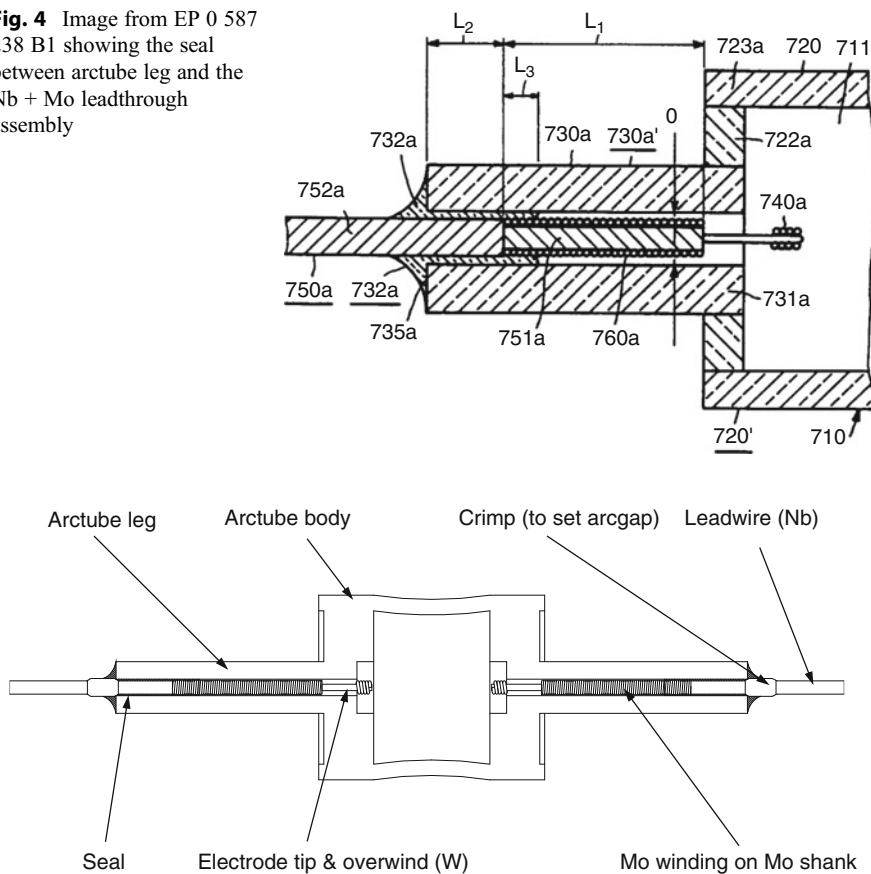


Fig. 5 Ceramic 5-part arctube assembly showing details of Nb and Mo components of the leadthroughs

Additionally, the process for making the seals must be robust and cost-effective, and the cycle time has to be compatible with the required throughput. A process that requires holding the components at high temperatures for long periods would not be suitable in most production environments. Seals for ceramic arctubes are usually formed from a frit ring composed of various metal oxides which form a number of crystalline phases in a glass matrix after processing. The frit ring is accurately located onto the components and the assembly then subjected to a well-defined and precisely controlled heating and cooling cycle. This temperature cycle is accompanied by a pressure cycle which is needed to control the extent of seal glass penetration over the Nb-Mo leadthrough structure. The requirements for the frit rings are specified by mass, composition, size, and shape.

The molten salt chemistry of the seals described above occurs during the process of heating and cooling to obtain the required morphology. For metal halide

lamps with ceramic arctubes, as shown in Fig. 3, the frit rings are based on the $\{\text{Al}_2\text{O}_3 + \text{Dy}_2\text{O}_3 + \text{SiO}_2\}$ system (Datta 1977). During the sealing process, the frit ring is subjected to temperatures above the melting temperature of the mixture and then cooled under controlled conditions to allow the appropriate crystal phases to form. Care has to be taken not only to ensure that the required crystal phases are present but also that the size and distribution of the crystals are correct and the composition of the glassy matrix is within predefined limits (Mucklejohn 1999; DeCarr et al. 2001). The sealing process is further constrained by the need to control the penetration of the glass along the leg and to ensure that residual stresses are minimized. Figure 6 shows a part of the seal glass between the niobium leadwire and the PCA leg. The PCA is represented by the dark area at the top of the picture, while the niobium is shown as the light area along the lower part.

During the sealing process, several chemical reactions occur while the glass is in the molten state; the extent of these reactions is governed by the temperature, time, and diffusion processes. The dissolution of PCA in the seal glass is clearly seen in Fig. 6 by the irregular nature of the inner wall of the leg and the deposition of crystals of dysprosium aluminum garnet, $\text{Dy}_3\text{Al}_5\text{O}_{12}$, adjacent to the PCA boundary. The dendritic crystals are dysprosium orthosilicate, Dy_2SiO_5 . The composition of the glassy matrix, seen above as a mid-gray shade, varies both radially and axially.

The glasses used to form the metal-to-ceramic seals, after a short time in the highly mobile liquid phase, remain in the immobile supercooled-liquid phase throughout the life of the lamp. This is an example of a transient molten salt process – the chemistry takes place during the course of a few minutes but dictates the nature of a structure that has to survive for many thousands of hours at high temperature.

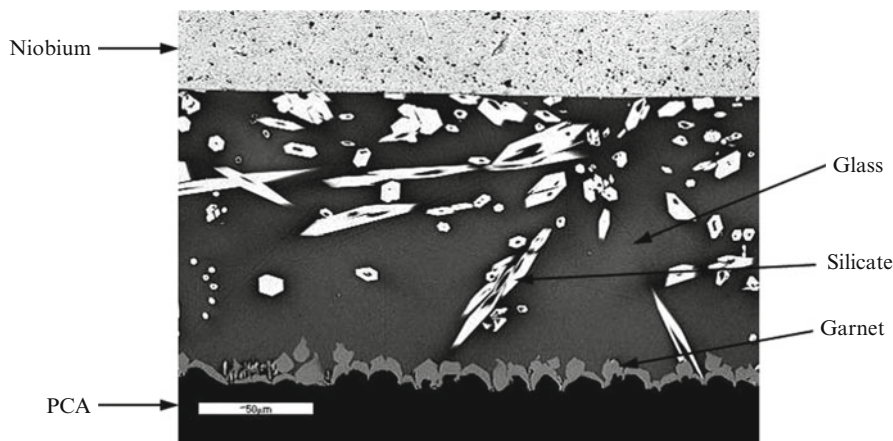


Fig. 6 SEM backscattered image of a PCA to niobium seal

The Evolution of Ceramic Metal Halide Lamps

The evolution of ceramic metal halide lamps from prototypes to highly successful commercial products may be summarized by considering four stages of development.

Generation 1 – early 1980s

Pre-commercialization

Prototypes confirm performance advantages, but corrosion limits survival

Special ballast needed to sustain high lamp voltage

Generation 2 – mid 1990s

Commercial products launched

Corrosion limited by low temperature at seal – metal halide interface

Retrofit to IEC 61167 electromagnetic ballasts plus ignitor

Limited product range, 35–150 W

Good efficacy confirmed

Excellent color control gains widespread recognition

Lumen maintenance identified as opportunity for improvement

Not classified as dimmable

Generation 3 – Late 1990s onwards

Product line well established

Growing range of products

Powers 20–400 W

Lumen maintenance improved

Suitable for electromagnetic and electronic ballasts

Lamps designed for optimum performance on electronic ballasts

Introduction of electronic ballasts with dimming capability

Generation 4 – 2010 onwards

Further improvements to lamp survival

Further improvements to lumen maintenance

Some products designed only for operation on electronic ballasts

Further reductions in color shift at 5,000 h

Some products designed for dimming

Not suitable for hot restrike

Lamp Performance

The advent of metal halide discharge lamps with ceramic arctubes in the mid- to late-1990s led to major changes in the customers' expectations for the performance of metal halide lamps. Many of the limitations of the early generations of metal halide discharge lamps with fused silica arctubes were overcome. At first the range of ceramic metal halide lamps was restricted to 70 and 150 W, but this was extended to a vast array of products from 20 to 400 W over the following 10 years.

Table 1 Color control of a group of 150 W ceramic metal halide lamps (Guest et al. 2008)

Burning time	Color temperature	Chromaticity		Color spread	Color spread	
	CCT/K	<i>x</i>	<i>y</i>	SDCM	CCT/K	SDCM
100 h	3,849	0.382	0.363	3	–	–
9,000 h	3,897	0.381	0.367	3	+48	2

A quantitative measure of lighting quality is that of color control which may be regarded as a combination of color spread and color shift. The former can be taken to be represented by the range of color appearances in a group of light sources after a specified burning time. The latter can be quantified by the difference in mean color appearances for a group of light sources between 100 h and a specified burning time. In this review, color spread is defined as the size of the Standard Deviation of Color Matching (SDCM) ellipse that will contain 90 % of the light sources of a particular group. Similarly, color shifts will be quantified both by the change in color point, again by reference to the size of the SDCM ellipse that contains 90 % of the light sources of a particular group, and by the change in average Correlated Color Temperature (CCT) with respect to the corresponding values at 100 h. Table 1 describes the color control of a set of 150 W ceramic metal halide lamps operated horizontally on electronic ballasts at 100 h and 9,000 h (Guest et al. 2008). This illustrates that metal halide discharge lamps can now match the color control demonstrated for many years by linear fluorescent lamps.

The first metal halide discharge lighting systems available commercially were based on 250 W and 400 W lamps with fused silica arctubes containing mixtures of (sodium iodide + scandium(III) iodide) or (sodium iodide + thallium iodide + indium iodide). It soon became apparent that there was substantial demand for lower powers, i.e., lower lumen output packages, for use in commercial environments. Products with improved color rendering ($R_a(8) > 75$), lamp efficacy > 70 LPW with lamp survival factor 0.50 at 6,000 h followed. However, the limitations of low wattage metal halide lamps with fused silica arctubes did restrict their adoption in some applications where color uniformity was particularly important. These limitations included:

- Poor color uniformity as a consequence of the relatively low maximum operating temperatures;
- Color shift through life due to sodium loss from the arctube;
- Short lives compared to high-pressure sodium lamps;
- Poor lumen maintenance.

Subsequently, improvements in the technology of fused silica arctubes gave rise to increased customer expectations with $R_a(8)$ above 80 considered essential for most applications. Thus, by the mid-1990s, the characteristics of the early ceramic metal halide lamps no longer offered such attractive attributes to the end user.

Brown et al. (1982) employed a mixed halide dose system containing sodium and tin together with chlorine and iodine; the next-generation metal halide lamps

with ceramic arctubes contained iodides of sodium, thallium, and lanthanide elements. Both dose systems generate continuum radiation which is responsible for the good color rendering properties ($R_a(8) >70$ and >80 , respectively) of the lamps.

These dose compositions give rise to different radiation generating species. The {Na + Sn + Cl + I} system promotes atomic emission lines from Hg, Na, and Sn with molecular bands from HgCl and HgI at 445 and 555 nm, respectively. Continuum radiation is thought to arise from the electronic transitions of the tin monohalides, SnCl and SnI. In the {Na + Tl + Ln + I} dose systems (Ln = lanthanide), discrete atomic emission lines arise from Hg, Na, and Tl. Many of these lines are pressure broadened supplementing the continuum. Pseudo-continuum radiation arises from the atomic emissions of the lanthanide atoms which provide an enormous number of overlapping lines in the visible region of the spectrum. These emissions are supplemented by molecular contributions from lanthanide monoiodides, LnI.

The progress in lamp performance from the 1980s to the first commercially available ceramic metal halide lamps in the 1990s and to the beginning of the second decade of the twenty-first century can be appreciated from the data listed in Table 2 which compares the properties of some typical 150 W lamps. Although the report by Brown et al. (1982) does not contain a comprehensive list of performance parameters, it does provide sufficient information to show that although the lamps of the 1980s and the 1990s have similar luminous efficacies at 100 h, the latter show immense advances in color uniformity, color stability, and color rendition properties. The lamps with ceramic arctubes show substantial improvements in performance over all metal halide lamps with fused silica bodies. The latter typically not only have a larger color spread at 100 h but also show a color shift of >500 K at 2,000 h. The former have efficacies approximately 20 % higher than the fused silica counterparts with lives at least 50 % longer.

The second generation of ceramic metal halide lamps, i.e., the first products to be made available commercially, had arctubes with cylindrical geometry with a 5-part construction (1 \times arctube body, 2 \times end plugs, 2 \times legs to house the leadwires and seals). This design was, in part, due to the need to use ceramic components fabricated by extrusion and/or pressing. Development of injection molding technology, such as described by Venkataramani et al. (1999), allowed the introduction of complex end chamber geometries thus enabling the ceramic component count to be reduced to 3 by combining the end plug and leg into one part. Further advances in ceramic processing gave rise to precisely shaped arctubes with one or two parts. These arctubes are specifically designed to provide an isothermal environment and to minimize the transport of arctube material by restricting the movement of the metal halide dose during lamp operation. Figure 7 shows ceramic metal halide lamps with 5-, 3-, and 2-part arctubes.

These changes to the design and construction of ceramic arctubes have extended lamp lives from 6,000 h to 15,000 h (to 50 % survival). Table 2 summarizes the evolution of CMH lamps.

Table 2 Summary of 150 W ceramic metal halide lamp performance

	Generation 1 ^a	Generation 2 ^b	Generation 3 ^c	Generation 4 ^d
Output/lm	13,500	14,000	14,000	15,000
Lamp power/ W	150	147	150	149
Lamp efficacy/ lm W ⁻¹	90.0	95.2	93.3	100.6
LLMF ^e				
at 2,000 h	0.90	0.75	0.75	0.91
at 5,000 h		0.70	0.70	0.84
LSF ^f				
to 50 % survival	4,000 h	6,000 h	12,000 h	15,000 h
to 90 % survival			10,000 h	10,000 h
Color temperature, CCT/K	3,800	3,000	3,000	3,000
x chromaticity	0.390	0.435	0.436	0.443
y chromaticity	0.400	0.400	0.396	0.391
Color uniformity	+/-150 K	+/-50 K	+/-50 K	<50 K
Color shift at 5,000 h	~300 K	<100 K	<100 K	<<100 K
R _a (8)	70	80	88	92
Control gear	Special ballast to sustain high lamp voltage	Electromagnetic ballasts to IEC 61167 with superimposed ignitor Approved electronic ballasts		Approved electronic ballasts only

^aLamp described in Brown et al. (1982)

^bTypical 150 W ceramic metal halide lamp from mid-1990s operated on an electromagnetic ballast

^cTypical 150 W ceramic metal halide lamp from 2010 operated on an electronic ballast

^dPhilips 150 W CDM-T Elite 930 operated on an electronic ballast

^eLamp lumen maintenance factor

^fLamp survival factor

CMH lamps now available equal the color control of fluorescent lamps, and the most recent developments suggest further improvements will be made in color spread, color shift, and lumen maintenance.

Readers should refer to manufacturers' product datasheets for lamp properties such as spectral power distribution, including UV emissions; polar light intensity curves; dimensions; warm-up characteristics; supply voltage sensitivity; allowed burning orientations; dimming characteristics; flicker; restrike times; fusing recommendations; compatible electromagnetic ballasts and ignitors; compatible electronic ballasts; and requirements for luminaires.



Fig. 7 Images of lamps with 5-, 3-, and 2-part arctubes. Images adjusted to have same G12 pin separation

Dimming

One of the limitations of ceramic metal halide lamps is that most of the product range is not rated as dimmable by the manufacturers. This can restrict the use of these lamps in installations with control systems that rely on dimming as part of an energy saving strategy. Although ceramic metal halide lamps can be operated at reduced powers for long periods without detrimental impacts on lumen maintenance and lamp survival (Guest et al. 2008), there is a significant change in color temperature as the lamp power is lowered. This color shift while being potentially beneficial in some applications, such as road lighting where the move to higher color temperatures can improve visual performance under mesopic conditions, is very undesirable in other applications.

Keijser (1998) showed that the changes in color temperature, luminous flux, and color rendering index with power reduction are considerably smaller for metal halide lamps with ceramic arctubes compared to similar lamps with fused silica arctubes.

A recent study by Li et al. (2013) describes the variations of efficacies, color temperatures, and color rendering indices with power for 70 and 150 W ceramic metal halide lamps. These studies are complemented by diagnostic measurements recording the changes in arctube wall temperatures as the input is varied.

End of Life Mechanisms

The dominant end of life failure mechanism for metal halide lamps with ceramic arctubes is corrosion of the wall leading to pinhole formation and/or cracks which allow leakage of the fill gas and dose into the outer jacket. This corrosion is brought about by the transport of arctube material by the metal halide species in the dose. Arctube leakage is manifested by a sudden and significant decrease in lumen output and a large change in color appearance. Lamps should be replaced as soon as this end of life mechanism is evident.

Leakage into the outer jacket of the lamp often promotes a discharge between, for example, parts of the arctube mounting frame. This discharge will give rise to asymmetry in the current drawn by the lamp, i.e., rectification. This rectification can lead to overheating of the ballast, and therefore, only electromagnetic ballasts conforming to IEC 61167 with thermal protection devices should be used with these lamps. Similarly, electronic ballasts used to operate ceramic metal halide lamps should have thermal protection and timed ignitor switch off.

Over the course of lamp's operating life, the lamp voltage will rise, mainly as a result of increased arc gap due to the gradual erosion of the electrode tips. If the lamp voltage rises to a value that is beyond the range that the ballast can supply, the lamp will extinguish. However, on cooling, it will restart when the required ignition voltage falls to the actual pulse voltage applied by the ignitor. Following warm-up, the lamp voltage will rise again causing extinction. This is known as end of life cycling and is a well-recognized phenomenon in high-pressure sodium lamps. Lamps should be replaced as soon as this end of life mechanism is evident.

The Next Generation of Ceramic Metal Halide Lamps

The most recent development in ceramic metal halide lamp technology is the move away from having a condensed mixture of metal halides present during lamp operation (i.e., saturated vapor phase conditions) to having all of the metal halides in the vapor phase during lamp operation (i.e., unsaturated vapor phase conditions). These developments are outlined in by Hendricx et al. (2010) and Rijke et al. (2012). All of the mercury is in the vapor phase during lamp operation. The unsaturated conditions while promoting further improvements to color spread, color shift, lumen maintenance, dimming characteristics, and lamp survival do place added demands on materials and manufacturing techniques.

Summary

Figures 8, 9, and 10 show typical installations for ceramic metal halide lamps reflecting their widespread use in commercial interiors, streetlighting, and city beautification.

The well-recognized advantages of ceramic arctubes for metal halide lamps have been successfully translated into a well-established and highly reliable light source technology. The improvements in performance compared with lamps with fused silica arctubes have been made possible by a combination of advances in materials science and processing, better understanding of discharge physics and lamp chemistry, and improved manufacturing techniques giving high precision and accuracy in fabricating and processing arctubes. The advent of reliable, low-cost electronic ballasts for high-intensity discharge lamps has given rise to products with color control matching that of linear fluorescent technology and lamps that are specifically designed to operate for prolonged periods at reduced powers. It has to be noted that not all of the benefits shown by ceramic metal halide lamps in the 20–150 W range have been translated into the higher power range, 250–400 W.

Fig. 8 Ceramic metal halide lamps in a retail installation. Image from Philips Lighting





Fig. 9 Ceramic metal halide lamps in streetlighting. Image from GE Lighting



Fig. 10 Ceramic metal halide lamps used for city beautification. Image from Osram

References

- Brown KE, Chalmers AG, Wharmby DO (1982) Tin sodium halide lamps in ceramic envelopes. *J Illumin Eng Soc* 11(2):106–114
- Carleton S, Seinen PA, Stoffels J (1997) Metal halide lamps with ceramic envelopes: A breakthrough in color control. *J Illumin Eng Soc* 26(1):139–145

- Datta RK (1997) Sealing materials for ceramic envelopes. United States patent 4076991 (Date filed: 6 May 1977)
- DeCarr SM, Grande JC, Gyor M, Laher I, Lovett DJ, Meszaros J, Mucklejohn SA, Toth Z (2001) Characterising metal-to-ceramic seals for high intensity discharge lamps. In: Proceedings of the 9th international symposium on science and technology light sources (LS9), Ithaca, pp 141–142. ISBN 0-9713422-0-2
- Geven ASG, Renardus MLP, Seinen PA, Stoffels JAJ, Wijenberg C, Diells HR (1993) High-pressure discharge lamp. European patent 0587238 (Date filed: 1 Sep 1993)
- Guest EC, Girach MH, Mucklejohn SA, Rast U (2008) Effects of dimming 150 W ceramic metal halide lamps on efficacy, reliability and lifetime. *Lighting Res Tech* 40:333–346
- Hendrixx J, Vrugt J, Denissen C, Suijker J (2010) Unsaturated ceramic metal halide lamps - A new generation of HID lamps. In: Haverlag M, Kroesen GMW, Taguchi T (eds) Proceedings of the 12th international symposium on science and technology light sources & 3rd international conference on white LEDs & solid state lighting, 11–16 July 2010, Eindhoven, pp 405–414. ISBN 9780955544521
- Keijser RAJ (1998) Dimming of ceramic metal halide lamps. In: Proceedings of the 8th international symposium on science and technology light sources (LS8), Greifswald, pp 226–227. ISBN 3-00-003105-7
- Li W, Shi SJ, Liu J, Lister GG, Zhang SD (2013) Dimming properties of medium and small wattage ceramic metal halide lamps. *Lighting Res Tech* 45:1477153513490589. doi:10.1177/1477153513490589
- Mucklejohn SA (1998) The evolution of low wattage metal halide lamps with ceramic arctubes. In: Proceedings of the 8th international symposium on science & technology light sources (LS8), Greifswald, pp 230–231. ISBN 3-00-003105-7
- Mucklejohn SA (1999) Molten salts in high-intensity discharge lamps. In: Proceedings of the international George Papatheodorou symposium, Patras, 17–18 Sept 1999, pp 63–67. ISBN 960-7839-01-3
- Mucklejohn SA (2012) Recent advances in lighting quality and energy efficiency with traditional light source technology. In: Proceedings of CIE 2012 – lighting quality and energy efficiency, 19–21 Sept 2012, Hangzhou, pp 38–48. ISBN 978-3-902842-42-8
- Mucklejohn SA, Preston B (2000) Low wattage metal halide discharge lamps with ceramic arctubes 1980 to 2000. Industry applications conference. Conference record of the 2000 IEEE, vol 5, pp 3326–3329. doi:10.1109/IAS.2000.882643
- Rijke AJ, Lemmens TL, Janessen JFJ, Nijdam S, Haverlag M, van der Mullen JJAM, Kroesen GMW (2012) Quantitative assessment of the energy balances of three generations of ceramic high intensity discharge lamps. In: Devonshire R, Zissis G (eds) Proceedings of the 13th international symposium on science & technology lighting, 24–29 June 2012, Troy, pp 183–184. ISBN 9780955544545
- Seinen PA (1995) High intensity discharge lamps with ceramic envelopes. In: Itatani R, Kamiya S (eds) Proceedings of the 7th international symposium on science & technology light sources (LS7), Kyoto, pp 101–109
- Venkataramani VS, Grescovich CD, Scott CE, Brewer JA, Ning C (1999) Ceramic discharge chamber for a discharge lamp. European patent 0954011 (Date filed: 8 Mar 1999)

Further Reading

Lamptech – The museum of electric lamp technology: <http://www.lamptech.co.uk/MBI%20Ceramic.htm>

Most major lightsource manufacturers and lighting specialists offer a range of informative articles about CMH lamps and their uses together with examples of installations on their websites

Electrodeless Lamps and UV Sources

Graeme Lister

Contents

Introduction	1142
Electrodeless Radio-Frequency (RF) Lamps	1143
Electromagnetic Interference (EMI) and Safety	1143
Power Supplies	1145
Classification of Electrodeless RF Lamps	1146
Discharge Properties	1146
Electrodeless Fluorescent Lamps	1148
Development of Electrodeless Fluorescent Lamps	1148
Benefits of Electrodeless Fluorescent Lamps	1148
The Physics of Electrodeless Fluorescent Lamps	1149
Reentrant Cavity Lamps	1150
Lamps with Outer Coils	1151
Toroidal Lamps	1151
Electrodeless High-Frequency HID Lamps	1153
Benefits of Electrodeless HID Lamps	1153
Microwave and High-Frequency Resonant Cavity (“Plasma”) Lamps	1154
Dielectric Barrier Discharge (DBD) Lamps	1160
Principles of DBD	1160
The Physics of DBD	1161
DBD Planar Light Sources	1163
Electrodeless Excilamps	1164
UV Light Sources	1164
The Physics of Excilamps	1166
DBD Excilamps	1167
Capacitively Coupled Excilamps	1168
Microwave Discharge Excilamps	1169
References	1169

G. Lister (✉)

Graeme Lister Consulting LLC, Wenham, MA, USA

e-mail: graeme_lister@graemelisterconsulting.com; graeme.lister@yahoo.com

Abstract

“Electrodeless” discharge lamps have several advantages compared to lamps containing electrodes, including longer life, improved lumen maintenance, lamp efficacy, flexibility of lamp design, the use of chemical doses that would interact with electrodes, and improved electrical control. Although the concept of electrode lamps was first demonstrated by Tesla in 1891, the first commercially viable electrodeless fluorescent lamps did not appear in the market until the early 1990s and electrodeless *HID* lamps a decade later. Prior to that time, the development of electrodeless lamps had been hindered by the lack of inexpensive and efficient electronics. Electrodeless lamps are also finding increasing use as UV sources, particularly so-called excilamps, for degradation of organic pollutants and microbe deactivation.

Introduction

As has been noted in previous chapters, conventional discharge lamps contain electrodes, which supply the electric current and electric field required to maintain the discharge. Since the 1990s, “electrodeless” light sources have been introduced in the market; these lamps operate either without electrodes, or with electrodes that are placed outside the discharge vessel. These lamps may be classified into two distinct categories:

Electrodeless Radio-Frequency (RF) Lamps: These lamps operate due to the presence of a radio-frequency electromagnetic field, supplied either from an external power source or by applying a radio-frequency voltage to electrodes placed outside the discharge vessel.

Dielectric Barrier Discharge (DBD) Lamps: The electrodes in *DBD* are placed outside the discharge. *DBD* is characterized by insulating layers on one or both electrodes, placed on dielectric structures inside the discharge gap.

There are a number of potential benefits to lighting from electrodeless operation of lamps (Wharmby 1993, 1997):

Lamp Life: Electrode failure is the major limitation on the operating life for both fluorescent and *HID* lamps. Further, electrodes are connected to the power supplies through glass-to-metal seals, which weaken the lamp structure and reduce lamp life.

Lumen Maintenance: Sputtering and evaporation of electrode emissive and substrate material result in dark regions on the lamp walls, which hinder the transmission of light.

Lamp Efficacy: Energy dissipated on, or in the region of, electrodes represents a net loss of lamp efficacy. However, the gain in lamp efficacy by electrodeless operation may be offset by energy losses in the power electronics used to run the lamp.

Lamp Design: The presence of electrodes places restrictions on discharge lamps in both size and shape, some of which can be overcome by electrodeless operation.

Chemical Dose: The chemicals that are introduced into a lamp to produce radiation are limited to those that attack neither the electrodes nor the envelope material. Electrodeless operation permits the use of chemicals with favorable radiative properties that could not be used in conventional lamps.

Electronic Control: The ability to dim discharge lamps to conserve electrical power when full light output is not required is enhanced by electrodeless operation. Starting and restarting times of lamps may also be reduced.

The following sections will discuss the operation of *RF* fluorescent and *HID* lamps, as well as *DBD* lamps, with emphasis on the underlying technology and product availability and development. Electrodeless *UV*-light sources are finding increasing commercial use, particularly through the development of “excilamps,” that produce *UV* radiation from rare-gas dimers, halogen dimers, and rare-gas halogen excimers, and a review of the current status of the technology of these sources is included in the final section of this chapter.

Electrodeless Radio-Frequency (RF) Lamps

The first experiments demonstrating that light could be produced using high-frequency electromagnetic fields were conducted by Tesla in 1891, but the commercial development of these lamps was hindered for almost a century by the size, cost, and reliability of the driving electronics. However, progress in semiconductor electronics and power switching technology in the late 1980s led to a number of new lamp products, firstly in fluorescent lighting and more recently for *HID* applications.

The benefits of electrodeless operation of discharge lamps have been briefly outlined above and will be discussed in some detail in the sections below. The following sections provide an overview of the important factors determining the technology and operation of electrodeless *RF* lamps.

Electromagnetic Interference (EMI) and Safety

Electrodeless lamps are required both to be safe and to avoid interference with radio communications (Wharmby 1997). Regulations for the control of *EMI* from

lighting devices are set by the International Electrotechnical Commission (*IEC*), and the main requirements for compatibility with these regulations are discussed below.

Radiated Electromagnetic (EM) Disturbance

Radiated electromagnetic fields from lamps are generated by plasma, coils, and circuits. In principle, these fields can be screened by a conducting enclosure, but shielding of the lamp to allowable levels while maintaining acceptable light means, in practice, that operating frequencies are limited to the industrial, scientific, and medical (*ISM*) bands allocated for noncommunication use:

- (i) 2.2–3.0 MHz: The *IEC* has allowed for relaxed standards relative to previous standards for this band, which lies between *MW* and *SW* radio bands. The bandwidth is sufficiently large to permit the use of simple low-cost self-oscillating circuits, and a number of electrodeless fluorescent products operate in this band.
- (ii) 13.56, 27.12, and 40.68 MHz: Lamps operated at the industrial frequency of 13.56 MHz have the advantage that the second and third harmonics are also in the allowed bands. However, the permitted bandwidth is extremely small (approximately 50 Hz), and the lamps operating at this frequency require expensive crystal control to achieve *EMI* compatibility.
- (iii) 433.075–434.775 MHz: This band is available only in International Telecommunication Union (*ITU*) region 1, which includes Europe, Africa, parts of the Middle East, and the former Soviet Union. This band is used for some electrodeless *HID* products, due to the ready availability of suitable solid-state electronics, but when used outside *ITU* region 1, these products require complete *EMI* shielding.
- (iv) 915 MHz: This frequency is available for lighting applications in the *USA*, but has not been used in a commercially available light source.
- (v) The 2.4–2.5 GHz band: This band includes the frequency used by microwave ovens, and the development of relatively low-cost magnetrons for this purpose has assisted in the development of practical microwave light sources, which will be discussed in section “[Microwave and High-Frequency Resonant Cavity \(“Plasma”\) Lamps](#).”

A further non-*ISM* frequency band, 100–300 kHz, is currently used for the electronic ballasts of electrode fluorescent and compact fluorescent (*CFL*) lamps and is also used for closed core (toroidal type) electrodeless induction fluorescent lamps (cf. section “[Toroidal Lamps](#)”).

Conducted Interference or Terminal Disturbance Voltage (TDV)

Conducted interference results from the presence of high-frequency currents in the mains supply due to the *RF* potential of the lamp relative to ground and *RF* energy feedback to the *AC* mains from the ballast. Currents which are induced in the lamp circuit itself (differential mode noise) are easily reduced by

including a blocking filter on the mains side of the power supply. In induction lamps, stray currents are induced by the coupling of the *RF* coils to the ground plane (common mode noise). This mode can be screened using a transparent conductive coating such as indium tin oxide (*ITO*) on the bulb (for reentrant cavity lamps) or operating the lamps in a grounded screen fitting or using a bifilar winding.

Safety

There are currently no international regulations covering safety to exposure of electromagnetic radiation, although a European standard detailing an assessment of lighting equipment related to human exposure to electromagnetic fields was recently published. Guidelines issued by the International Non-ionizing Radiation Committee recommend that the level should not exceed 1 W/m^2 in the frequency range 10–400 MHz for uncontrolled areas of public access, with higher limits for lower frequencies, as the wavelength becomes greater than the dimensions of the human body. All existing electrodeless lamps fall well within these limits. In the USA, some local regions, e.g., California, impose standards which are more stringent than the federal regulations. Devices intended for use in these localities must comply with all applicable local regulations.

Power Supplies

The power supply, often referred to as the *RF* generator, or the electronic control gear (*ECG*) that operates an electrodeless discharge lamp performs the dual functions of igniting the lamp and maintaining the system at the correct power level, analogous to the ballast in an electrode lamp, and is designed to operate at a nominal input frequency of 50–60 Hz alternating current (*AC*). The term *ECG* is in common usage in Europe, whereas the term *RF* generator is preferred in the USA.

The *RF* generator or *ECG* consists of two components: a device to convert *AC* power from the mains to *DC* and a device to convert *DC* to *RF* power. The device may also include a function to dim the lamp, or a function to continuously monitor the lamp and switch the lamp off if an operating failure is detected. External messages can also be addressed by accessing the serial port and using DALI-type or other proprietary messaging protocols, enabling individual lamps or sets of lamps to be controlled from a central operating system.

The overall efficiency of a lamp power source is defined as the fraction of power delivered by the generator at the source input that is absorbed by the plasma (Zakrewski et al. 1992). In addition to reflection at the line source interface, other possible loss mechanisms that can divert *RF* power from the discharge include losses in the dielectric and metal components of the plasma source and *RF* radiation leaking into the surroundings. However, this power loss can be made negligibly small by a suitable power source design. Reflected power at the line source must be minimized, both to maximize the electrical efficiency of the system and to prevent unwanted power being dissipated in the power source itself, limiting its life. This can be

accomplished by using an appropriate matching network of the field applicator. In general, two tuning methods are required: one to remove the imaginary part and the other to bring the real part close to the transmission line characteristic impedance. The lamp should be designed such that its steady-state impedance is close to that of the microwave transmission line or driving source impedance.

Classification of Electrodeless RF Lamps

In principal here are four distinct types of excitation possible for electrodeless *RF* lamps (Wharmby 1993, 1997; Lister 1999) but only two of these methods (inductive and microwave discharges) have been successfully applied to commercial light sources, for reasons which are discussed below.

Inductive or H Discharges: The plasma in an *H* discharge forms a single turn secondary of a transformer, excited with a primary coil which may be placed in or around the discharge. Provided sufficient power is applied to maintain the *H* discharge, efficient coupling may be achieved at low frequencies, with reduced electromagnetic interference (*EMI*) and less expensive electronics.

Capacitive or E Discharges: The simplest *E* discharge consists of a gas-filled vessel placed between the plates of a capacitor. Capacitive *RF* discharges are electrodeless if the *RF* electrodes are placed outside the discharge vessel. The applicator couples to the discharge through the sheaths next to the electrodes, resulting in a strong dependence of discharge characteristics on frequency. In contrast to *H* discharges, higher power densities are achieved only at higher frequencies, where cost of electronics and electromagnetic interference (*EMI*) issues are more serious. Efficient capacitively coupled *HID* lamps have been developed, but to date there is no commercial application (Lapatovich 2012).

Resonant Cavity Discharges: These discharges operate in the frequency range of 400 MHz and above, where the wavelength of the electromagnetic field is comparable to the dimensions of the exciting structure and vessel. These lamps are particularly suitable for operation at high electrical power (Waymouth 1993).

Traveling Wave Discharges: In a *traveling wave discharge* (such as a surface wave, Moisan and Zakrewski 1992), the plasma is created by an electromagnetic wave traveling along a slow-wave structure formed by the plasma column itself. There was considerable interest in the potential of surface waves for application to fluorescent lighting in the late 1980s but these suffer from some of the problems of capacitive discharges (Lister et al 2004; Doell and Lapatovich 1990).

Discharge Properties

All high-frequency discharges currently used for lighting are “overdense” (Waymouth 1993), i.e., the applied frequency ω is smaller than the plasma frequency ω_{pe} , where

$$\omega_{pe} = 56.4n_e^{1/2} \text{ s}^{-1} \quad (1)$$

and $n_e \text{ m}^{-3}$ is the electron density. In fluorescent lamps, typical electron densities are 10^{17} – 10^{18} m^{-3} , with $\omega_{pe} \geq 2 \times 10^{10} \text{ s}^{-1}$ ($\sim 3 \text{ GHz}$), while in *HID* lamps, electron densities are 10^{21} – 10^{22} m^{-3} with $\omega_{pe} \geq 2 \times 10^{12} \text{ s}^{-1}$ ($\sim 300 \text{ GHz}$).

In plasmas excited by *RF* electric fields, the “effective field,” E_{eff} , is the field that provides the same power input to the plasma as would a *DC* field, E , i.e.,

$$P_{RF} = \bar{\sigma}_e E_{eff}^2 \quad (2)$$

where $\bar{\sigma}_e$ is the *RF* electron conductivity. Following Brown (1959) we then have

$$E_{eff}^2 = E^2 \frac{\nu_m^2}{\nu_m^2 + \omega^2} \quad (3)$$

where $\nu_m \text{ s}^{-1}$ is the electron momentum transfer collision frequency (Note that Eq. 3 is an approximation and does not account for the variation of the electron frequency with electron energy (Lister et al. 1996). However, this approximation remains valid in the limit.). For fluorescent lamps, *DC* values of $\nu_m = \nu_{dc}$ are typically 10^8 – 10^9 s^{-1} compared to 10^{12} – 10^{13} s^{-1} in *HID* lamps. Since in all cases $\nu_m^2 \gg \omega^2$, from Eq. 3, $E_{eff}^2 \approx E^2$, and the particle kinetics of these discharges can be adequately described by a steady-state *DC* model (Waymouth 1993).

In contrast, the electrodynamics must take into account the inhomogeneity of the electromagnetic field, exemplified by the “skin effect” (Weibel 1967). The collisional skin depth, δ_s , is given by

$$\delta_s = \frac{c}{\omega_{pe}} \left(\frac{2\nu_m}{\omega} \right)^{1/2}. \quad (4)$$

In a typical fluorescent lamp discharge at 3 Torr and operating at 2.65 MHz, $\delta_s \sim 10 \text{ cm}$, while for a microwave-excited mercury discharge at 8 atmospheres operating at 2.45 GHz, $\delta_s \sim 2 \text{ mm}$. In all discharges, there is a region near the wall of low electron density, and therefore low conductivity, through which the field will penetrate before being attenuated by the plasma.

In a real plasma discharge, the concepts “overdense” and “skin depth” in the bulk plasma are more complex. Many of the electrodeless discharges have dimensions which are small compared to the excitation wavelength, and the boundaries often comprise the major extent of the discharge. In these regions, especially in *EHID* lamps, the temperature profile varies rapidly with radial position, often as much as 1000 K/mm. The plasma frequency in a Saha equilibrium plasma is a strong function of temperature, and so conditions for absorption and reflection of incident electromagnetic power can vary strongly over short distances. The relation between skin depth and discharge dimensions for different electrodeless lamps will be discussed in the appropriate sections.

Electrodeless Fluorescent Lamps

Development of Electrodeless Fluorescent Lamps

Inductively coupled discharges (*ICDs*) are currently the only form of electrodeless fluorescent lamp to be exploited commercially. In common with *ICD* for other industrial applications, the discharge starts as a capacitively coupled *E* discharge, until sufficient poloidal electric field is present to create the high-density *H* discharge required to provide adequate light output.

The first commercially available electrodeless *FL* were based on the reentrant cavity approach of Bethenod and Claude (1936). The geometry of these lamps resembles that of a regular incandescent bulb, and the electromagnetic field is supplied from an inductive coil contained within the lamp (section “[Reentrant Cavity Lamps](#)”), analogous to the filament in an inductive lamp. These lamps were introduced into the market in the early 1990s. A later development made use of the “tokamak” principle, in which the electromagnetic field is induced in a toroidal-shaped lamp (section “[Toroidal Lamps](#)”) via a coil wrapped around a ferrite ring outside the lamp. In recent years, toroidal induction lamps have found particular applications for outdoor lighting, where high output lighting is required in places where the cost of lamp replacement is high. Reentrant cavity lamps have to date achieved less penetration in the market place.

An overview of the potential benefits of electrodeless lamp operation was given in section “[Introduction](#).” The particular benefits that have been obtained in commercial electrodeless fluorescent lamps are discussed in the following section. It should be noted, however, that the increase in efficacy of the light source due to the absence of electrodes in these lamps is generally more than offset by the reduction in efficiency of the power electronics used to run the lamp. As a result, in common with *LED* and *HID* electrodeless discharge lamps (section “[Electrodeless High-Frequency HID Lamps](#)”), the energy that is not coupled to the discharge is dissipated as heat in the immediate region of the lamp, and extensive thermal management is often required.

Benefits of Electrodeless Fluorescent Lamps

CRI

Since these lamps are based on fluorescent lamp technology, a suitable choice of phosphor can provide *CRI* in excess of 80, and products are available with a variety of color temperatures.

Lamp Life

Evaporation of emitter material from fluorescent lamp electrodes during operation (particularly starting) causes unsightly darkening of the end of the lamp, accelerated lumen depreciation, and eventually results in lamp failure. Arc attachment at “hot spots” along the filament and movement thereof can cause visible flicker near

the ends of the lamp. On the other hand, recent develops in electrode technology for conventional fluorescent lamps have resulted in products with lifetimes of up to 100,000 hours.

Lamp Design

Fluorescent lamps are long in order to provide sufficient illuminance for general applications and to reduce the fraction of energy loss to the electrodes. Compact fluorescent lamps (*CFL*) are made by folding the discharge tube, with electronics integral to the base of the lamp which enable them to fit into a regular Edison-type screw base or bayonet incandescent lamp socket. Inductively coupled fluorescent lamps have been developed with shapes resembling the incandescent lamp, with improved aesthetic appeal.

Chemical Dose

Fluorescent lamps contain mercury because Hg has a relatively high vapor pressure at room temperature and a high efficiency of converting electrical power to *UV* radiation. However, conversion of a *UV* photon to a single visible photon at the maximum luminous efficacy of 555 nm reduces the usable radiation by 50 % or more. Longer-wavelength *UV* or direct white-light emitters could bring enormous benefits. Many potential candidates are highly reactive with electrode materials and have very sensitive vapor pressure characteristics. Direct white-light emitters are better suited for use in induction lamps.

Electronic Control

The starting phase in electrodeless *FL* lasts only a few milliseconds and is therefore effectively instant for the user (Wharmby 1993, 1997). Unlike either incandescent or conventional fluorescent lamps, however, they perform well under rapid switching. Many induction lamps can be dimmed to 40 % of electrical power, either by step or continuous dimming. When dimmed, induction lamps do not have the same extreme color shift issues as *HID* lamps. The integral electronics package is also well suited to introducing additional electronic controls beyond dimming. Motion or occupancy sensors can be incorporated within the electronics.

The Physics of Electrodeless Fluorescent Lamps

The underlying physics in electrodeless *FL* is similar to that of electrode *FL*. However, absence of electrodes enables fluorescent lamps to operate under conditions which would be impractical in conventional fluorescent lamps:

- (i) Electrodeless lamps often operate at much higher current and electrical power densities than would be possible in conventional fluorescent lamps.
- (ii) In conventional lamps, the buffer gas protects the electrodes. Electrodeless lamps can operate at lower gas pressures, giving better efficacy, particularly at high power loading.

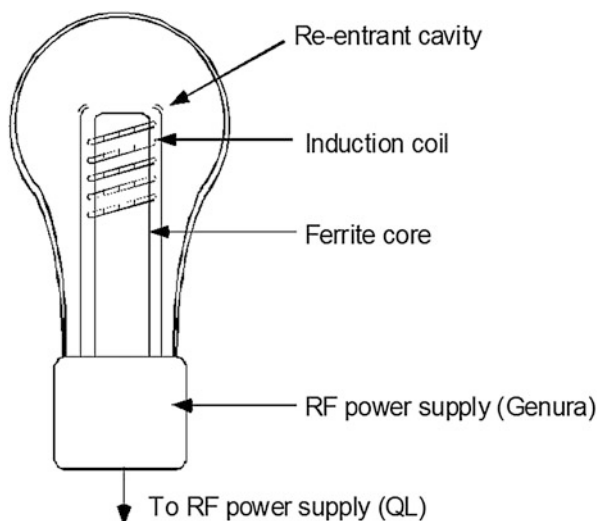
At optimum operating pressure, the sum of all nonradiative losses is a minimum. For low values of discharge current ($<1\text{A}$) in argon, the maximum efficacy for a fluorescent lamp is obtained for a gas pressure between 50 and 150 Pa. However, buffer gas pressure in conventional lamps must be greater than 300 Pa (2.5 Torr) to ensure acceptable electrode life. As discharge current increases, both elastic losses and losses to the wall (mainly ion flux) increase, requiring higher buffer gas pressures, but optimum pressure remains below that of electrode lamps.

Reentrant Cavity Lamps

The concept of the reentrant cavity lamp was introduced by Bethenod and Claude (1936) and is illustrated schematically in Fig. 1. An insulated coil is wound several turns around a ferrite core and placed within the reentrant cavity, inside a lamp similar in shape to a regular incandescent bulb. The application of an *RF* current activates the discharge. The high permeability of the ferrite core increases the magnetic flux surrounding the discharge and leads to more efficient coupling, but power losses in the ferrite increase at high temperatures and electric fields, although these are usually less than losses in a coil without a ferrite core. Phosphor is coated on the interior of the bulb section, and an additional phosphor coating on the reentrant cavity can be used to enhance the production of visible light from *UV* produced inside the lamp. These lamps typically operate at 2.65 MHz, with a gas pressure of 50–150 Pa argon or krypton.

The *QL*[®] lamp (Netten and Verheij 1991) was the first commercially available electrodeless *FL* and was introduced into the market in 1992. Heating of the ferrite is reduced by a copper rod connected to the lamp base, and the electronics to ignite

Fig. 1 Schematic of a reentrant cavity lamp (*QL*[®] and Genura[®])



and maintain the discharge are connected to the lamp via a cable. System lifetime is up to 100,000 h, limited by the life of the electronics. Ignition and reignition times are 0.5 s, while 70 % of maximum light output is reached in 10 min. The original *QL* lamp operated with a system power of 55 W and was somewhat larger than an incandescent bulb (8.4 cm in diameter and 15 cm tall). Currently, there are a number of products in the market using this technology, for a system power of 35–165 W and a range of *CCT* from 2700 to 5000 K. Operation at high system power is particularly attractive for applications where high illuminance is required for use in inaccessible areas, such as high-bay and roadway lighting.

Genura[®] (Wharmby 1997) was the first compact electrodeless fluorescent lamp with integrated electronics. It has the appearance of an incandescent reflector lamp, with a system power of 23 W, an efficacy of 21 lpw, and 15,000 h life. *Genura* has an electrically transparent *ITO* coating to reduce *EMI* and a further titania reflector coating on the reentrant and neck of the bulb.

As noted above, a ferrite core is used to enhance the magnetic flux in an induction lamp. However, ferrites are expensive and fragile and their properties degrade with temperature. Lamps without ferrite have been examined, using air core coils for excitation. These open core exciters work best at frequencies above 2.65 MHz, such as 13.65 MHz, where ferrites do not perform well. At the time of writing there are no commercial induction lamps in the market containing this technology.

Lamps with Outer Coils

Everlight[®] (Shinomaya et al. 1991), an electrodeless fluorescent lamp with outer coils, was introduced in the Japanese market in 1991. The coil is wound outside a 4.5 cm diameter bulb (cf. Fig. 2) and requires increased electromagnetic screening compared to other inductively coupled lamps, which is provided by a mesh screen outside the lamp. The lamp operates at 13.56 MHz with neon as a buffer gas to provide visible light during starting.

Toroidal Lamps

Electrodeless fluorescent lamps can also be operated at low frequency (100–300 kHz), by adopting the principle of a ring discharge, similar to that of a tokamak used in thermonuclear fusion research (cf. Fig. 3). The tube ring penetrates a closed ferrite core that contains the magnetic flux; the core is wound with a primary winding to which *RF* power is applied, with the ferrite ring providing the single turn secondary. The voltage in the lamp is thus induced by a closed magnetic path. Low-frequency operation of electrodeless lamps is attractive because of the low cost of electronics and easier restrictions on *EMI*.

The concept was patented in 1970 (Anderson 1970), but the first designs required a large quantity of ferrite, and considerable power was dissipated in the

Fig. 2 Schematic of Everlight lamp

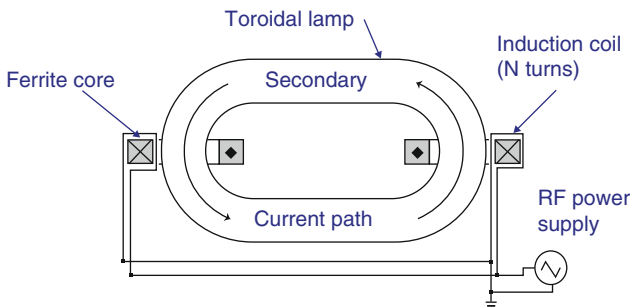
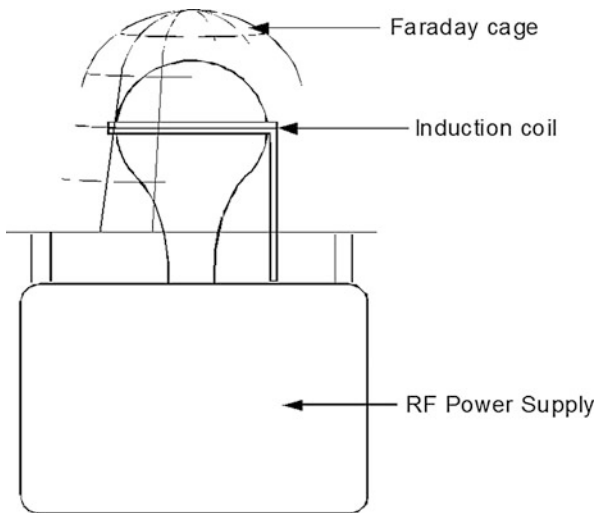


Fig. 3 Schematic of a toroidal (ring) electrodeless fluorescent lamp

ferrite ring, resulting in unacceptably low system efficacy. Godyak and Schaffer (1998) demonstrated that ferrite losses could be minimized provided the power loading was sufficiently high. They showed that the ratio of core losses P_c to power in the discharge P_d follows

$$P_c/P_d \propto I^{-1.5} \tag{5}$$

resulting in a rapid reduction in the fraction of ferrite losses as the discharge current increased.

ICETRON[®] (ENDURA[®]) was the first lamp introduced into the market based on this principle and operates at 250 kHz, similar to the electronic ballast in an electrode FL. There are currently a number of lamps of this type available in the

market, operating at frequencies of 100–300 kHz. Since the high-permeability magnetic core completely encloses the discharge current, coupling of RF power to the discharge is close to 100 %. Power supply efficiency for these frequencies is typically in the mid-90 % range.

These lamps are particularly applicable where high-power, high-illuminance lamps are required for use in inaccessible areas. The life of these induction lamps is rated at 100,000 h, limited by the life of the electronics, especially the electrolytic capacitors used in the power supply, and the lumen maintenance of the phosphors.

Further advantages of these light sources are that they produce low glare and good vertical illuminance, which is an important safety issue in areas where identification of occupants is required, such as parking facilities. On the other hand, the relatively large size of the light source presents difficulties in controlling and directing the light output with the lamp optics. The lamps also have relatively high backlight and poor horizontal illuminance uniformity, which is disadvantageous for some street lighting applications.

Electrodeless High-Frequency HID Lamps

Benefits of Electrodeless HID Lamps

Lamp Life

The tungsten electrodes in *HID* lamps are subjected to a high cathode fall during the starting phase of the lamp (the glow-to-arc transition, Lister et al. 2004) which enhances sputtering and evaporation of tungsten, while diffusion and convection during steady-state operation may transport tungsten evaporated at the electrode to other parts of the discharge, much of which reaches the walls causing wall blackening and reduction in light output. Further, overheating of the walls due to misalignment of electrodes and leaks at glass-to-metal seals can shorten lamp life. Electrode assemblies are not pure tungsten, but combinations of refractory metals, usually niobium, molybdenum, and tungsten to match the coefficient of thermal expansion of the envelope (fused silica or ceramic). Chemical reactions between the fill chemistry and electrode assemblies can cause rapid transport of the metals in strong thermal gradients leading to erosion of the electrode structure and subsequent failure.

Lamp Design

For many general lighting applications, the source of light should be unobtrusive while providing optimum illumination. Smaller lamps are preferred because they can be more accurately positioned within the optical configuration of a luminaire, resulting in high luminaire efficiency.

Although the arc tubes in *HID* lamps are very compact, they are imprecisely located in glass outer jackets, resulting in both a bulky device, which cannot be accurately positioned in an optic, and less than optimum luminaire efficiency. These jackets are necessary to prevent exposure of niobium and molybdenum

electrical leads to the atmosphere, since oxidation in air would destroy the leads in a few hours. Extremely small “white-light” sources with very high efficacies using microwave excitation are now commercially available [section “[Microwave and High-Frequency Resonant Cavity \(“Plasma”\) Lamps](#)”].

Chemical Dose

Conventional *HID* lamps have a complex chemical composition, and condensation of molten salts in the crevices near electrodes can lead to color variations and corrosion. Metal halides used in current *HID* lamps are principally iodides, but use of more volatile chlorides and bromides can lead to the formation of stable radicals at high temperatures, with enhanced broadband emission (Wharmby 1997).

Electronic Control

Starting and restarting times for electrodeless *HID* lamps are considerably shorter than for electrode lamps. Many electrodeless lamps can also be dimmed, without significant loss of efficiency, which can provide energy savings for applications where maximum lighting levels are not always required, such as for street and area lighting.

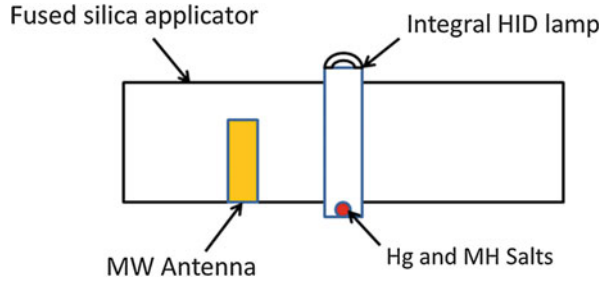
Microwave and High-Frequency Resonant Cavity (“Plasma”) Lamps

Principles of Operation

A number of resonant cavity configurations for lighting applications were reviewed by Waymouth (1993), and there is a large patent literature on the subject, much of which is cited in that paper. Advances in radio-frequency and light source technology in the last decade have led to the most recent generation of highly efficient white-light electrodeless metal-halide discharge lamps (Babykumar et al. 2007; Gilliard et al. 2011; Neate and Lister 2012). Such lamps are generally classified in the scientific literature as *electrodeless HID* (EHID) lamps, but in the lighting industry, they are referred to as *plasma lamps*. The reason the term “plasma” has been adopted to describe this type of lamp is not clear, since all discharge lamps contain plasma. However, the term is in current usage and included in Illuminating Engineering Society (IES) and American National Standards Institute (ANSI) documents in the USA.

In a plasma lamp, *DC* power is converted to *RF* using a magnetron or a solid-state circuit device. *RF* power is then coupled to the lamp through a connector (a waveguide in the case of magnetron excitation). The lamp is enclosed in an applicator, which may consist of air, fused silica, or ceramic. An example of a fused applicator is shown in Fig. 4, where the *RF* power is coupled to the power source via an antenna. The applicator acts as a resonant cavity during ignition, supplying the high electric field required to start the lamp. After starting, the applicator supplies the necessary radio-frequency power to maintain the discharge during steady-state operation.

Fig. 4 Schematic of a fused silica resonant cavity, surrounding an MH lamp and connected to the microwave input power via an antenna



Technology of Resonant Cavity Lamps

There are two distinct phases in the operation of cavity discharges (Waymouth 1993). The behavior of these two phases can best be thought of in terms of the quality factor Q :

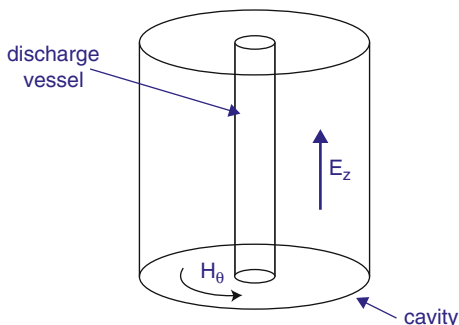
$$Q = \omega \frac{\text{Energy Stored}}{\text{Power Loss}} \quad (6)$$

where ω is the angular excitation frequency.

- (a) *Ignition*: In common with conventional discharge light sources, a much higher electric field is required to establish the discharge than to maintain it. When no plasma is present, the Q of the unloaded cavity is high (typically 100–150), and application of a high external voltage ensures a resonant coupling of power to the lamp, providing a sufficiently high electric field to ionize the constituent atoms of the buffer gas and metal atom mixture and create a discharge and ultimately a plasma. In cases where the high Q is insufficient to ignite the discharge, an additional starting electrode may be employed to provide the necessary electric field.
- (b) *Steady State*: As the plasma forms, metal atoms and salts vaporize, and the plasma becomes increasingly dissipative, absorbing power from the microwave field. Eventually, as electron density increases, the plasma becomes increasingly shielded from the microwave field due to the skin effect (see section “Physics of Resonant Cavity Lamps”), the value of Q reduces to order unity, and the microwave fields reduce to the value at the surface required to maintain the discharge.

A number of electromagnetic configurations are possible in radio-frequency discharges: transverse electric or TE modes, in which all electric field components are transverse to the direction of propagation; transverse magnetic or TM modes, in which all magnetic field components are transverse to the direction of propagation; and transverse electric and magnetic or TEM modes, in which all electromagnetic field components are transverse to the direction of propagation. A detailed analysis of these modes is beyond the scope of this chapter, and we shall concentrate on the

Fig. 5 Schematic of model cavity applicator lamp in the TM_{010} mode, showing the electric and magnetic field configurations



simplest TM mode, TM_{010} , illustrated in Fig. 5, which is the most relevant to the current discussion. A detailed discussion of other cavity configurations used for $EHID$ light sources is given by Waymouth (1993).

The dimensions of a resonant cavity are determined by the resonance wavelength λ , given by

$$\lambda = \frac{c}{\omega \sqrt{\epsilon_c \mu_c}} \quad (7)$$

where c is the speed of light in vacuum and ϵ_c and μ_c are the relative permittivity and permeability of the cavity material, respectively. Many resonant cavities contain air, but Eq. 7 shows that use of a material with higher permittivity allows for a smaller-sized cavity. In order to use alternative materials in the cavity, dielectric losses must also be small. At 20°C and 1 MHz, for air, $\epsilon_c = 1$, for fused silica, $\epsilon_c = 3.9$, and for PCA $\epsilon_c = 9.8$. Note that sapphire could be used instead of polycrystalline alumina but at much greater expense. There are also other ceramics and glasses which could be used for dielectric resonators, but they are generally expensive and often colored and act as an optical filter to the generated light.

Physics of Resonant Cavity Lamps

For the discharges described above, we find that the “skin depth” (cf. Eq. 4) is in the range $1.5\text{ mm} < \delta_c < 1.5\text{ cm}$. Equation 4 implies that the skin depth is reduced as electron density (i.e., lamp power) or applied frequency increases. The influence of gas pressure is somewhat more subtle. From Eq. 4, $\delta_s \propto \left(\frac{\nu_m}{n_e}\right)^{1/2} \propto \left(\frac{N}{n_e}\right)^{1/2}$. From the Saha equation, at constant gas temperature, $\frac{n_e}{N} \propto N^{-1/2}$ so $\delta_s \propto \sqrt{N}$, and the skin depth also decreases as pressure decreases, provided all other lamp parameters are kept constant.

In one dimension, this means that the electric field at a distance r from the wall w will decrease from its value E_w ; to a first approximation, the electric field can be calculated from the local skin depth (Waymouth 1993) according to

$$E(r) = E_w \exp \left(- \int_w^r \frac{dr'}{\delta_s(r')} \right). \quad (8)$$

This is in contrast to conventional *HID* lamps where the electric field is effectively constant across the discharge. The consequences of this behavior in *EHID* lamps for the lamp physics are discussed below.

The energy balance equation is crucial in understanding the behavior of microwave-powered lamps. The gasses in all discharge lamps are heated by collisions between the constituent electrons, ions, atoms, and molecules, described by the energy balance equation (in the absence of gas convection), Elenbaas–Heller equation,

$$\nabla \cdot \kappa_g \nabla T_g + \sigma_e E^2 - U_{rad} \quad (9)$$

where T_g is the gas temperature, κ_g is the coefficient of thermal conductivity of the gas, and U_{rad} represents the net energy transported by radiation from each point in the discharge.

Offermanns (1990) and Waymouth (1993) have both developed models to describe the power balance in microwave discharges but the physics described is applicable to all radio-frequency lamps. The crucial parameter in this regard is the skin depth, δ_s . As the value of δ_s becomes smaller, the radial temperature profile becomes flatter, with the principal effect of allowing normally opaque resonant and molecular radiation to escape from the lamp, thereby enhancing efficacy (Waymouth 1993). As noted above, δ_s can be influenced by three lamp design parameters, the electrical power input to the lamp, the gas pressure, and the operating frequency, although the choice of frequency for practical applications is greatly limited by *ISM* regulations.

There are three components to the energy balance in Eq. 9: the first is the joule heating term, $\sigma_e E^2$; the second is the net radiant energy loss, U_{rad} ; and the third is the loss by thermal conduction, $\nabla \cdot \kappa_g \nabla T_g$.

Waymouth (1993) considered a pure Hg discharge, for which the *DC* properties were well known (Zoillweg et al. 1975). He assumed that the experimental values of $\kappa_g(T_g)$, $\sigma_e(T_g)$, and $U_{rad}(T_g)$ had the same temperature dependence in the microwave discharge as in the *DC* discharge and calculated temperature profiles from Eqs. 7 and 8 for microwave discharges in the TM_{010} mode, for various values of total electric power.

From the results of these calculations, Waymouth concluded that as the electric power in a microwave discharge is increased:

- (a) The position of the maximum temperature shifts from the center of the discharge toward the wall, because the short skin depth prevents the microwave power from penetrating into the center of the discharge.
- (b) The fraction of electric power converted to radiation in the discharge, W_{rad}/W_{elec} , increases.

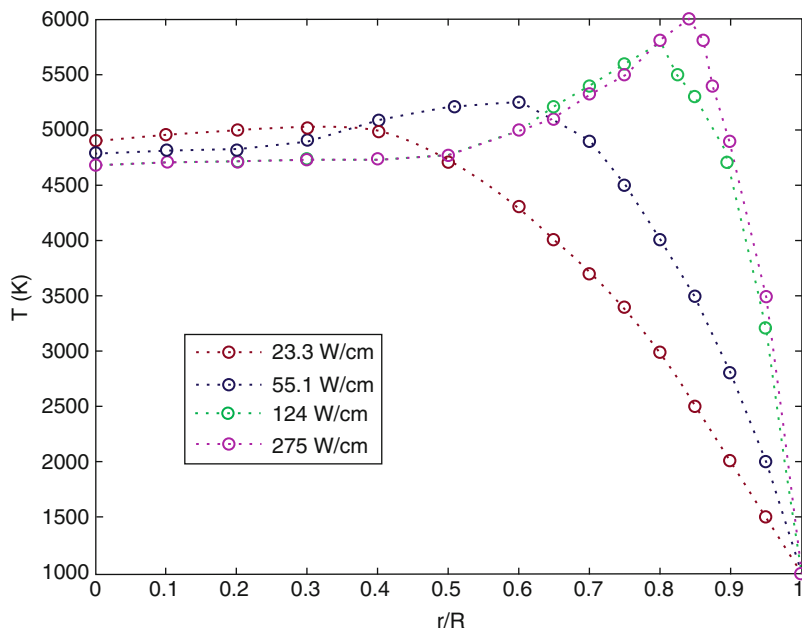


Fig. 6 Radial temperature profiles for microwave-excited discharges at 3.0 atmospheric pressure mercury vapor, ID = 2.0 cm and frequency = 2.45 GHz, for various values of total electrical input power (After Waymouth 1993)

The shift of the temperature maximum away from the discharge axis with increasing power is a common observation in microwave plasma devices (Offermanns 1990; Waymouth 1993) and is a direct result of the skin effect discussed above. At high powers, the electron density near the axis increases and the skin depth δ_s becomes comparable to the dimensions of the discharge, effectively screening the central region from the external electromagnetic fields. In the center of the discharge, the net emission coefficient of the plasma is much higher than the deposited microwave power, resulting in a temperature minimum at the axis, as shown in Fig. 6. The temperature distribution is maintained by heat flux down the temperature gradient toward the axis for $r < r_{\max}$, where r_{\max} is the radial position of the temperature maximum in the distribution.

The calculations by Waymouth were limited to the *total* radiation emitted by the discharge. Since this radiation consists of *UV*, visible and *IR* radiation, it was not possible to compute the efficacy of visible radiation from these results. The results of calculations by Hamady et al. (2013, 2015) have indicated that in some circumstances, although the total radiation efficiency may increase with increasing electrical power, the luminous efficacy may actually decrease. An increase in efficacy was predicted for ultrahigh-pressure Hg lamps as electrical power increased, but for the lamps considered by Waymouth, luminous efficacy was calculated to decrease as electric power increased.

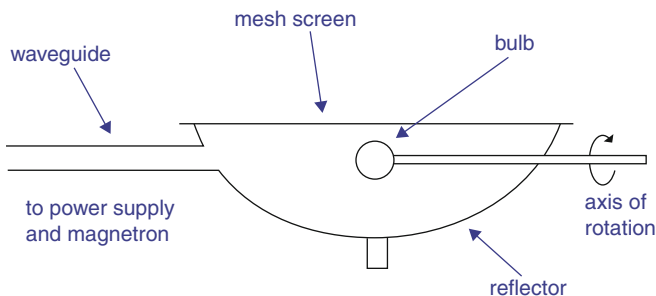


Fig. 7 Schematic of the Solar 1000[®] lamp

Commercial Examples of Resonant Cavity Lamps

The Sulfur Lamp

The first commercial plasma lamp for general lighting was the *Solar 1000* (Wharmby 1993, 1997; Lister et al. 2004), which produced light largely from diatomic sulfur. This lamp is unique in that the principal emission species are molecules not atoms. The lamp was operated by transmitting microwave power at 2.45 GHz from a magnetron through a waveguide to the lamp, which was contained in a resonant cavity (cf. Fig. 7).

The original lamp contained argon and a small amount of sulfur providing a very “white” light, mainly from the sulfur molecules in a 5900 W lamp, with a wall power loading of 250 Wcm^{-2} , which is a factor of 10 higher than conventional *HID* lamps. The excess heat was removed by rotating the lamp and by supplementary air cooling. The *Solar 1000* was developed for general lighting applications and operated at 1425 W, with a bulb diameter of less than 3 cm. Due to the greatly reduced wall loading (30 Wcm^{-2}), air cooling was not required, but the lamp had still to be rotated.

Sulfur molecules have a broad continuum of radiation in the visible spectrum with very little in either the *IR* or *UV*, and it is these molecules which were presumed to provide the major source of radiation (Johnston et al. 2002). Since sulfur is chemically reactive with most metals, it cannot be used in conventional electrode lamps (Wharmby 1993). A modification of the sulfur lamp is currently marketed by LG, particularly for sport arena applications, with electric powers of 1000 and 2000 W.

Metal-Halide Resonant Cavity Lamps

A number of MH resonant cavity lamps are currently available in the market. A lamp powered by a magnetron at 2.45 MHz and using a fused silica resonator has been developed by Ceravision. LUXIM has developed a lamp powered by a solid-state circuit at 433 MHz, using a ceramic resonator, and Topanga has a lamp operating with a solid-state circuit at 433 MHz using an air resonator. These lamps are all effectively point light sources, with consequent advantages for good

optical control in a luminaire and the advantages of long life and operation at high electrical power (>250 W). Plasma lamps are finding application for high-bay industrial lighting, high-mast roadway lighting, and lighting of sporting arenas.

Dielectric Barrier Discharge (DBD) Lamps

Principles of DBD

Dielectric barrier discharges (*DBD*) are electrical discharges between two electrodes separated by an insulating dielectric barrier. These discharges, originally called silent (inaudible) discharges (Siemens 1857), have found a wide variety of industrial applications over the past two decades (cf. Kogelschatz 2003; Kogelschatz and Salge 2008), in particular for the production of *UV* radiation, which, as in the case of fluorescent lamps, may be converted to visible radiation by means of a phosphor. Usually *DBDs* at high pressure consist of many miniscule parallel current filaments, referred to as *microdischarges*. More recently, it has been shown that under special conditions, this configuration can also produce homogeneous discharges, and this mode of operation has proved successful in developing planar light sources in Xe, with efficiencies of 172 nm *VUV* production as high as 60 % (Vollkommer and Hitzschke 1998).

Figure 8 is a schematic diagram showing a typical construction of a *DBD*, in which one of the two electrodes is covered with a dielectric barrier material. The lines between the dielectric and the electrode represent discharge filaments, which are normally visible to the naked eye.

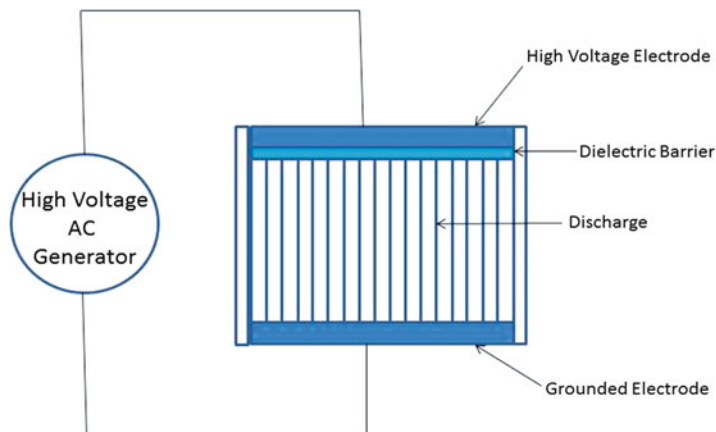


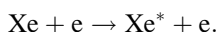
Fig. 8 Schematic of a dielectric barrier discharge (DBD)

The Physics of DBD

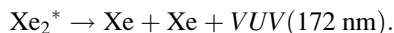
Principles of Operation

Since the dielectric barrier in a *DBD* is an insulator, it is unable to sustain a *DC* current and these discharges require *AC* operation. The amount of displacement current that can be passed through the dielectric(s) depends on its dielectric constant and thickness and the time derivative of the applied voltage, dU/dt . To transport current (other than capacitive current) in the discharge gap, the electric field must be sufficiently large to cause breakdown in the gas. In most applications, the dielectric limits the average current density in the gas space. It thus acts as a ballast which, ideally, does not consume energy. Glass or fused silica (quartz) is the preferred material for the dielectric barrier, although ceramic materials and thin enamel or polymer layers are used for special applications. At very high frequencies, the current limitation by the dielectric becomes less effective, so *DBDs* generally operate between line frequency and about 10 MHz. In most gasses, when the electric field is sufficiently high for breakdown to occur, a large number of microdischarges are observed when the pressure is of the order of 10^5 Pa, which is the preferred pressure range for most applications (cf. section “[Filamentary Discharges](#)”). On the other hand, pulsed operation of *DBD* may be used to produce a homogeneous discharge, which is beneficial for the production of *VUV* radiation in Xe.

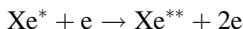
The principal processes in creating the Xe_2^* excimer are electron excitation of ground-state Xe and the subsequent three-body interaction to create Xe_2^* , i.e.,



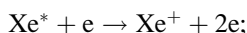
Xe_2^* then spontaneously decays to produce 172 nm radiation:



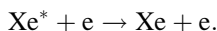
There are three main competing non-radiation-producing atomic processes: stepwise ionization of Xe^* ,



where Xe^{**} represents higher excited states of Xe; stepwise ionization of Xe^* ,



and electron quenching of Xe^* (collisions of the second kind)



These processes, together with electron–electron collisions, should be minimized in order to maximize the radiation efficiency of the discharge.

Filamentary Discharges

When a sinusoidal voltage is applied to the electrodes of a *DBD*, there are four distinct phases of operation (Kogelscahtz and Salge (2008)):

- (i) At t_1 an electric field is established in the discharge.
- (ii) At t_2 microdischarges are initiated, which short-circuit the gas gap: if the microdischarges are simultaneous, surface discharges cover the whole dielectric barrier. In some discharges, in spite of the steep voltage pulse, not all the microdischarges are initiated at the same time, and those first ignited arrive first at the barrier, and gliding discharges can spread undisturbed over a large area. The microdischarges extinguish and the voltage across the gap is close to zero, the entire process lasting about 10^{-9} s. However, channels of the extinguished microdischarges remain in the gap and as the recovery continues, their conductivity decreases. These channels provide preferred locations for the ignition of new microdischarges when the applied voltage is reversed.
- (iii) At t_3 the voltage across the gap is then dominated by the memory charges deposited at the surface.
- (iv) At t_4 new microdischarges of opposite polarity are ignited, the return path being preferentially along the residual channels of previous microdischarges, if these channels have not recovered sufficiently in the meantime.

Homogeneous Discharges

Investigations have also shown that dielectrically controlled homogeneous or uniform discharges can be obtained under special conditions (Kogelschatz and Salge 2008). Volkommer and Hitzschke (1998) showed that, by applying electrical pulses to the discharge at the appropriate repetition rate, they could obtain homogeneous discharges with an efficiency of 172 nm *VUV* production of 60 %.

Volkommer and Hitzschke found that, by inducing an extreme non-LTE electron energy distribution function (*EEDF*) in the discharge, the production of metastable Xe could be enhanced while minimizing the quenching by excitation to higher states. In order to prevent the *EEDF* from rapidly relaxing to a Maxwellian distribution, the electron density in the discharge was also required to be low. It is not possible to achieve these conditions in standard *AC* operated *DBD*, due to the filamentary nature of the discharge. However, optimum efficiency of *VUV* production can be obtained by pulsing the discharge at a specific repetition rate, chosen such that during the idle time between pulses, the plasma is able to relax to the state existing prior to the excitation. In this regard, the parameters of the *DBD* resemble those of a glow discharge. The specific voltage shape to be applied depends on the specific fill gas, the pressure and the electrode configuration.

DBD Planar Light Sources

Planar DBD Lamps

Xe is the most appropriate gas for use in general lighting, since 172 nm *VUV* radiation from the Xe_2^* excimer can be efficiently produced and converted to visible radiation using a phosphor, as in a standard fluorescent lamp. Flat electrode fluorescent lamps are inefficient for aspect ratios greater than 3:1 (Ingold 1991), but *DBD* fluorescent lamps have no such dimension limitation (Vollkommer and Hitzschke 1998; Hitzschke and Vollkommer 2001; Herring et al. 2012). Although the phosphor conversion to visible light is less efficient than for the Hg 254 nm, in the absence of mercury, the degradation of the phosphor is reduced, and the lamp reaches full light output almost immediately. Lamps typically operate at full luminance for ambient temperatures from $-30\text{ }^\circ\text{C}$ to $+85\text{ }^\circ\text{C}$, and there are no environmental issues with lamp disposal.

Pulsed-Excitation DBD

Pulsed-excitation *DBD* is the basis for the PLANON[®] lamp, developed by OSRAM (Hitzschke and Vollkommer 2001). The lamp is an 8.5 mm thick flat panel. The arrangement consists of two parallel glass sheets with a gap between them. The inner surface of each is coated with a phosphor, and the space between them is sealed and filled with xenon gas at a pressure of 10^4 Pa.

The anode and cathode electrodes are printed onto the lower glass sheet in a regular array, with several tens of electrode pairs spread throughout the panel. The anodes are point-like tips and the cathodes are continuous lines, with the result that the discharge occurring between each electrode pair is triangular in shape. Lamp life is rated at 100,000 h, and the lamps are dimmable to 20 % light output, with maximum luminance 1200 cd/m^2 . PLANON[®] lamps have been applied in *LCD* panel backlights, in medical *X-ray* viewers, in architectural lighting installations, and on the *KPN* Telecom tower in Amsterdam.

High-Pressure Microdischarges

High-pressure microdischarges generated in spatially confined cavities have been investigated for a number of applications since the mid-1990s: Schoenbach et al. (1997) were the first to report stable high-pressure ($>4 \times 10^4$ Pa) discharges in cylindrical hollow cathode geometries. Later developments showed that *2D* arrays of individual microdischarges could be operated in parallel, or in series or both (Becker and Schoenbach 2008). Microdischarges can be operated in parallel without individual ballast resistors if the discharges are operated in the range where the current voltage (*I-V*) curve has a positive slope (Schoenbach et al. 1997).

Microplasma planar lamps have recently been developed by Eden Park Illumination (Herring et al. 2012) using flat panel *DBD* discharges based on high-pressure microdischarge technology. Microplasma lamps operate by confining the gas discharge inside the lamp into thousands of microcavities formed into the glass. This confinement allows stable operation at higher pressures than conventional cold discharges, which results in higher efficiencies. Figure 9 shows a cross section of

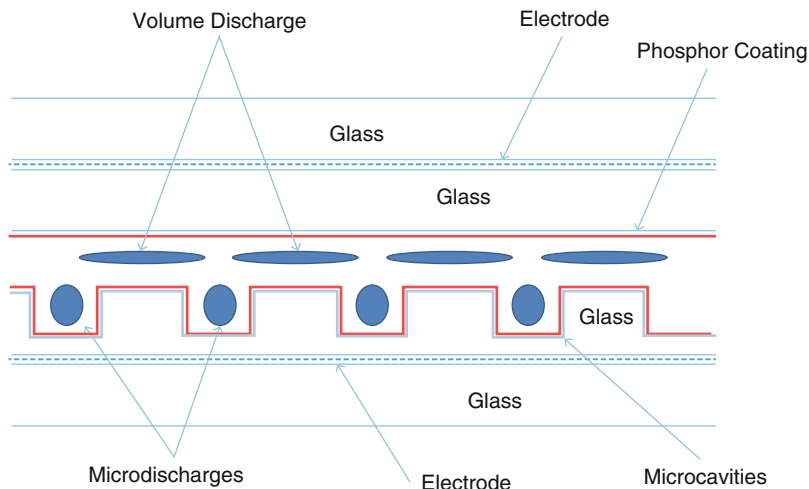


Fig. 9 Cross-sectional view of a flat panel lamp showing microcavities inside the glass envelope

the glass envelope comprising the lamp structure. The microcavities are etched into the glass before the application of phosphor and before the lamp is assembled. The lamps are dimmable and, in common with other Xe-based lamps, require no warm-up time and operate over a wide range of ambient temperatures.

As shown in Fig. 9, the lamp is composed of four layers of glass, each layer being 1.1 mm thick, and the total lamp thickness, including the gas discharge layer and electrodes, is <math>< 5\text{ mm}</math>. Since the glass itself serves as the mechanical structure of the lamp, there is no need for additional external structures for support or power delivery, so the lamp can maintain the 5 mm thickness for lighting tile areas $> 1\text{ m}^2$. In principle, the lamps can be operated with a variety of shapes, provided the discharge gap remains constant. The discharge gap spacing is determined by using spacers strategically positioned throughout the lamp structure, forming a grid pattern. Lamps operate at 40 lm/W, with luminance 80,000 cd/m². The life of the lamp is 50,000 h, the time at which the lamp luminance is 70 % of its initial value, determined by the degeneration of the phosphor.

Electrodeless Excilamps

UV Light Sources

As has been discussed in the previous sections (“[Electrodeless Fluorescent Lamps](#),” “[Dielectric Barrier Discharge \(DBD\) Lamps](#),” “[Electrodeless Excilamps](#)”), light sources that produce *UV* radiation are applied to general lighting by the use of a phosphor to convert the UV photons to visible photons. However, UV and VUV

light sources have a number of other applications, including advanced semiconductor treatment technologies, synthesis of new materials and modification of material properties, chemical processes in medicine and biology, and disinfection of water, air, and industrial waste.

UV photons, with energies of 5–10 eV, are capable of splitting most chemical bonds (Gellert and Kogelschatz 1991), and they have proved particularly effective in the deactivation of biological systems. For this reason, *UV* sources have been widely used as a method of disinfection, particularly in the water industry. Traditionally, mercury lamps have been the main *UV* source used for this purpose, due to their simple power supply systems and easy maintenance. However, the radiation power of low-pressure Hg lamps is very sensitive to variations in ambient temperature, and the use of mercury has a negative environmental impact.

In the last two decades, much attention has been given to a class of gas discharges which produce a relatively narrow spectrum of ultraviolet (*UV*, $200 \text{ nm} < \lambda < 400 \text{ nm}$) and vacuum ultraviolet (*VUV*, $100 \text{ nm} < \lambda < 200 \text{ nm}$) radiation from excimer molecules (rare-gas molecules, e.g., Xe_2^* or halogen molecules, e.g., Cl_2^*) or exciplex molecules (typically rare-gas halide molecules, ArX^* , KrX^* , XeX^* , where $\text{X}=\text{Br}, \text{Cl}, \text{I}$); the choice of molecule is dependent on the radiation wavelength appropriate for the particular application. In the literature, these lamps are usually referred to as “excimer lamps,” “exciplex lamps,” or more generally “excilamps” (Sosnin et al. 2006; Lomaev et al. 2012).

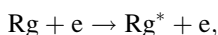
Excilamps are sources of spontaneous radiation that employ non-equilibrium radiation of excimer and exciplex molecules. One of the advantages of these molecules, compared to mercury, is the absence of a strong molecular bond in the ground state, and this eliminates radiation self-absorption, thus providing favorable conditions for radiation transport exiting the lamp. A further advantage is the independence of the discharge ignition on the temperature of their operating media and the short time of discharge ignition.

Most excilamps today are electrodeless discharges, in particular dielectric barrier discharges (*DBD*), but electrodeless UV sources have also been produced in capacitively coupled discharges (*CCD*, Lomaev et al. 2012) and microwave discharges (Kumagai and Obara 1989). This enables excilamps to operate with many of the benefits of electrodeless lamps discussed in the introduction to this chapter: the wide range of bulb geometries, instantaneous triggering to maximum radiation power, variable tuning of the photon flux, and increased lamp life.

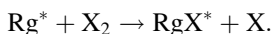
The major limiting factor on the lifetime of excilamps is the decrease of halogen concentration in the gas mixture due to interactions with glass or fused silica (quartz) walls (Lomaev et al. 2012). This can be mitigated to some extent by passivation of the walls by active passivation of the walls with halogen molecules (Tarasenko et al. 1998). Further increase in lifetime can be achieved at lamp sealing by increasing the fraction of halogen to rare gas 10–20 % above its optimum value, typically 1:200–1:100. Under optimum conditions, lifetimes of 1000–2000 h have been reported (Lomaev et al. 2012; Tarasenko et al. 1998).

The Physics of Excilamps

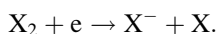
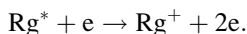
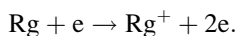
The atomic and molecular processes in excilamps are highly complex, but the creation of exciplexes and subsequent dissociation and production of VUV or UV photons involve a limited number of simple processes. At low pressure, the rare-gas atoms Rg (Rg=Ne, Ar, Kr, Xe) are excited into metastable states by electrons



and the rare-gas halide exciplex is formed from the “harpoon” reaction between the excited rare-gas atom and a halogen molecule X_2 (where X=Cl, Br, I, F)



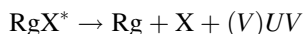
As pressure increases, the production of rare-gas ions Rg^+ increases, by ionization or two-step processes, as does the production of negative halogen ions X^- :



The excimer is then formed by combination of the ions in a three-body process:

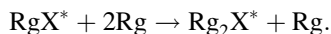


The dissociation of the exciplex

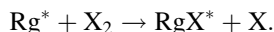


is thus dependent on the twin processes, the harpoon reaction and recombination of ions, which in general means there is an optimum pressure in each discharge for maximum production of $(V)UV$. As in the case of UV production in rare-gas discharges (section “[The Physics of DBD](#)”), there are a number of competing mechanisms, such as quenching (Zhang and Boyd 1996) and electron–electron collisions, that dissipate energy in the discharge; these processes must be minimized in order to optimize UV efficiency.

At low pressures, the dominant quenching mechanism is direct quenching by the halogen-bearing species:



However, at high pressures, three-body reactions involving the rare-gas atoms quench the excited rare-gas halides by forming triatomic species:



As has been noted above, the absence of a stable ground state for the exciplex molecule means that there is no self-absorption of the radiation.

The excilamp spectrum as a rule is a rather narrow, intense emission band of half-width 2–15 nm for RgX^* molecules and up to 30 nm for Rg_2^* molecules (Lomaev et al. 2012). In addition to the B–X emission band, the spectrum of the plasma of the RgX^* exciplex molecules can contain the D–X, C–A, and D–A transition bands of the same molecule. However, their contribution to the radiation power at high gas pressures is very small, and excilamps are normally characterized by emission in the B–X band. The emission spectra may also include radiation from the halogen dimers. For example, the luminescence spectra of a krypton mixture with molecular bromine at moderate pressures exhibit the emission bands of the Kr_2^* and KrBr^* molecules whose contributions to the total radiation power can be varied by varying the gas concentration ratio in the mixture. Nevertheless, this emitting system is generally referred to as a KrBr excilamp.

Excilamps may also contain pure halogen gasses, which produce radiation from the D→A band of X_2^* halogen dimers, with similar excitation and quenching processes to those in rare-gas halogen mixtures.

The parameters influencing the efficiency of *UV* production in excilamps include:

- (i) Excitation method (waveform, pulse rate, etc.)
- (ii) Gas pressure
- (iii) Gas mixture (ratio of X_2/Rg partial pressures)
- (iv) Geometry of the discharge
- (v) Total electric power in the discharge

Due to the complexity of the physical processes in excilamps, it is difficult to provide generalizations as to the optimization of *UV* production, as each lamp will have a set of unique properties that require thorough investigation. There are, however, a number of important trends, and these will be summarized in the following sections.

DBD Excilamps

DBD excilamps are traditionally excited using generators producing a sinusoidal voltage pulse shape. In this type of device, the efficiency of converting electric power to *UV* radiation is typically 10–15 % (Lomaev et al. 2012; Zhang and Boyd 1996). The conditions for the formation of an individual microdischarge, the discharge as a whole, and the efficiency and radiation power of an excilamp are greatly influenced by the pulse repetition rate *f*. For many discharges, there is an

optimum repetition rate f_0 (Lomaev et al. 2012). For $f < f_0$, chaotically distributed diffuse cylindrical channels of low brightness are observed in the gas discharge. As the pulse repetition rate is increased ($f > f_0$), the density of the microdischarges and the size of their feet and brightness increase. The excitation power and, hence, the radiation power thus increase but the efficiency of converting electric power to radiation decreases. For *DBD* discharges in XeCl, the optimum frequency f_0 was found to be ~ 1 kHz (Lomaev et al. 2012).

For *DBD* in Xe₂, Kr₂, KrCl, and XeCl excilamps, a sinusoidal voltage pulse shape was found to be preferable to a short high-voltage pulse of duration 50–100 ns for excitation by a barrier discharge (Lomaev 2012). The low efficiency with short high-voltage pulses has been attributed to the fact that the energy deposited in the gas discharge plasma is lower compared to that stored in the discharge. Another cause is a significant overvoltage across the discharge gap, and the reduced electric field strength in the gas discharge plasma is therefore not optimal for the formation of exciplex molecules. However, it has also been shown that the efficiency of a *DBD* Xe₂ excilamp increases when excited by a short voltage pulse compared to that excited by a sinusoidal voltage pulse (Volkommer and Hitzschke 1998; Mildren and Carman 2001]. Because exciplex and excimer molecules are formed in different ways, there are grounds to suppose that the conditions for attaining maximum efficiency can also differ.

The gas pressure is one of the main factors influencing the efficiency of *UV* production. As indicated above, decreasing the mixture pressure leads to an increasingly more homogeneous discharge. The optimum pressure and Rg//X₂ ratio are strongly dependent on the discharge parameters, but optimum pressures are typically of the order of a few tens of kPa and Rg//X₂ ratios between 100:1 and 400:1.

The spectral and energy characteristics of *DBD* plasmas generating spontaneous emission of D→A band radiation excited by X₂* halogen dimers in homonuclear chlorine, bromine, and iodine gasses have also been investigated (Lomaev et al. 2012) with efficiencies of 2–4 %, achieved with the addition of Ar buffer gas.

Capacitively Coupled Excilamps

The discharge in an electrodeless *CCD* is separated from the electrodes by dielectric layers. As in the case of *DBD*, the presence of the dielectric layer limits the energy delivered to the plasma during an individual excitation pulse. *CCD* for excilamps suffer from many of the issues that have prevented their use for general lighting. *CCD* excilamps operate at low pressure (up to a few kPa) with a much longer electrode gap (a few tens of centimeters) than for *DBD* lamps.

CCD excilamps provide a relatively low (< 10 mW/cm²) radiation power, due to the low excitation level at low pressures of the operating medium. These lamps are used where high radiation power densities are not required. Studies by Lomaev et al. (2012) indicate that the lamp design, the electrode dimensions and arrangement, the pressure of the operating medium, and the excitation mode influence the

time and specific energy characteristics of the lamps. An advantage of air-cooled *CCD* is that these lamps exhibit longer lifetimes than *DBD* lamps (Lomaev et al. 2012; Tarasenko et al. 1998).

In *CCD* excilamps, as in *DBD* excilamps, the most intense transition is the $B \rightarrow X$ transition of the exciplex molecules. However, due to the reduced rate for collisional relaxation processes at low pressure, the radiation spectra include clearly defined $D \rightarrow X$ and $C \rightarrow A$ transitions of the exciplex molecules (see Lomaev et al. 2012). Spectra from XeBr^* and XeCl^* molecules may also include a broad intense blue wing of the $B \rightarrow X$ transition. At very low pressures (200 Pa) in *XeI CCD* lamps, the most intense line is the iodine transition at 206 nm, but as the pressure is increased, the ratio of the spectral energy densities of the iodine line and that of the $B \rightarrow X$ band varies, and at a pressure of 1 kPa spectra of the two lines become comparable.

Microwave Discharge Excilamps

To date there has been only one report of an excilamp operated by microwave power (Kumagai and Obara 1989). A *DC* voltage of 4 kV was applied to a magnetron, producing microwave power at a frequency of 2.45 GHz. Microwaves propagate through directional couplers, a three-stub tuner, and a coupling hole into a cylindrical cavity made of a metallic mesh, creating a strong electromagnetic field in the TE_{111} mode. The gas discharge tube contained a mixture of Kr and F_2 , diluted with either Ar or Ne, producing KrF^* UV radiation at 248 nm. The diameter of the discharge tube was 30 mm, chosen to prevent electromagnetic field inhomogeneity and ensure a uniform discharge. The discharge was operated at total gas pressure of 13 kPa and achieved a maximum efficiency of 12.3 % at 120 W average microwave power deposition, falling to 8.3 % at 678 W.

Acknowledgements I am grateful for helpful comments and advice from Gary Eden, Walter Lapatovich and John Waymouth during the preparation of this chapter.

References

- Anderson JM (1970) US Patent # 3,500,118
- Babykumar V, Neate A, Odell E, Preston B, Sadiq A, Devonshire R (2007) A compact microwave resonant cavity electrodeless light source system. In: Liu MQ, Devonshire R (eds) Proceedings of 11th international symposium on the science and technology. FAST-LS, Sheffield, p 110
- Becker K, Schoenbach KH (2008) High-pressure microdischarges. In: Hippler R, Kersten H, Schmidt M, Schoenbach KH (eds) Low temperature plasma physics, vol 2, 2nd edn. Wiley, Weinheim, pp 463–493
- Bethenod J, Claude A (1936) US Patent # 2,030,957
- Brown SC (1959) Basic data of plasma physics, 1st edn. MIT Press, Cambridge, MA, p 142
- Doell GW, Lapatovich WP (1990) US Patent 5,889,368
- Gellert B, Kogelschatz U (1991) Generation of excimer emission in dielectric barrier discharges. *Appl Phys B* 52:14–21

- Gilliard RP, DeVincentis M, Afidi A, O'Hare D, Hollingsworth G (2011) Longitudinally mounted light emitting plasma in a dielectric resonator. *J Phys D Appl Phys* 44:224008 (8pp)
- Godyak V, Schaffer J (1998) Endura: a new high output electrodeless fluorescent light source. In: Proceedings of 8th international symposium on the science and technology of light sources, Greifswald. KIEBU-DRUCK GmbH, pp 14–23
- Hamady M, Lister GG, Stafford L (2013) Emission spectra from direct current and microwave powered Hg lamps at very high pressure. *J Phys D: Appl Phys* 46:455201 (10pp)
- Hamady M, Lister GG, Zisis G (2015) Calculations of visible radiation in HID lamps, *Lighting. Res Technol.* 10.1177/1477153515571678 (14 pages)
- Herring C, Bulson J, Dugan M (2012) Microplasma planar lighting http://edenpark.com/wp-content/uploads/Microplasma_Planar_Lighting-by_Eden_Park_Illumination.pdf
- Hitzschke L, Vollkommer F (2001) Product families based on dielectric barrier discharges. In: Proceedings of 9th international symposium on the science and technology of light sources, Cornell (R S Bergman ed.) Cornell University Press Cornell Ithaca, pp 411–421
- Ingold JH (1991) Ambipolar diffusion theory of the rectangular positive column with quadratic ionization. *J Appl Phys* 69:6910–6917
- Johnston CW, van der Heijden HWP, Janssen GM, van Dijk J and van derMullen JJAM (2002) A self consistent LTE model of a microwave-driven high pressure sulfur lamp *J. Phys. D:Appl Phys* 35:342–351
- Kogelschatz U (2003) Dielectric-barrier discharges: their history, discharge physics, and industrial applications. *Plasma Chem Plasma Proc* 23(1):1–46
- Kogelschatz U, Salge J (2008) High pressure plasmas: dielectric-barrier and corona discharges – properties and technical applications. In: Hippler R, Kersten H, Schmidt M, Schoenbach KH (eds) *Low temperature plasma physics*, vol 2, 2nd edn. Wiley, Weinheim, pp 439–462
- Kumagai H, Obara M (1989) New high-efficiency quasi-continuous operation of a KrF (B-X) excimer lamp excited by microwave discharge. *Appl Phys Lett* 55:1583–1584
- Lapatovich WP (2012) Electrodeless lamp technology overview. *Light sources 2012; Proceedings of 13th international symposium on science and technology lighting. Fast-LS, Sheffield*, pp. 193–207
- Lister GG, I Y-M and Godyak VA (1996) Electrical conductivity in high frequency plasmas. *J. Appl. Phys.* 79:8993–8997
- Lister GG (1999) Electrodeless discharges for lighting. In: Schlüter H, Shivarova A (eds) *Advanced technologies based on wave and beam generated plasmas*. Kluwer, Dordrecht
- Lister GG, Lawler JE, Lapatovich WP, Godyak VA (2004) The physics of discharge lighting. *Rev Mod Phys* 76:541–598
- Lomaev MI, Sosnin EA, Tarasenko VF (2012) Excilamps and their applications. *Prog Quantum Electron* 36:51–97
- Mildren RP, Carman RJ (2001) Enhanced performance of a dielectric barrier discharge lamp using a short pulsed excitation. *J Phys D Appl Phys* 34:6
- Moisan M, Zakrewski Z (1992) Surface-wave plasma sources. In: Moisan M, Pelletier J (eds) *Microwave excited plasmas*. Elsevier, Amsterdam, pp 123–180
- Neate AS, Lister GG (2012) Microwave powered metal halide discharge lighting systems, In: *Light sources 2012; Proceedings of 13th international symposium on science and technology lighting. Fast-LS, Sheffield*, pp 185–192
- Netten A and Verheij C M (1991) *The Operating Principles of the Philips QL Lamp System* (Eindhoven: Philips Lighting)
- Offermanns S (1990) Electrodeless high-pressure microwave discharges. *J Appl Phys* 67:115
- Schoenbach KH, El-Habach A, Shi W, Ciocca M (1997) High-pressure hollow cathode discharges. *Plasma Sources Sci Technol* 6:468–477
- Shinomaya M, Kobayashi K, Higashikawa M, Ukegawa S, Matsuura J, Tanigawa K (1991) Development of the electrodeless fluorescent lamp. *J Illum Eng Soc* 20–1:44–49
- Siemens W (1857) Poggendorff's. *Ann Phys Chem* 102:66–122
- Sosnin EA, Oppenlander T, Tarasenko VF (2006) Applications of capacitive and barrier discharge excilamps in photoscience. *J Photochem Photobiol C* 7:145–163

- Tarasenko VF, Chernov EB, Erofeev MV, Panchenko AN, Skakun VS, Sosnin EA (1998) Reliability and lifetime of UV excilamps pumped by glow, barrier and capacitive discharges. In: SPIE conference on laser applications in microelectronic and optoelectronic manufacturing IV. San Jose, Jan 1999, pp 425–432
- Vollkommer F, Hitzschke L (1998) Dielectric barrier discharge. In: Proceedings of 8th international symposium on the science and technology of light sources, Greifswald. KIEBU-DRUCK GmbH, pp 51–60
- Waymouth JF (1993) Applications of microwave discharges to high power light sources. In: Ferreira CM, Moisan M (eds) Microwave discharges, fundamentals and applications, NATO ASI series B: physics. Plenum Press, New York, pp 427–443
- Weibel ES (1967) Anomalous skin effect in a plasma. *Phys Fluids* 10:741–748
- Wharmby DO (1993) Electrodeless lamps for lighting: a review. *IEE Proc A* 140:465–473
- Wharmby DO (1997) Electrodeless lamps. In: Coaton JR, Marsden AM (eds) *Lamps and lighting*, 4th edn. Arnold, London, pp 216–226
- Zakrewski Z, Moisan M, Sauvé G (1992) Physical principles of microwave plasma generation. In: Moisan M, Pelletier J (eds) *Microwave excited plasmas*. Elsevier, Amsterdam, pp 11–52
- Zhang J-Y, Boyd IW (1996) Efficient ultraviolet sources from a dielectric barrier discharge in rare gas/halogen mixtures. *J Appl Phys* 80:633–638
- Zollweg RJ, Lowke JJ, Liebermann RW (1975) Arc constriction in lamps containing mercury and iodine. *J Appl Phys* 46:3828–3838

Index

A

- Absolute scotoma, 763
- Access Point (AP) 2, 684
- Accommodation, 759
- Acidification, 939, 944
- Acoustic impedance, 509
- AC ripple, affecting light sensor, 526
- AC transformer, 24
- Action spectrum, 831
- Activation energy, 187
- Activators
 - 3d–3d transitions of TM ions, 189–190
 - 4f–4f transitions of RE ions, 190–193
 - 4f–5d transition of RE ions, 193–197
- Actuation model, 539
- Adaptive control
 - commercial building lighting, 586–590
 - dimming, 590–594
 - energy savings potential, 604
 - energy savings strategies, 594–599
 - technology trends, 585
- Adaptive sampling, 542–546
- Adhesives, 259–260
- Age-related macular degeneration (ARMD), 871
- Aging population
 - daylight, 853–856
 - lighting design for, 851–853
 - senior care facilities, 856–861
 - vision, 849
- AlGaAs heterojunctions, 46
- AlInGaP alloy, 48
- Alternate thermal transfer methods, 424–426
- Alternate vapor deposition methods, 421–424
- Alzheimer's disease, 858
- Ambient light, 852
 - measurement, 517
 - sensor architecture, 523–524
- Ambient light sensors (ALS)
 - basic operation, 612–618
 - CCD imaging sensors as, 630–632
 - commissioning, 624–629
 - control algorithms, 622–624
 - daylight responsive dimming systems, 609–611
 - maintenance control strategy, 611–612
 - reaction of occupants, 629
 - spatial response, 618–620
 - spectral response, 620
- Analog dimming, 450
- Analog dimming ballasts, 591
- Analog to digital converter, 523
- Angle of arrival (AOA), 672–675
- Angular CCT deviation (ACCTD), 284
- Anode dissolution, 233
- Aqueous humor, 759
- Architecture for control networks (ACN), 566–567
- Arctube seals, 1128–1132
- Art conservation, 731, 734
- Artificial light at night (ALAN)
 - arthropods, 967–971
 - birds and reptiles, 973–976
 - fish and amphibians, 971–973
 - human perception, 959–963
 - mammals, 977–979
 - plants, 964–967
 - uses, 958
- Artificial lighting
 - economical impact, 927–932
 - energetic impact, 922–926
 - environmental impact, 926–927
- Artificial polarization, 962
- Astronauts, 841
- Asymmetric traffic distribution, 686
- Atmospheric pressure metal-organic chemical vapor deposition (AP-MOCVD), 96

- Atomic force microscopy (AFM), 76, 108
 Autoionization, 204
 Automotive lamps, 1047
 Autonomous underwater vehicles (AUVs), 540
- B**
- Backhaul network, 693–695
 Ballast, 25, 27, 31, 1071
 Bamboo fibers, 12
 Bary's room model, 640
 Base Station (BS), 2, 684
 Biological effects, 836–838
 Bioluminescence, 183
 Bis(2,4,5-trichloro-6-carboxyphenoxyphenyl) oxalate, 183
 Black body radiation, 470–471
 Blackbody, 1017–1018
 Blue, 839
 Blue light effective radiance dose, 874
 Blue light exposure data, 889–891
 Blue light hazard, 872–876
 limitations, 888–889
 radiance, 878
 Blue light hazard efficacy
 definition, 883
 luminous radiation, 883
 Blue-light weighted radiance, 873
 Bluetooth Low Energy (BLE), 667
 Bonding wire, 226, 232–233
 Bottom-emitting OLEDs, 366
 Bragg-mirror, 127, 129, 134, 137
 Bragg scattering, 375
 Bright light, 839
 Brightness, 993, 1000, 1002
 Bright white light, 838
 Broca-Sulzer effect, 775
 Brown adipose tissue (BAT), 978
 Building automation and control networks (BACnet), 567–569
- C**
- Camera coordinate, 674
 Capacitively coupled excilamps, 1168–1169
 Carbon-filament lamps, 18
 Carotenoids, 705
 Cataphoresis, 1069
 Cathodoluminescence, 184
 CCD imaging sensors, 630–632
 Ce-doped lutetium yttrium aluminum garnet, 185
 Ceiling sensor pattern, 478
 Cellular networks, 686
 Center for Medicare & Medicaid Services (CMS), 857
 Ceramic arc tube, 1126, 1128, 1134
 Ceramic metal halide lamps, 1132, 1137
 CFL's. *See* Compact fluorescent lamps (CFL's),
 Channelizing photodiode, 520
 Charge generation layer (CGL), 336
 Chemical dose, 1143
 Chemiluminescence, 183
 Chinese association of lighting industry (CALI), 931
 Chip-on-board (COB) package, 259
 Chromaticity, 789–791, 912, 914
 diagram, 1002
 CIE chromaticity chart, 452
 Circadian
 clock, 998
 light intensity, 834–836
 light wave length, 830–832
 neural pathways, 832–834
 phototransduction, 832
 rhythm support, 858–860
 Clocking, 631
 Closed loop
 algorithms, 625
 daylight responsive functions, 529
 mode, 536
 CMH lamps. *See* Ceramic metal halide lamps
 Color appearance models (CAMs), 791
 Color discrimination, 840
 Color discrimination index, 804
 Colored filters, 901
 Color fidelity scale, 812
 Color harmony rendering index (HRI), 809
 Colorimetric properties, 1002
 Colorimetry, 777, 791, 793
 Color matching functions (CMFs), 787, 792
 Color metrics, 794–795
 Color psychophysics, 776–780
 Color quality scale (CQS), 810–812
 Color rendering, 792–794
 Color-rendering capacity (CRC), 805
 Color rendering index (CRI), 281–284, 323,
 793, 795, 801–803
 Color rendition metrics
 beyond color fidelity, 803–806
 color rendition vectors (CRVs), 814
 CQS, 810–812
 CRI, 801–803
 drawbacks, 806, 808
 engineering, 816–821
 Color rendition vectors (CRVs), 813–816

- Color sensing, light sensor, 521
 Color shift, 452, 744, 1133, 1136
 Color spread, 1133
 Color temperature, 1000, 1002, 1003
 Commercial IRC lamps, 1054–1056
 Commission Internationale de l'Éclairage, 993
 Communications infrastructure, 463
 Communications protocol, 458, 459
 Compact fluorescent lamps (CFL's), 8, 35
 Composite buffer layer structure (CBLS), 105, 106, 108, 110
 Conducted interference/terminal disturbance voltage (TDV), 1144–1145
 Conduction shape factor, 244–246
 Cones, 832
 Configurational coordinate model, 186–188
 Conservation, 998
 Contrast ratio, 449, 450, 458
 Contrast sensitivity function (CSF), 773
 Control, 447, 449, 458
 Controlled laboratory study, 903
 Control network protocol (CNP), 569
 Cortisol pattern, 903
 Covalency, 185
 Crack, 232
 Crew quarters, 840
 Critical flicker fusion frequency (CFF), 774–776
 Cross-relaxation process, $4f$ - $4f$ transition, 191
 Cryptochrome, 705
 Cryptochrome photosystem, 705
 Cultural heritage, 723–724, 729
 Curfew hour, 997, 998
 Current efficiency (CE), 327
 Cyclometalated iridium (III) complex, 303
- D**
- Damp location, 747
 Danner machine, 20
 Darkness, 992, 994, 998, 1006
 Dark-Sky Association, 996
 Data backhaul, 463
 Dawn simulator, 838
 Daylight, 898
 and artificial lighting, 899
 harvesting, 527
 and lighting distribution, 903
 and man, 899
 responsive dimming system, 609–611
 responsive luminaire block diagram, 529–531
 responsive luminaire components, 528
 responsive luminaire integration, 529
 0–10 V DC, 563–564
 DC bus, 462
 DC power distribution, 462
 Decree, 997, 999, 1004
 Delamination, 230, 231
 Delayed handover, 697
 Delay-spread model accuracy, 657
 Demand response, 595–596
 Dementia, 858
 Device description language (DDL), 567
 DeVries Rose law, 771
 Dexter energy transfer, 297
 Die attach, 226, 229–230
 Dieke's diagram, 190, 191
 Dielectric barrier discharge (DBD) lamps, 1142, 1160–1164
 Digital addressable lighting interface (DALI), 564, 593–594
 Digital control, 593
 Digital signage, 677
 Dimming, 590–594, 1136–1137
 0–10 V, 459–460
 analogO, 452–454
 control technique, 617–618
 decision flow chart, 454–455
 importance, 447
 LED, 447–451
 PLC, 455–458
 PWM0, 452–454
 switched, 448
 system integrator, 454–455
 1,2-Dioxetanedione decomposition, 184
 Diphenyl oxalate oxidation, 184
 Directional wireless, 687–688
 Direct linear transformation, 673, 674, 679
 Disability glare, 867
 Discharge properties, 1146–1147
 Distributed sensor networks (DSN), 540
 Distribution, 937, 938
 DMX512, 565
 Doppler effect, 505
 Doppler shift, 504–506
 Drive current, 225, 230, 235
 Driver, 448, 451, 459, 462, 590
 Dumet wire, 13
 Duv value, 909
 Dye relaxation, 184
 Dynamic multi-scale adaptive sampling (DMSAS) approach, 545
- E**
- Ecological, 993, 1004
 Edisonian research procedure, 14

- Educational lighting
 adverse effects of lighting, 901
 ambient lighting and horizontal illuminance levels, 900
 colored filters, 901
 color temperature, 900
 daylight, 898 (*see also* Daylight)
 electrical lighting, 900
 empirical studies, 898
 full-spectrum lamps, 901
 man and daylight, 899
 non-visual effects, 899
 pre-programmed schemes, 902
 vertical illuminance levels, 902
 visual perception, 900
- EDX analysis, 163
- Effective field of view (FOV)
 angle, 873
- Eight-dimensional matrix model-based approach, 550
- Electrical lamp market, 31–38
- Electrical lighting, 900
- Electroded fluorescent lamps. *See* Fluorescent lamps (FL)
- Electrodeless excilamps, 1164–1169
- Electrodeless fluorescent lamps
 benefits, 1148–1149
 development of, 1148
 physics of, 1149–1150
- Electrodeless high frequency HID lamps,
 benefits, 1153–1154
- Electrodeless lamps, 1067
- Electrodeless radio frequency (RF) lamps,
 1142, 1143
 electromagnetic interference (*EMI*) and safety, 1143–1144
 classification, 1146
- Electroluminescence, 43, 297–299
 in GaP, 45
 in SiC, 44
 ZnS, 45
- Electromagnetic induction, 6
- Electron blocking layers (EBLs)
 and MQWs (*see* Multiple quantum wells (MQWs))
 QSLs-EBL, 105, 110, 114, 116
- Electronic ballast for fluorescent lamps,
 587–588
- Electronic configurations of trivalent RE ions, 188
- Electronic control, 1143
- Elenbaas-Heller equation, 8
- Embedded-contact VTF (EC-VTF), 54
- Emission and excitation spectra, $\text{Y}_3\text{Al}_5\text{O}_{12}$, Ce^{3+} ,
 197–201
- Empire Machine Company, 19
- Enclosed fixture rating, 747
- End-of-life, 937, 944, 951
- End of life mechanisms, 1137
- Energy consumption, 940, 941, 945, 948,
 950, 951
- Energy efficiency, environmental impact,
 714–717
- Energy levels, organic molecule, 296
- Energy use. *See* Energy consumption
- Engineering tool software (ETS), 572
- EnOcean technology, 578
- Environmental impact
 categories, 944
 of light, 945
- Environmental zones, 994
- Environment stability, 205
- Equivalent noise model, 488, 489
- Euclidian distance, 791
- European Home Systems Protocol
 (EHS), 570
- Eutrophication, 944
- Everlight[®], 1151
- Excilamps, 1164–1169
- Exhibition lighting, 721–735
- Exposure limit values (ELV), 876
- External quantum efficiency (EQE), 79, 80,
 104, 203, 327
- Eye, 837
- F**
- Fabrication (of LED die), 947
- Favorable color (Bikou-shoku)
 Duv value, 909
 feeling of contrast index (FCI), 911
 general spectral radiation distribution, 908
 LED spectral radiation distribution,
 912–916
 preference index of skin color, 909
- Fechner's law, 771
- Feeling of contrast index (FCI), 911
- Femtocell, 687
- Field emission displays (FEDs), 184
- Filter circuits, 29
- Finite element analysis (FEA), 262, 265
- Finned heat sinks, 260–261
- Flashcube[™], 9
- Flicker, 447, 448, 450, 458, 460
- Flip-chip architecture (FCLED), 54
- Flip-Flash[™], 9

- Fluorescent lamps (FL), 6, 20, 33, 38, 840
 descriptions, 1069
 energy balance, 1071–1075
 fabrication technology, 1073–1074
 material limitations, 1073
 products, 1074–1075
 radiation mechanism, 1070–1071
- Fluorescent materials, 300–302
- Fluorescence/phosphorescence hybrid white
 OLEDs, 346–352
- Fluorescence white OLEDs
 EL spectrum, 337
 energy level diagram, 334
 PIN structure, 332
 two-stack tandem, 336
- Footcandles, 747
- Förster energy transfer, 297
- Forward phase dimming, 456
- Forward voltage method, 172
- Four integral module LED sources, 742
- Foveola, 761
- Franck Condon principle, 190
- Frenkel exciton, 325
- Free-standing GaN (FS-GaN) substrate, 96–105
- Fresnel lens, 476
- Full cut off luminaires, 997, 1003
- Full width at half maximum (FWHM), 82, 97,
 104, 109, 116
- Functional unit, 938, 939, 941, 953
- G**
- GaAs, 45
- GaAsP LEDs, 47
- Gallium nitride (GaN)
 FS-GaN, 96–105
 MQWs (*see* Multiple quantum wells
 (MQWs))
 Si substrate, 105–116
- Galvanic cell, 233
- GaN light-emitting diodes
 conventional GaN/sapphire LEDs,
 153–154, 158
 MOCVD-grown LED, 152–153
 vertical n-side-up GaN/mirror/Si LEDs,
 154–157
 wafer transfer application in, 150–151
- GaN/AlInGaN materials and devices, 49–52
- Ganglion cell layer, 762
- Gas fill, 1024–1027
- General electric lighting (GEL), 930
- General lighting service (GLS) lamp,
 1033–1037
- Genura[®], 1151
- German Osram, 930
- Glare control, 738
- Glare phenomena, 867
- Global lighting market segmentation, 928
- Global Positioning System (GPS), 668
- Global warming, 939, 944
- Goal and scope definition, 938–939
- Graded-composition electron blocking layer
 (GEBL), 91–95
- Graded-composition multiple quantum barriers
 (GQB), 86–91
- Graded-thickness multiple quantum wells
 (GQWs), 81–86
- Grassman's laws, 777, 778
- Grating acuity, 773
- Greenhouses, 1004
- Growth demand, 927
- Guidelines, 993
- H**
- Habitat conservation, 993
- Halogen lighting, 742, 744, 746
- Handover, 696–698
- Handshaking, 695
- Hard-scape landscape, 738
- Heat capacity, 244
- Heat dissipation, 241
- Heat transfer, 240
- Heterogeneous network (HetNet),
 688–689
- Hg lamps, 1106, 1112
- Hierarchical radial basis functions
 (HRBF), 544
- High intensity discharge (HID), 1004
- High-pressure mercury vapor lamp, 8
 basic working principle, 1080
 energy balance, 1091
 future research, 1093
 historical overview, 1080
 ignition and breakdown phase, 1085
 Lamptech website, 1093
 optically thick plasma radiation,
 1089–1091
 optically thin plasma radiation, 1089
 quartz technology, 1084
 radiation properties, 1087–1089
 stationary operation phase, 1085–1087
 technological aspects, 1081
 types of, 1080, 1082
 UHP lamps, 1092–1093
 warm-up phase, 1085

- High-pressure sodium-vapor lamps, 9, 16, 28, 36, 713, 1000, 1002, 1004
 electrical and thermal conductivities, 1100–1101
 electrode phenomena, 1101
 gas lantern used for street lighting, 1103
 historical overview, 1098
 with improved color rendering, 1102
 LEDs penetration in outdoor-illumination market, 1103
 photography of, 1098
 power balance, 1101
 principle, 1099
 pulsed current waveform, 1101
 radiation properties, 1099–1100
 technological aspects, 1099
 types, 1098–1099
- High-pressure xenon incandescent lamps, 1056
- High-pressure xenon lamps, 1108
- High reflection mirror, 152
- HomePlug, 573
- Horizontal handover, 696
- Horizontal illuminance level, 900
- Horticultural lighting, 704
 quantification and characterization, 707–710
 sources, 710–714
- HPS lamps. *See* High-pressure sodium-vapor lamps
- Human vision and perception
 absolute sensitivity, 768–770
 color psychophysics, 776–780
 discrimination thresholds, 770–772
 photoreceptor density and spectral sensitivity, 764–766
 receptive field organization, 766–768
 spatial vision, 772–774
 spectral sensitivities, 766
 temporal vision/critical flicker frequency, 774–776
 visual system, anatomy of, 758–764
- Humidity, 224, 225, 230, 231, 233
- Hygro-mechanical stress, 230, 232
- Hypothalamus, 832
- I**
- ICETRON® (ENDURA®), 1152
- IECC 2012 sensor requirements, 468
- IEC TR 62417–2
 hazard-related risk group labeling of lamp systems, 879–880
 labelling information and guidance, 880–881
 maximum permissible risk group of products, 882
- IEC TR 62778, 882–888
- IEEE 802.15.7, 681
- IESNA. *See* Illuminating Engineering Society of North America (IESNA)
- Illuminance, 518, 834, 839, 994
- Illuminating Engineering Society (IES), 794
- Illuminating Engineering Society of North America (IESNA), 234, 996
- Image sensor, 672–675, 678, 680, 681
- Immediate handover, 697
- Impact assessment, 939
- Incandescent lamps, 6, 32
 and energy levels, solid state, 1016
 blackbody, 1017–1018
 electronic properties and spectral emissivity, 1018–1020
 filament coiling and thermal management, 1027–1028
 history, 1015
 light bulb, 6
 gas fill atmosphere, 1024–1027
 general lighting applications, 1033–1037
 improvements in, 1053–1056
 life expectancy, 1057–1061
 optical emission and energy balance, 1028–1030
 operation, 1023
 radiative properties, 1018–1023
 sources, 1020–1021
 tungsten characteristics, thermal radiation source, 1021–1023
 tungsten metallurgy and filament production, 1030–1033
 special applications, 1045–1053
 technology, 13
- Incremental deployment algorithm, 542
- Indium tin oxide (ITO), 128, 140
- Individual fixture dimming, 750
- Indoor localization
 applications of, 675–680
 technologies, 666–667
 using visible light communication, 667–675
- Indoor navigation
 using visible light communication, 675–676
 for visually impaired, 675–677
- Inductively coupled discharges (ICDs), 1148
- Inductively coupled plasma (ICP), 106
- Industrial, Scientific and Medical (ISM) bands, 1144
- Infrared radiators, 1049–1051

- InGaN, 124, 139
Inside-out model, 687
Institution of Lighting Professionals (ILP), 994, 995
Integral reset algorithms, 625
Intelligent lighting, 456, 460–464
Intelligent street light control systems, 998
Interelectronic repulsion parameters, 195
Interface roughness (IRN), 99
Intergeniculate leaflets, 833
Interlayer, 135–137
Internal quantum efficiency (IQE), 55
Internal scattering layer, 369–371
International Commission on Illumination (CIE), 787–795
International Dark Sky Places, 998
International Electrotechnical Commission (IEC), 681
International Historical Mechanical Engineering Landmark, 19
International Space Station, 840
Internet, 599
Interpretation, 939
Inter-symbol interference (ISI), 650
Intraocular pressure (IOP), 759
Intrinsically photosensitive retinal ganglion cells, 832
Inventory analysis, 939, 947
IR filtering, light sensor, 519
Irradiance, 834, 835
 measurement, 886
Irreversible degradation, 205
ITO. *See* Indium tin oxide (ITO)
- J**
JEITA CP-1221, CP-1222, CP-1223, 680, 681
Judd's flattery index, 804
Junction heating effect, 172
Junction temperature, 224, 225, 227, 228, 230, 231, 234, 235
- K**
Kalman filter, 545
Kirkendall void, 232
KNX, 570–572
Kriging technique, 547
- L**
Lambertian distribution, LED, 744
Lamp design, 1143
Lamp efficacy, 1143
Lamp life, 1142
Lamps with outer coils, 1151
Landfill, 944, 946, 952
Landscape lighting
 glare control, 738
 hardscape, 738
 LED (*see* Light emitting diode (LED))
 power distribution, 739–741
 soft-scape, 738
Langevin type recombination, 325
Laser induced thermal imaging (LITI), 425
Laser lift-off (LLO) process, 151
Lateral geniculate nucleus (LGN), 763
Lateral inhibition, 767
LCA. *See* Life cycle assessment (LCA)
LCC. *See* Life cycle costing (LCC)
Lead-peak ballasts, 27
LED. *See* Light emitting diode (LED)
Legislation, 1004
Lens, 226, 230–232, 236, 837
Life cycle assessment (LCA)
 economic input output (EIO/LCA), 938
 hybrid, 938
 process, 938
 social, 940
Life cycle cost (LCC), 940
Life cycle, stages
 distribution, 937
 end-of-life, 951–952
 manufacturing, 937
 raw material acquisition, 937
 use, 937
Lifetime prediction, 234
Light, 830
 decay, 255–256
 dome, 1001
 extraction, 126–131, 137, 140, 141
 pollution, 991–1007
 quality (light spectrum), 706
 sensor architecture, 523–524
 sensor energy savings, 527
 sensor, field of view, 525
 sensor networking, 531
 sensor optical path, 525
 sensor selection considerations, 524–525
 shelf, 854
 therapy, 838–841
Light emitting diode (LED), 12, 17, 23, 31, 38, 42, 246, 269, 447, 834, 861, 1000
 backlight, 677
 chromatic performance, 281–285
 components cost, 42

- Light emitting diode (LED) (*cont.*)
- drivers, 591–593
 - efficacy, 449, 451
 - encapsulation lens, 285–287
 - evolution, 43
 - filters, 1000
 - first products of, 47
 - flexibility, 42
 - four integral module sources, 741
 - fundamentals, 447
 - gallium nitride related materials
(*see* Gallium nitride (GaN))
 - heat sources, 247–251
 - lamp selection, 743–746
 - light extraction efficiency, 270–275
 - lighting technology, 714
 - modules with landscape lighting fixtures,
749–752
 - optical modeling, phosphors, 279–281
 - packaging, 60–65
 - performance, continual improvement,
52–65
 - practical implementations of, 45
 - rapid adoption, 42
 - replacement lamps, 746–749
 - silicone, 255
 - source modeling, 275–279
 - temperature-dependent properties, 252
 - thermal management, 256–261
 - thermal resistance, 261–266
 - transient behavior, 256
 - visible III-V technology evolution, 46–49
- Lighting, 183, 830, 836, 838
- applications, 837
 - design, 836
- Lighting energy use intensity (EUI), 601
- Lighting-related greenhouse gas (GHG)
- emissions, 927
 - biological effects, 836–838
 - intensity, 834–836
 - measurement, 841–843
 - therapy, 838–841
- Light-sensitive species, 998
- Linear combination of atomic orbitals
(LCAO), 294
- Line-of-sight, 666, 669
- Lambertian transmission, 641
- Liquid phase epitaxy (LPE) growth, 46
- LM-80, 234, 235
- Local thermal equilibrium (LTE), 1108
- Location-based advertising, 676–677
- LonWorks, 569–570
- Louvers, 994
- Low-index grids (LIG), 371–372
- 6LoWPAN, 576–577
- Low pressure discharge (LPD) lamps, 1066
- Low pressure metalorganic chemical vapor
deposition (LP-MOCVD), 74, 75
- Low pressure sodium (LPS) lamps, 1001,
1003, 1075
- Low vision, 849
- Low voltage AC, 458–459
- Low wattage HID lamps, 1126
- LucaloxTM, 16
- Lumen maintenance, 1142
- Luminaire(s), 994, 997, 1004, 1006
- Luminaire level lighting controls (LLLC)
- commissioning, 598
 - definition, 597
 - demonstrations, 600–602
 - second generation, 599
- Luminance, 993, 1001
- values, 886
- Luminescence efficiency, 188
- Luminous
- efficacy, 1000
 - flux, 999, 1003
 - intensity model, 550
- Lutron sensor, 490, 494, 497, 498
- Lux, 834
- M**
- Macrocells, 687
- Macula, 761
- MagicubeTM, 9
- Magnetoreception, 974
- Manufacturing, 943, 946, 947, 949, 950, 953
- Mass flow controller (MFC), 76
- Master–slave/token-passing (MS/TP)
protocol, 567
- Measurements, 999, 1001
- light, 841–843
- Melanopsin, 832, 963
- Melatonin, 830, 834, 840
- suppression, 840
- Mercury, 941, 951
- Mercury-free high-pressure lamps, 1120
- Mercury-free illumination, 208
- Mesopic, 1000
- vision, 834
- Metal core-printed circuit board, 258
- Metal halide (MH) lamps, 9, 10, 16, 744,
1000, 1004
- resonant cavity lamps, 1159
 - technology, 11, 16

- Metal-organic chemical vapor deposition (MOCVD), 76, 88, 91, 93, 95, 108
- Metal sources, 433–435
- MH lamps. *See* Silica metal-halide lamps
- Microcontroller, 450, 451, 456, 460, 462, 463
- Micro lens approach, 367
- Microphonic noise, 489
- Microwave and high frequency resonant cavity (“Plasma”) lamps, 1154–1160
- Microwave discharge excilamps, 1169
- Milky Way, 992
- Mirror, 127, 141
- MOCVD-grown LED, 152
- Model accuracy analyses, 651–652
- Modulation transfer function (MTF), 773
- Molecular orbital model, 294–295
- Monochromatic light, 837
- Moonlight, 999, 1006
- Multimodal function, 536
- Multipath propagation, 511
- Multiple input multiple output (MIMO) model, 641–644
- Multiple quantum barriers (MQBs), 86
- Multiple quantum wells (MQWs), 128, 131, 140
 - GEBL, 91–95
 - QWB, 86–91
 - GQWs, 81–86
 - p-type MQWs, 74
 - wider InGaN quantum well, 76–81
- Multi-scale adaptive sampling algorithm (MSAS), 544
- Multi-scale surrogate model, 543
- Munsell color space, 786
- Munsell color system, 911
- Museum lighting, 722
 - artworks storage area, 734–735
 - characteristics of, 723
 - conservation and restoration workshops, 734
 - LEDs, 727
 - maintenance workshop, 734
- Myopia, 900
- N**
- Nano-silver paste, 230
- N,N*-bis-(1-naphthyl)-*N,N*-diphenyl-1,1-biphenyl-4,4-diamine, 300
- National Electric Lamp Company, 32
- Natural lithography method, 158
- Negative color mixing, 787
- Neon lamps, 7
- Nephelauxetic effect, 185, 194
- Networked lighting systems, 596
- Neurobehavioral regulation
 - light intensity, 834–836
 - light wave length, 830–832
 - phototransduction, 832
- Neuroendocrine system, 831
 - light intensity, 834–836
 - phototransduction, 832
- Night-lights, 852
- Nocturnal melatonin. *See* Melatonin
- Noise equivalent power (NEP), 488
- Non-phonon transition, 191
- Non-radiative electron–hole recombination, 254
- Normalized indices, 867
- Notch effect, 774
- Novak lamp, 10
- Nursing homes, 856
- O**
- Obtrusive light, 993, 994, 997, 1004
- Occupancy sensor, 468, 600
- Ohmic contact, 152
- OLED. *See* Organic light-emitting diode (OLED)
- Open automated demand response, 595–596
- Open-loop mode, 536
- Open protocol, 562
- Opsins, 962
- Optically inert anions, 185
- Optimal light source configurations, 548–551
- Organic light-emitting diode (OLED)
 - device performance, 418
 - displays, 418
 - manufacturing process lighting panels, 389–393
 - panel design considerations, 386–389
 - process challenges for white, 403–410
 - white device configurations, 393–403
 - yield and reliability, 410–411
- Organic semiconductors, 296–299
- Organic source configurations, 429–433
- Original equipment manufacturer, 37
- Oscillator, 832
- OSRAM, 931
- Outdoor rating, 747
- Ozone depletion, 939, 945

P

Pacific gas and electric (PG&E) emerging technology, 599

Passive infrared sensing, 469–470

Performance, 840

Personal exposure to artificial optical radiation, 876

Perspective-n-Point, 673

Phase-control, 28

Phase difference of arrival (PDOA), 670–671

Phase dimming, 459

Phase shift, 834

Philips lighting, 930

Phosphorescence white OLEDs, 337–346

Phosphors, 15

- activator, 189–197
- applications, 207–208
- classification for pc-WLEDs, 205–207
- configurational coordinate model, 186–188
- coating, 226, 228, 229
- conversion quantum efficiency, 254
- definition, 182
- host lattice, 185
- luminescence phenomena, 183–185
- packaging technology, 213–217
- position, 236
- red-emitting nitridoaluminate, 209–213
- requirements for LED-used, 197–205
- selection rule, 188–189
- terminology, 182–183

Photobiological safety

- definition, 866
- SSL products, 867–868
- standard assessment, 877–888

Photocell, 516

Photochemical ozone creation, 945

Photochemical retinal damage, 870–872

Photochromatic interval, 770

Photodiode, energy absorption, 520

Photometric SI units, 518

Photopic, 999, 1001

- illuminance values, light sensor, 521
- response, 517

Photonic crystals (PCs), 128, 137, 374

Photopsin, 962

Photoreception, 843

Photoreceptor, 762, 836, 841

Photosensitive syndroms, 869

Photosynthesis action spectrum, 1005

Photosynthetic system, 705

Photosystems, 705

Physical layer, 463

Phytochrome, 705, 962

- photosystem, 705

Pioneer commercialized the first passive-matrix OLED (PMOLED) displays, 385

Planar homography, 673

Planckian locus, 1002

Planckian radiator, 201, 202

Plant growth and development, influence of light, 704–707

Point spread function (PSF), 761

Popcorn noise, 469, 486, 487

Pose estimation, 673, 674

Pot-flyer machine, 22

Power cycling, 230

Power distribution plan, 739–741

Power efficiency (PE), 327

Power line communications, 455–458, 462

Preferred chromaticities, 804

Presbyopia, 760

Presence detection sensors, 998

Primaries, 777, 778, 787

Product category rules (PCR), 938

Proprietary protocol, 562

Purkinje shift, 1000

Pulse dual slope modulation (PDSM), 638

Pulse width modulation (PWM), 447, 448, 450

Pyroelectric detector, 481

Pyroelectric sensors, 480

Q

QL[®] lamp, 1150

Quality lighting, 850

Quantification and characterization, horticultural lighting, 707–710

Quantum confined Stark effect (QCSE), 84, 99

Quantum dot (QD) structures, 101

Quantum efficiency, 201

Quartz technology, 1084

Quasi experimental field studies, 902

Quaternary superlattices electron-blocking layers (QSLs-EBL), 105, 110, 114, 116

R

Radiated electromagnetic (EM) disturbance, 1144

Radiation, 224, 230, 234, 235

- pattern, 503

Radiation induced sublimation transfer (RIST), 425

Radioactivity, 184

Rapid start system, 26

- Raw material acquisition, 944
 Rayleigh scattering, 1001
 Reabsorption phenomenon, 201
 Received Signal Strength Indicator (RSSI), 667, 669, 670
 Recreation, 1006
 Recursive simulation method, 647
 Recycling, 950, 952
 Red-emitting nitridoaluminate phosphor
 crystal structure, 209
 electroluminescent spectra, 212
 excitation and emission spectra, 210
 in HF solution, 211
 Re-entrant cavity lamps, 1150–1151
 Reflector lamps, tungsten halogen, 1044
 Regulator ballasts, 27
 Relative scotoma, 763
 Remote device management (RDM), 565
 Remote-phosphor product, 213
 Residential lighting, 924
 Resonant cavity lamps, 1154–1160
 Resource depletion, 945
 Retinal pigment epithelium (RPE), 761–762, 870
 Retinohypothalamic tract, 832
 Retinotopic organization, 763
 Reuse, 937, 950, 952
 Reverse phase dimming, 456
 Rhodopsin, 962
 Robot control, 675, 678–681
 Root layer protocol (RLP), 566
 Rotary-indexing machines, 21
 RSSI. *See* Received Signal Strength Indicator (RSSI)
- S**
- Sand-blasting, 369
 Scattering, 510
 Sealing glass, 1128
 Seasonal affective disorder, 838
 Selection rule, 188–189
 Self-deployment methods, 541
 Self-heating, 235
 Semiconductor lasers, 61
 Sensitive population, 888
 Sensor(s)
 amplification stages, 481–483
 application, 497–501
 black body radiation, 470–471
 diffraction, 507–509
 driven luminaire schematic, 531
 IECC 2012 requirements, 468
 infrared radiation, 478–480
 integration times, 524
 lenses, 474–478
 lux steps, 523
 material behavior, 471–474
 node network, 541
 noise, 483–489
 occupancy sensor, 467
 passive infrared (PIR) motion, 469
 propagation parameters, 506–507
 reflection coefficient, 509
 responsivity, 480–481
 system response, 491–495
 temperature effects, 489–491
 testing, 492, 495–497
 ultrasonic occupancy sensing, 470
 Series resistance, 139
 Series switched PWM dimming, 452–454
 SiC, electroluminescence in, 44
 Silica metal-halide lamps, 1120
 physical principles, 1120
 technological aspects, 1115–1116
 types and applications, 1112–1114
 Silicone, 255
 encapsulant, 226, 228, 230–232
 Silicon and photopic response, compared, 518
 Silicon photosensor, 518
 Single input single output (SISO) model, 640–641
 Single lenslet, 478
 Single view geometry, 674
 Sintered-tungsten-filament lamps, 13
 Skyglow, 965
 Smart space testbed, 551
 Smoke-coating machine, 22
 SNVT_lux, 570
 Social life cycle assessment, 940
 Soda-lime glass, 13
 Sodium line, 967
 Soft-scape landscape, 738
 Solder, 230
 Solid-state HID ballasts, 30
 Space flight, 839
 Spatial CCT uniformity, 284–285
 Spatial summation, 772
 Speciated parameter adaptive differential evolution (SPADE) algorithm, 550
 Spectral power distribution (SPD), 792
 Spectrum control technology, 916
 Spectrum fidelity, 450, 452, 456
 Specular reflection, 509
 Spherical semivariogram model, 547
 SSL sources, maximum luminous efficacy, 59

Stage and studio lamps, 1047–1049
 Standardized protocols, 563
 Standard reflector lamps, 1036–1037
 Stark-Einstein law, 708
 Step-up transformer, 26
 Steven's power law, 772
 Stokes shift, 187
 Strain engineering, 132–139
 Substrate, production, 947
 Sulfur lamp, 713, 1159
 Surface of carborundum (SiC), 43
 electroluminescence, 43
 Surface plasmon (SP) mode, 373
 Sustainability, 940
 Sustainability assessment, total. *See* Total sustainability assessment
 Switching technique, 617
 System noise model, 485

T

Talbot brightness, 775
 TALQ, 578–579
 Tanabe-Sugano energy level diagrams, 190
 Tandem white OLEDs, 352–356
 Task lighting, 852–853
 Temperature, 224, 225, 227, 228, 230, 231, 233, 234
 gradient, 225, 226, 231, 235
 Temporal summation, 772
 Thermal degradation (TD) effect, 204
 Thermal energy
 generation, 240
 and temperature, 240
 Thermally activated delayed fluorescence (TADF) organic molecule, 308–312
 Thermally-induced luminescence, 183–185
 Thermal quenching, 203–205
 Thermal resistance, 229, 234, 235, 241–244
 Thermo-mechanical stress, 230, 232
 Thin-film-flip-chip architecture (TFFC), 54
 Thin film LED, 129, 133, 141–143
 Threading dislocation density (TDD), 74, 110, 115, 116
 Three-band phosphor system, 15
 Threshold versus intensity function (TVI function), 771
 Time difference of arrival (TDOA), 671–672
 TM-21, 234, 235
 Tokamak principle, 1148
 Top-emitting OLEDs, 366
 Toroidal lamps, 1151–1153
 Total light reflective effect, 157

Total sustainability assessment, 940
 Toxicity, 939, 945
 Traffic distribution, 695–696
 Translucent polycrystalline alumina, 16
 Transmission electron microscopy (TEM), 100
 Transmissivity, 473, 474
 TRIAC dimming, 456
 Triboluminescence, 183
 Tungsten filament lamps, Physics and technology, 1023–1033
 Tungsten halogen (TH) lamps, 10, 1037–1045
 Tungsten metallurgy and filament production, 1030–1033
 Two-phosphor PiG color converter, 214
 Two-way spreading loss, 504

U

Ultra-high-pressure (UHP) lamp, 1092–1093
 Ultrasonic field patterns, 502–504
 Ultrasonic sensing, 470, 501–502
 Ultrasonic source, 510
 Ultrasonic wave incident, 511
 Ultrasound positioning, 667
 Ultrasound wave diffraction, 508
 Ultra-violet ozone (UVO) treatment, 364
 Unified glare rating (UGR), 867
 Uniform sampling, 542
 Up-conversion mechanism, 326
 User Device (UD) 2., 684
 Utility function, 697
 UV filtering, light sensor, 519
 Ultra wide band (UWB), 667

V

Vacancy sensor, 468
 Vacuum thermal evaporation (VTE), 418–421
 equipment configurations, 426–429
 equipment maintenance, 438–439
 equipment productivity, 435–438
 Vanity lighting, 852
 Variable lighting scenarios, 902
 Varshni parameters, 253
 Veiling phenomenon, 867
 Vello machine, 20
 Vertical handover, 696
 Vertical lights, 853
 Vertical n-side-up nitride-based LEDs characteristics of, 169–175
 design of mirror structure for, 165
 surface texturing for, 165–169
 Vibronic transitions, 190

- Visible III-V LED technology, 46–49
 - Visible light beacon system, 681
 - Visible light communication (VLC), 562, 689–690
 - indoor optical wireless model, 640–645
 - LED, 636
 - model error analyses and calibration, optical model, 646–655
 - vs. radio frequency wireless communication, 636, 638
 - Vision, 834
 - Visual clarity, 805
 - Visual ergonomics, 730–731
 - Visually comfortable luminaire, 867
 - Visual semiotics, 732
 - Vitamin D₃, 860–861
 - Vitreous humor, 759
 - Vivid color, 916
- W**
- Wafer, manufacturing, 948
 - Wall box sensor, 496
 - Wall lens, 477
 - Waste, 945, 952
 - Water use, 945–946
- Wavelength division multiplexing (WDM) techniques, 689**
- Wavelength shift, 225
 - Weber's law, 771
 - White light-emitting polymer, 312
 - Window comparator, 484
 - Wireless home link (WHL), 636
 - Wireless load controllers, 594
 - Wireless local area networks (WLANs), 686
 - Wireless network system, 596
 - Works with Sora program, 747
 - World coordinate, 674, 675
- X**
- X10, 572–573
 - Xenon lamps, 1108
- Y**
- Yellowing, 231, 232
- Z**
- ZigBee, 574–576
 - Z-Wave, 577