NanoScience and Technology

Gadi Eisenstein
Dieter Bimberg *Editors*

Green Photonics and Electronics

NanoScience and Technology

Series editors

Phaedon Avouris, Yorktown Heights, USA
Bharat Bhushan, Columbus, USA
Dieter Bimberg, Berlin, Germany
Cun-Zheng Ning, Tempe, USA
Klaus von Klitzing, Stuttgart, Germany
Roland Wiesendanger, Hamburg, Germany

The series NanoScience and Technology is focused on the fascinating nano-world, mesoscopic physics, analysis with atomic resolution, nano and quantum-effect devices, nanomechanics and atomic-scale processes. All the basic aspects and technology-oriented developments in this emerging discipline are covered by comprehensive and timely books. The series constitutes a survey of the relevant special topics, which are presented by leading experts in the field. These books will appeal to researchers, engineers, and advanced students.

More information about this series at http://www.springer.com/series/3705

Gadi Eisenstein · Dieter Bimberg
Editors

Green Photonics and Electronics

Editors
Gadi Eisenstein
Departmenmt of Electriocal Engineeering and Russell Berrie Nanotechnology Institute
Technion – Israel Institute of Technology
Haifa
Israel

Dieter Bimberg
Center of NanoPhotonics
Technical University of Berlin
Berlin
Germany

ISSN 1434-4904 ISSN 2197-7127 (electronic)
NanoScience and Technology
ISBN 978-3-319-67001-0 ISBN 978-3-319-67002-7 (eBook)
DOI 10.1007/978-3-319-67002-7

Library of Congress Control Number: 2017951174

© Springer International Publishing AG 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. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Preface

The energy consumption associated with the Internet has become a major concern of the scientific and technological communities, the economic and political establishment as well as the media and the public at large. The vast increase in the Internet usage, driven by a continuous introduction of new applications, integration and reliance of public services and an increase in the number of users worldwide brought upon a situation where the ability of modern society to supply the required energy for the predicted consumption associated with future computing and communication came into question. While several apocalyptic predictions published in the early years of the third millennium will definitely not materialize, the problem is real and requires a solutions to which the technological community must dedicate itself.

Achieving an energy-efficient Internet requires a multi-facet solution that addresses all aspects of the complex data network. An important part of the solution will involve photonic devices and systems. The photonics technologies have traditionally not emphasized sufficiently energy issues but once it became clear that efficiency is paramount, large efforts aiming at reducing the power required by photonic devices, mainly lasers, have started. A new field called *Green Photonics* has emerged and quickly became popular.

However, photonic devices are driven by and feed electronic circuits, which are often energy inefficient and hence have to also be improved. Above and beyond this, the energy consumption is affected by the complete computing and communication systems. It is obvious, therefore, that what is needed are not photonic devices and systems that are *Green* but rather, the technological community should seek to develop a *Green Internet* where each part is as energy efficient as possible.

The development of a *Green Internet* requires a dialog and collaborations between experts of many fields including photonics, VLSI circuit design, computer architecture, networks, switching and information sciences. An additional aspect of the *Green Internet* is the use of renewable energy sources whenever possible.

The significance of holding this multidisciplinary dialog was recognized by the directors of the Russell Berrie Nanotechnology Institute at Technion – Israel Institute of Technology and the Center for Nano Photonics at the Technical

University of Berlin who joined forces and established a series of three annual *Green Photonics* symposia held alternatively in Haifa and Berlin between 2014 and 2016. These symposia which were funded by the Reinhart Frank Foundation brought together world famous experts in all fields related to the *Green Internet.* This book summarizes the three *Green Photonics* symposia highlighting the most important topics that were covered.

Optoelectronic semiconductor devices are addressed extensively. Energy-efficient VCSELs for the common 850 nm range as well as for 1550 nm are covered in two separate Chaps. 1 and 2, respectively. The advantages of quantum dots are highlighted in two chapters one, Chap. 3 dealing with optical amplifiers and the second, Chap. 4 with mode-locked lasers. Low-energy fast switching is described in Chap. 5 which addresses nonlinear photonic crystal waveguides. The electronic aspects of a *Green Internet* is covered in a few chapters. Low-energy logic design is presented in Chap. 6, while power management in the so-called Network on Chip devices is presented in Chap. 7. The higher level system aspects cover the topics of optimization of large interconnect networks is addressed in Chap. 8. Finally, we include two chapters that deal with renewable energy sources; Chap. 9 is an extensive survey of the global impact of photovoltaics and Chap. 10 deals with futuristic solar cells based on thin film organic semiconductors.

The diverse issues covered by the 10 chapters of the book highlight the need for an extensive multidisciplinary dialog and for collaborations between experts from different fields. This joint effort is needed to ensure a future *Green Internet*, which will enable all the applications needed by the modern society at an energy cost that is affordable.

Haifa, Israel — Gadi Eisenstein
Berlin, Germany — Dieter Bimberg

Contents

1 Energy-Efficient Vertical-Cavity Surface-Emitting Lasers for Optical Interconnects 1
Philip Moser, James A. Lott and Dieter Bimberg
1.1 VCSEL Energy Efficiency 1
1.2 Energy Efficiency Figures of Merit 2
1.3 Resonance Frequency and Modulation Bandwidth 4
1.4 Energy Efficiency Analysis 7
1.5 Energy Efficient Data Transmission Results 10
1.6 Summary 13
References 14

2 High-Speed InP-Based Long-Wavelength VCSELs 17
Silvia Spiga and Markus C. Amann
2.1 InP-Based VCSELs 18
2.1.1 Active Region 18
2.1.2 Hybrid-Cavity Concepts 19
2.1.3 Tunnel-Junction Laser 22
2.2 Single-Mode 1.55-μm Short-Cavity VCSELs 23
2.2.1 Hybrid Dielectric-Semiconductor VCSELs 24
2.2.2 Stationary Characteristics 25
2.2.3 Dynamic Characteristics 28
2.3 VCSEL Arrays and Advanced Modulation Formats 29
2.3.1 Data Communication 29
2.3.2 Telecommunication 31
2.4 Conclusion 32
References 33

3 Quantum-Dot Semiconductor Optical Amplifiers for Energy-Efficient Optical Communication 37
Holger Schmeckebier and Dieter Bimberg
3.1 Introduction 38
3.2 Basics of Quantum-Dot Semiconductor Optical Amplifiers 40
3.2.1 Parameters of SOAs 40
3.2.2 Dynamics of Conventional and QD SOAs 42
3.2.3 Design and Static Characteristics of QD SOAs 44
3.2.4 QD SOA Sample Series 45
3.3 Phase Modulation of QD SOAs 46
3.3.1 Introduction of the Concept 46
3.3.2 Prove of the Concept 47
3.4 Concept of Dual-Communication-Band Amplifiers 51
3.4.1 Introduction of the Concept 51
3.4.2 Proof of Concept 52
3.5 Signal Processing—Wavelength Conversion 56
3.5.1 Non-linearities of SOA Gain Media 57
3.5.2 Four-Wave Mixing in QD SOAs 58
3.5.3 Optimization of Static Four-Wave Mixing in QD SOAs 61
3.5.4 FWM of D(Q)PSK Signals 64
3.6 Summary 68
References 68

4 Quantum-Dot Mode-Locked Lasers: Sources for Tunable Optical and Electrical Pulse Combs 75
Dejan Arsenijević and Dieter Bimberg
4.1 Quantum-Dot Mode-Locked Lasers 75
4.1.1 Device Structures 76
4.1.2 Passive Mode-Locking 77
4.2 Jitter Reduction and Frequency Tuning 79
4.2.1 Hybrid Mode-Locking 81
4.2.2 Optical Injection 84
4.2.3 Optical Self-Feedback 86
4.3 Applications 91
4.3.1 Millimeter-Wave-Signal Generation 91
4.3.2 Optical Communication 96
4.4 Conclusion 101
References 101

5 Nanophotonic Approach to Energy-Efficient Ultra-Fast All-Optical Gates 107
Grégory Moille, Sylvain Combrié and Alfredo De Rossi
5.1 Introduction: A Case for All-Optical Signal Processing 107
5.2 Integrated All-Optical Gate 109

5.2.1 Technologies for Integrated On-Chip All-Optical Processing . . . 109
5.2.2 Energy-Efficient All-Optical Gates . . . 111
5.2.3 III–V Photonic Crystals Resonators . . . 113
5.3 Nonlinear Dynamics in PhC Resonators . . . 115
5.3.1 Microwatt Nonlinear Response . . . 115
5.3.2 Fast Optical Nonlinearities in Semiconductors . . . 116
5.3.3 Nonlocal Nonlinear Response of PhC Cavities . . . 118
5.4 PhC All-Optical Gate . . . 118
5.4.1 Photon Molecule . . . 121
5.4.2 The Role of the Carrier Lifetime . . . 121
5.4.3 InP . . . 123
5.4.4 P-Doped InP . . . 124
5.4.5 Passivated GaAs . . . 126
5.4.6 Integration with Silicon Photonics . . . 128
5.5 Application Example: All-Optical Signal Sampling . . . 129
5.5.1 All-Optical Sampling . . . 130
5.6 Conclusions . . . 133
References . . . 133

6 Alternative Logic Families for Energy-Efficient and High Performance Chip Design . . . 139
Itamar Levi and Alexander Fish
6.1 Introduction . . . 139
6.2 Background . . . 141
6.3 DML Basics . . . 155
6.4 DML Utilization for Increased E-D Flexibility . . . 158
6.5 Summary . . . 167
References . . . 167

7 Secure Power Management and Delivery Within Intelligent Power Networks on-Chip . . . 173
Inna Partin-Vaisband and Eby G. Friedman
7.1 Power Network on-Chip for Distributed Power Delivery and Management . . . 176
7.1.1 Concept of Power Network-on-Chip . . . 177
7.1.2 Power Network-on-Chip Architecture . . . 177
7.1.3 Challenges in Distributed Power Delivery . . . 180
7.2 Power Routing in SoCs . . . 181
7.2.1 Power Routers . . . 181
7.2.2 Locally Powered Loads . . . 182
7.2.3 Power Grid . . . 182
7.2.4 Case Study . . . 183
7.3 Stable Distributed Power Delivery Systems . . . 185

7.3.1 Experimental Evaluation of Stability Criterion 187
7.4 Secure Power Delivery and Management 191
7.5 Automated Design of Stable Power Delivery Systems 191
7.6 Summary 196
References. 198

8 Energy Efficient System Architectures 203
Avinoam Kolodny
8.1 Power Issues in Computing Systems 203
8.2 Characteristics of the Power Reduction Problem 204
8.2.1 The Disruption Principle 204
8.2.2 The Locality Principle. 205
8.2.3 The Challenge of Parallelism 208
8.2.4 A Unified Machine Model 210
8.2.5 Memory Intensive Systems 211
8.2.6 Applying the Principles in Large Data Center Computing 213
References. 213

9 Low-Cost Harvesting of Solar Energy: The Future of Global Photovoltaics 215
Eicke R. Weber, Stefan Glunz, H. -M. Henning, A. Palzer and R. Schindler
9.1 Introduction: The Needed Transformation of Our Energy System, Limited Fossil Fuels, the Climate Problem 215
9.2 The Role of Photovoltaics in Our Future Energy System, Based on Simulation of the German Energy System for 80% and More of Renewable Energy 219
9.3 Crystalline Silicon Photovoltaics. 228
9.3.1 Al-Back Surface Field Technology 230
9.3.2 Partial Rear Contact Technologies (PRC, PERC, PERL, PERT). 230
9.3.3 Solar Cells on n-type Si 233
9.3.4 Heterojunction Solar Cells. 233
9.3.5 Crystalline Si PV Beyond the Shockley-Queisser Limit . . 235
9.4 High-Concentration PV: CPV Technology 236
9.5 Thin Film PV Technologies 245
9.5.1 Thin Film Silicon Solar Cells 247
9.5.2 Cupper Indium Diselenide (CIS) Solar Cells. 249
9.5.3 Cadmium Telluride Solar Cells 250
9.6 The Future of PV: Further Market Development, PV Going into the Terawatt Range. 250
References. 255

10 Novel Thin-Film Photovoltaics—Status and Perspectives 263
Benjamin Oesen, Sascha Ullbrich, Johannes Widmer and Karl Leo
10.1 Introduction: Status of Photovoltaics in General 263
10.2 Organic Photovoltaics 266
10.2.1 Basics of Organic Photovoltaics 266
10.2.2 Thin Film Optics and Interference 271
10.2.3 Morphology of the Blend Layer 272
10.2.4 Optimized p-i-n Cells 274
10.2.5 Tandem and Multi-junction Cells—Maximizing the Power Output 275
10.3 Perovskite Photovoltaics 278
10.4 Application of Different Solar Cell Technologies 279
10.4.1 Application Scenarios 279
10.4.2 Energy Harvesting Under Real Application Conditions 281
10.5 Conclusion 285
References 285

Index 289

Contributors

Markus C. Amann Walter Schottky Institut, Technische Universität München, Munich, Germany

Dejan Arsenijević Department of Solid-State Physics, Technische Universität Berlin, Berlin, Germany

Dieter Bimberg Institut für Festkörperphysik, Technische Universität Berlin, Berlin, Germany; King Abdulaziz University Jeddah, Jeddah, Saudi Arabia

Sylvain Combrié Thales Research and Technology, Palaiseau, France

Alfredo De Rossi Thales Research and Technology, Palaiseau, France

Alexander Fish Faculty of Engineering Bldg, Bar-Ilan University, Ramat-Gan, Israel

Eby G. Friedman Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL, USA

S. Glunz ISE Fraunhofer Institute, Freiburg, Germany

H. -M. Henning ISE Fraunhofer Institute, Freiburg, Germany

Avinoam Kolodny Electrical Engineering department, Technion, Haifa, Israel

Karl Leo Dresden Integrated Center for Applied Physics and Photonic Materials (IAPP) and Institute for Applied Physics, TU Dresden, Dresden, Germany

Itamar Levi Faculty of Engineering Bldg, Bar-Ilan University, Ramat-Gan, Israel

James A. Lott Institut für Festkörperphysik, Technische Universität Berlin, Berlin, Germany

Grégory Moille Thales Research and Technology, Palaiseau, France

Philip Moser Institut für Festkörperphysik, Technische Universität Berlin, Berlin, Germany

Benjamin Oesen Dresden Integrated Center for Applied Physics and Photonic Materials (IAPP) and Institute for Applied Physics, TU Dresden, Dresden, Germany

A. Palzer ISE Fraunhofer Institute, Freiburg, Germany

Inna Partin-Vaisband Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL, USA

R. Schindler ISE Fraunhofer Institute, Freiburg, Germany

Holger Schmeckebier Department of Solid-State Physics, Technische Universität Berlin, Berlin, Germany

Silvia Spiga Walter Schottky Institut, Technische Universität München, Munich, Germany

Sascha Ullbrich Dresden Integrated Center for Applied Physics and Photonic Materials (IAPP) and Institute for Applied Physics, TU Dresden, Dresden, Germany

Eicke R. Weber ISE Fraunhofer Institute, Freiburg, Germany; Berkeley Education Alliance for Research in Singapore BEARS, Singapore, Singapore

Johannes Widmer Dresden Integrated Center for Applied Physics and Photonic Materials (IAPP) and Institute for Applied Physics, TU Dresden, Dresden, Germany

Chapter 1
Energy-Efficient Vertical-Cavity Surface-Emitting Lasers for Optical Interconnects

Philip Moser, James A. Lott and Dieter Bimberg

Abstract General rules that describe how to achieve extremely energy-efficient data transmission with oxide-confined VCSELs are derived, explained, and verified by data transmission experiments. We demonstrate that VCSELs with smaller oxide-aperture diameters are more energy-efficient than similar VCSELs with larger oxide-aperture diameters and introduce a new method for analyzing the suitability of different VCSELs for application in different optical interconnect configurations by introducing the modulation factor M. Applying the derived rules for energy-efficient VCSEL operation enables record energy-efficient data transmission with less than 100 femto-Joules per bit in a wide range of bit rates and multimode optical fiber lengths.

1.1 VCSEL Energy Efficiency

Energy-efficient operation with low dissipated and consumed energy per transmitted bit is a prerequisite for future optical interconnects. Vertical-cavity surface-emitting lasers (VCSELs) are a key component for optical computer interconnects. Although additional performance requirements such as long transmission distances across multimode optical fiber, high error-free bit rates, and operation at elevated temperatures over a wide range of temperatures exist, all these goals must be achieved at a simultaneously high energy efficiency.

In the fields of optical data transmission and semiconductor lasers the term "energy efficiency" is not a firmly set term with one established definition. For a complete optical interconnect the energy efficiency could for instance refer to the complete consumed or dissipated energy to transfer one unit of information from

P. Moser (✉) · J.A. Lott · D. Bimberg
Institut für Festkörperphysik, Technische Universität Berlin, Berlin, Germany
e-mail: p.moser@tu-berlin.de

J.A. Lott
e-mail: james.lott@tu-berlin.de

D. Bimberg
e-mail: bimberg@physik.tu-berlin.de

© Springer International Publishing AG 2017

G. Eisenstein and D. Bimberg (eds.), *Green Photonics and Electronics*,
NanoScience and Technology, DOI 10.1007/978-3-319-67002-7_1

the transmitter to the receiver and would then have the units of energy per bit. The true energy efficiency depends on the overall performance of the entire VCSEL-based optical communication system including for example the transmitter (that is the VCSEL and the VCSEL driver circuit), the receiver, the characteristics of the interconnecting optical medium, and the attributes of the particular modulation scheme employed to encode the information. In this Chapter we focus on the intrinsic energy efficiency of the VCSEL.

Large progress has been achieved in the past years in increasing the modulation bandwidth of oxide-confined VCSELs resulting in modulation bandwidths exceeding 25 GHz [1, 2]. Since the major concern of the past research was on increasing the modulation bandwidth and the error-free bit rate, these record results are achieved at large bias currents and large power dissipations and energy consumptions that exceed the practical limits for commercial applications in future novel optical interconnects. Even larger modulation bandwidths up to 37 GHz have been demonstrated with coupled-cavity VCSELs [3, 4]. The largest bit rate achieved with transversely coupled VCSELs is 25 Gb/s [5, 6], thus the large modulation bandwidths of these coupled-cavity devices does not yet fully translate into large bit rates. Thus at present, optically coupled VCSELs have not yet been demonstrated that achieve bit rates superior to conventional directly current-modulated VCSELs and the large modulation bandwidths therefore do not lead to energy-efficient data transmission with low consumed energy per bit.

1.2 Energy Efficiency Figures of Merit

In order to compare the suitability of VCSEL designs for low energy consuming optical interconnects we define the electrical energy-to-data ratio *EDR* (fJ/bit), and the dissipated heat-to-bit rate ratio HBR (expressed in units of mW/Tbps or fJ/bit) as [7]:

$$EDR = \frac{I \cdot V}{BR} = \frac{P_{el}}{BR} = \frac{I \cdot V}{M \cdot f_{3dB}} \tag{1.1}$$

$$HBR = \frac{I \cdot V - P_{opt}}{BR} = \frac{P_{el} - P_{opt}}{BR} = \frac{P_{diss}}{BR} = \frac{I \cdot V - P_{opt}}{M \cdot f_{3dB}} \tag{1.2}$$

where V and I are the direct current (DC) operating voltage and bias current of the VCSEL, $IV - P_{\mathrm{opt}}$ is the dissipated power P_{diss}, P_{opt} is the optical output power of the VCSEL, BR is the bit rate of error-free data transmission, and f_{3dB} is the modulation bandwidth. The modulation energy that in a typical laboratory test set up is delivered from a bit pattern generator and that finally generates the bits is not included in the *EDR* and *HBR* definitions given by 1.1 and 1.2. The term M in 1.1 and 1.2, where $BR = M \cdot f_{\mathrm{3dB}}$ is the *modulation factor* (M-factor) that relates the

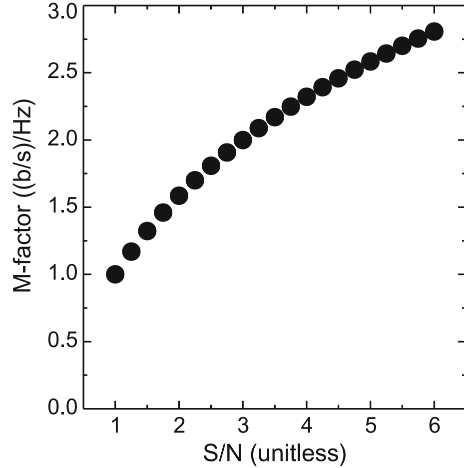


Fig. 1.1 Calculated modulation factor *M* for a digital communication system versus the signal-to-noise ratio S/N as determined from the Shannon-Hartley theorem [11]

intrinsic f_{3dB} to the error-free bit rate of a given optical interconnect technology [8]. We note that *M* simply replaces the term $log_2(1 + S/N)$ in the classic Shannon-Hartley Theorem [9] in the field of Information Theory. The maximum channel capacity *C* of a communication channel subject to additive white Gaussian noise (i.e. white thermal noise) and with an "*arbitrarily small frequency of errors*" (see [9], p. 43) is written as $C = B \cdot log_2(1 + S/N)$, where *B* is the bandwidth of the channel ($B = f_{3dB}$ in our present case; written as *W* by Shannon [9]), *S* (written as *P* by Shannon [9]) is the average signal power in Watts, and *N* is the average noise power in Watts. The channel capacity *C* is equal to the bit rate for our purposes since in our case the VCSELs are employed to transmit simple binary data using a two-level nonreturn-to-zero (NRZ) coding scheme and thus $C = BR$ is naturally given in units of bit/s. The modulation factor *M* (in units of (bit/s)/Hz) is then equivalent to the "spectral efficiency" given by $\eta = M = C/B$ [10]. We also note that $\eta = log_2(1 + M \cdot E_b/N_0)$, where E_b is the energy (in Joules) to transfer a bit and N_0 is the noise spectral density (also in Joules) [10]. Thus we may relate the energy efficiency to the spectral efficiency by writing $N_0/E_b = (2^M - 1)/M = SNR/M$, where the signal-to-noise ratio $SNR = S/N$. The calculated M-factor for a digital communication system is plotted versus the signal-to-noise ratio in Fig. 1.1.

Assuming that a constant minimum average noise power exists in a given information channel, then energy efficient data transmission, i.e. low energy consumption per transmitted bit, corresponds to a small average signal power and thus a small signal-to-noise ratio and consequently a small M-factor. Thus information can be transported very energy efficiently for systems with small M-factors only [11]. Therefore, if the M-factor is small, according to 1.1 and 1.2, the VCSEL needs to provide a large f_{3dB} at simultaneously low consumed energy and low dissipated power. In this Chapter we demonstrate that VCSELs with smaller oxide-aperture diameters and therefore smaller optical mode volumes achieve larger f_{3dB} at significantly lower consumed energy and dissipated power than similar VCSELs of the same epitaxial

design but with larger oxide-aperture diameters. We present general rules applicable to all oxide-confined VCSELs that describe how to operate such small oxide-aperture diameter VCSELs in order to achieve record low *EDR* and *HBR* values.

1.3 Resonance Frequency and Modulation Bandwidth

The relaxation resonance frequency f_r of a semiconductor laser increases with the square-root of the current above the threshold current I_{th} [12]. The rate at which f_r increases is given by the D-factor. In order to achieve a large relaxation resonance frequency f_r and thus a potentially large modulation bandwidth $f_{3\mathrm{dB}}$ at simultaneously low bias currents the D-factor is desired to be as large as possible.

$$f_r = D\sqrt{(I - I_{th})} \tag{1.3}$$

The D-factor of a semiconductor laser depends on the epitaxial design and the geometrical design of the device and is given by 1.4:

$$D = \frac{1}{2\pi}\sqrt{\frac{\eta_i \Gamma v_g}{qV_a}\frac{\partial g/\partial n}{\chi}} \tag{1.4}$$

where η_i (unitless) is the internal quantum efficiency, Γ (unitless) is the optical confinement factor, v_g (cm/s) is the group velocity of light, q is the elementary charge (coulomb), V_a (cm^3) is the active volume, $\partial g/\partial n$ (cm^2) is the differential gain, and χ (unitless) is the transport factor. For an oxide-confined VCSEL the active volume V_a is the product of the aperture area and the thickness of the active material, given for example by the cumulative quantum well thickness. The main factors that determine the D-factor are the differential gain and the active volume. While the differential gain is determined by the active region design and therefore cannot be changed for a given epitaxial design, the active volume V_a can easily be varied by changing the oxide-aperture diameter. The oxide-aperture restricts the area of the active material that is electrically pumped and therefore the area that can provide positive optical gain. In addition, the oxide-aperture may provide a lateral optical confinement due to the lateral refractive index step, depending on the placement of the oxide layer relative to the nodes and antinodes of the resonant optical field intensity for a given optical mode, and on the thickness of the oxide layer. Figure 1.2 shows the two-dimensional and one-dimensional optical field intensity distributions in an example VCSEL with a double-mesa structure for the fundamental LP01 lasing mode. The schematics show two oxide-aperture layers that simultaneously constrict the charge carrier flow in the device and provide lateral optical confinement.

The D-factor increases with decreasing oxide-aperture diameter and therefore VCSELs with smaller oxide-aperture diameters achieve larger relaxation resonance frequencies at a given bias current compared to neighbor VCSELs from the same area on the same wafer but with larger oxide-aperture diameters [13]. In order to

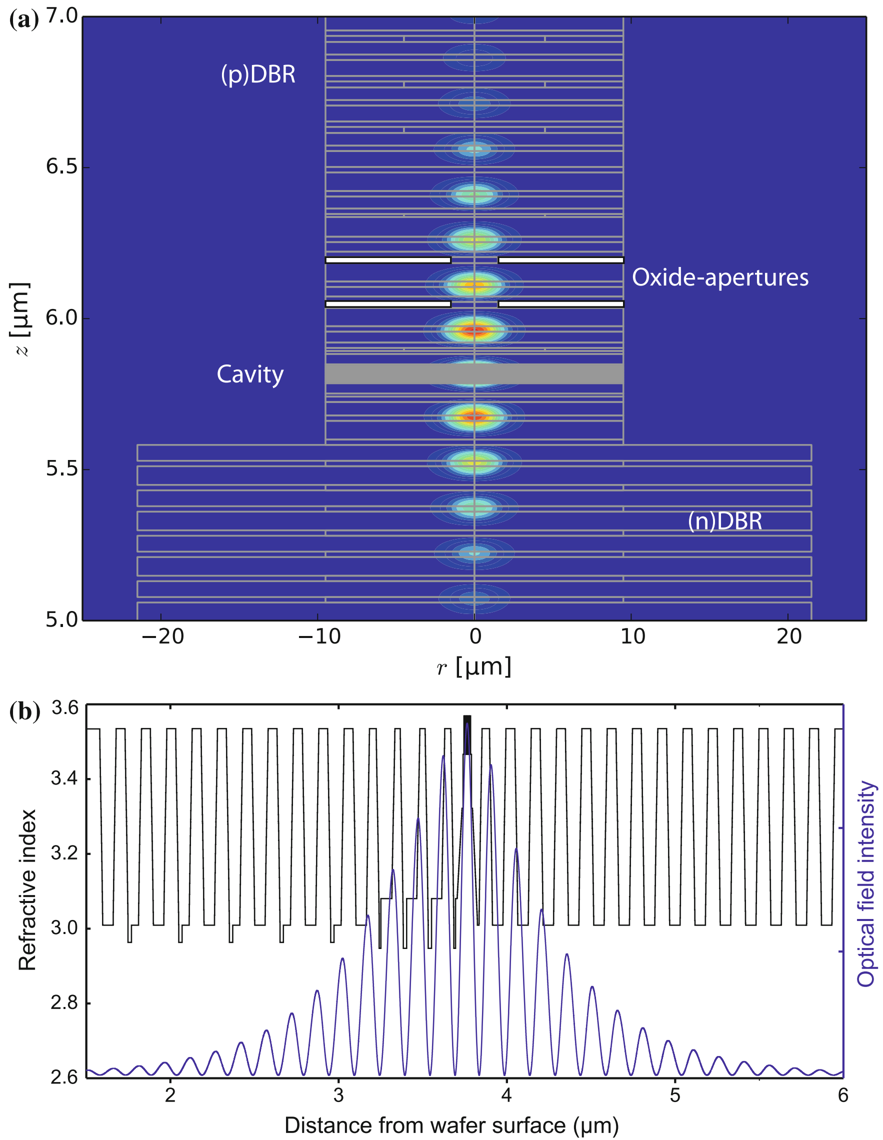


Fig. 1.2 Optical field intensity distribution in an example GaAs-based vertical-cavity surface-emitting laser with a double mesa structure in two dimensions (**a**) and in one-dimension (**b**) along a rotationally-symmetric center line where the top mesa radius r = 10 μm. Two oxide-apertures in the top p-doped DBR confine the light laterally

estimate the potential for error-free operation at high bit rates we use the modulation bandwidth f_{3dB} rather than the relaxation resonance frequency, because f_{3dB} is directly related to the maximum possible bit rate that the VCSEL is able to support.

The modulation bandwidth f_{3dB} is the bandwidth at which the small-signal modulation response is −3 dB. As with the relaxation resonance frequency f_r, the modulation bandwidth f_{3dB} increases with the square-root of the current above the threshold current, until thermal and other effects cause the modulation bandwidth to saturate. The proportionality factor for f_{3dB} that is equivalent to the D-factor is the modulation-current efficiency-factor *MCEF* [14] as shown in 1.5:

$$f_{3dB} = MCEF\ \sqrt{(I - I_{th})} \tag{1.5}$$

The *MCEF* cannot be analytically expressed by geometrical or epitaxial VCSEL design parameters but must be determined experimentally by small-signal modulation response measurements.

At bias currents smaller than the thermal rollover current where the relaxation resonance frequency f_r and the modulation bandwidth f_{3dB} both follow the linear theory given by 1.4 and 1.5, the modulation bandwidth is larger than the relaxation resonance frequency and we have:

$$f_{3dB} > f_r \Rightarrow MCEF > D \Rightarrow X = \frac{MCEF}{D} > 1 \tag{1.6}$$

$$f_{3dB} = X \cdot f_r \tag{1.7}$$

Since both the D-factor and the *MCEF* are constant values at lower forward bias currents before the saturation of f_r and f_{3dB} occurs the ratio X of the *MCEF* and the D-factor is a constant as well and can be used to calculate f_{3dB} from f_r. Therefore, a large D-factor also leads to a large f_{3dB} and thus to large bit rates at low bias currents above the threshold current. The dependence of the D-factor and *MCEF* on the oxide-aperture diameter is shown in Fig. 1.3 for neighbor VCSELs with

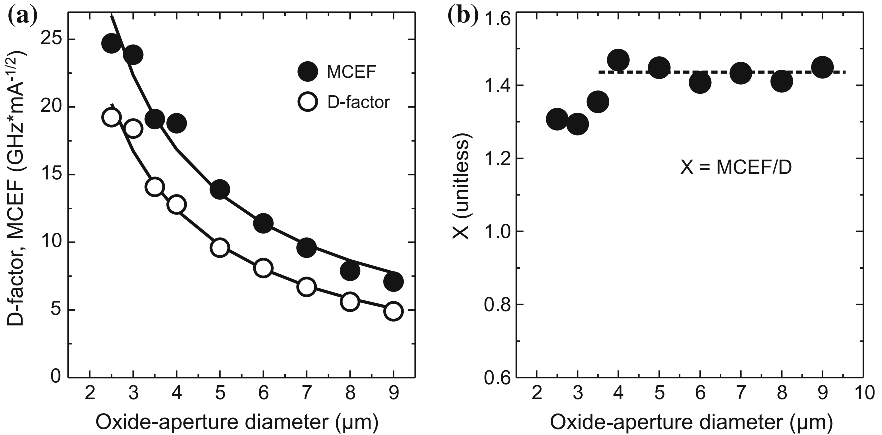


Fig. 1.3 **a** D-factor and modulation-current efficiency factor (*MCEF*) versus the oxide-aperture diameter. **b** Ratio $X = MCEF/D$ versus the oxide-aperture diameter [11]

oxide-aperture diameters ranging from 2.5 to 9.0 μm. For oxide-aperture diameters larger than approximately 4 μm, e.g. for typical multimode VCSELs, *X* does not vary with the diameter of the oxide-aperture. For smaller oxide-aperture diameters *X* slightly decreases with the oxide-aperture diameter [11] but *X* still remains above 1.2.

1.4 Energy Efficiency Analysis

In order to compare the potential of VCSELs for application in energy-efficient optical interconnects not only the absolute value of f_{3dB} but the ratio of the modulation bandwidth to the electrical and dissipated power is important. At a given bias current the electrical power increases with decreasing oxide-aperture diameter d_A because the differential resistance increases as d_A decreases, roughly by a factor of the inverse oxide-aperture radius squared (r^{-2}). For commercial applications of oxide-confined VCSELs the current density ***J*** is more important than the bias current, because ***J*** is directly related to the reliability of the VCSEL [15]. For a given current density the electrical power of a VCSEL decreases with decreasing oxide-aperture diameter. The modulation bandwidth f_{3dB} and the electrical power P_{el} are plotted versus the current density for two VCSELs with oxide-aperture diameters of 3 and 5 μm in Fig. 1.4. The VCSEL with the smaller oxide-aperture diameter reaches the maximum modulation bandwidth at a smaller current density and at a significantly smaller electrical power than the VCSEL of the same epitaxial design but with a larger oxide-aperture diameter. At low current densities as required for commercial data communication applications the modulation bandwidth of the smaller oxide-aperture diameter VCSELs exceeds that of the VCSEL with the larger oxide-aperture diameter. The VCSEL with the larger oxide-aperture

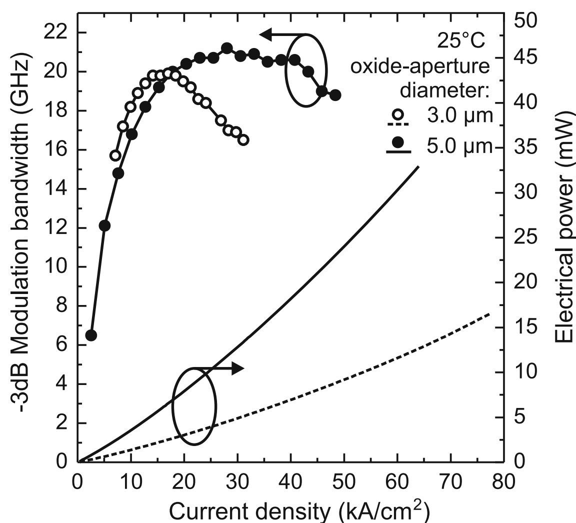


Fig. 1.4 Modulation bandwidth and electrical power versus the forward bias current density for 3 and 5 μm-diameter oxide aperture VCSELs [11, 16]

diameter achieves a larger maximum modulation bandwidth, but at a larger current density and at a significantly larger operating power as compared to the VCSEL with the smaller oxide-aperture diameter [11, 16]. Since the energy-efficiency of the VCSEL is at least as important as the modulation bandwidth, it is not a viable strategy to make use of the larger modulation bandwidth of the larger oxide-aperture diameter VCSEL. The 5 μm oxide-aperture diameter VCSEL of the given example has a 5% larger modulation bandwidth, but at the cost of an almost 6 times larger operating power [11, 16]. At a current density of approximately 17 kA/cm^2 both VCSELs achieve about the same modulation bandwidth f_{3dB}, but at a power of 2.5 and 6.6 mW for the VCSELs with oxide-aperture diameters of 3 and 5 μm, respectively [11, 16].

For energy-efficient data transmission the ratio of the modulation bandwidth and the driving electrical power is important. The bandwidth-to-electrical power ratio is given by the modulation bandwidth of a VCSEL at a given bias current divided by the electrical power. Figure 1.5 shows the bandwidth-to-electrical power ratio of VCSELs with different oxide-aperture diameters versus the current density [11]. At a given current density VCSELs with smaller oxide-aperture diameters have a larger modulation bandwidth-to-electrical power ratio than VCSELs of the same epitaxial design but with larger oxide-aperture diameters. The advantage of using VCSELs with a small oxide-aperture diameter is especially significant at small current densities (***J*** values) as small ***J*** values are required for reliable commercial applications.

As already illustrated in Fig. 1.4, the modulation bandwidth of a VCSEL reaches a maximum at a certain current density and a further increase in the bias current will eventually even lead to a decrease of f_{3dB}. The ratio of the modulation bandwidth to the electrical power consumption f_{3dB}/P_{el} decreases monotonically with increasing bias current, whereas f_{3dB} first increases, saturates and then decreases again with increasing bias current. This dependence of f_{3dB} limits the practical bias current range for application of the particular VCSEL as an energy-efficient light source for optical

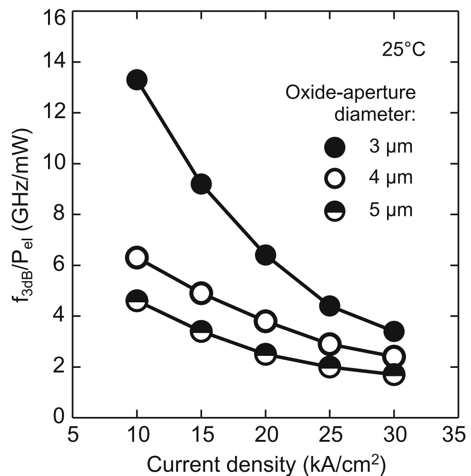


Fig. 1.5 Modulation bandwidth-to-electrical-power ratio versus the current density for VCSELs with oxide-aperture diameters of 3, 4, and 5 μm [11, 16]

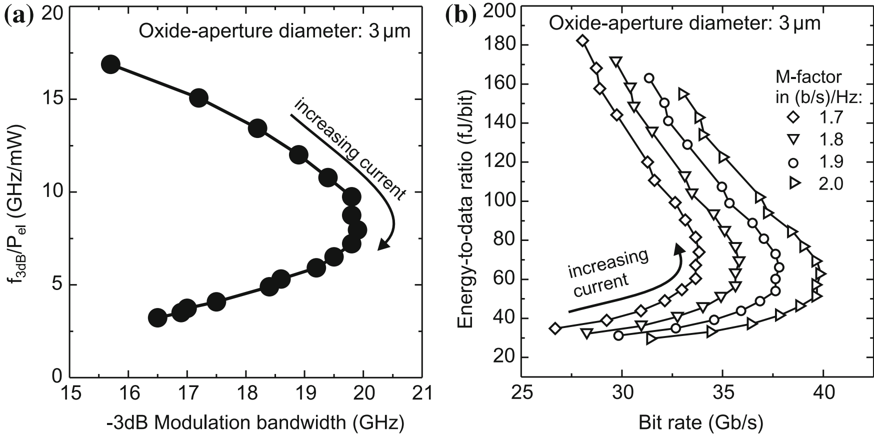


Fig. 1.6 Modulation bandwidth to electrical power ratio **a** versus the modulation bandwidth and the calculated energy-to-data ratio (*EDR*) versus the calculated bit rate **b** for M-factors of 1.7, 1.8, 1.9, and 2.0 (b/s)/Hz, respectively

interconnects. Thus ideally VCSELs must be operated at bias currents before the saturation of f_{3dB} sets in, in order to be energy-efficient and achieve a high bit rate.

The ratio f_{3dB}/P_{el} is plotted versus the modulation bandwidth f_{3dB} of a VCSEL with an oxide-aperture diameter of 3 μm in Fig. 1.6a. The electrical power P_{el} and the modulation bandwidth f_{3dB} both reflect purely intrinsic properties of the VCSEL without any impact of the measurement setup or configuration. That is why we refer to the bandwidth-to-electrical power ratio as the *intrinsic dynamic energy-efficiency*. From these measured intrinsic values of f_{3dB} and P_{el} we can calculate the bit rate and the *EDR* of a VCSEL for any given M-factor, assuming 1.1 is valid. The *EDR* is plotted in Fig. 1.6b versus the bit rate for M-factors of 1.7, 1.8, 1.9, and 2.0 (b/s)/Hz, respectively, calculated from the f_{3dB}/P_{el} ratio shown in Fig. 1.6a.

Given an M-factor that serves to characterize the performance of a complete optical interconnect system and the measured f_{3dB} and P_{el} characteristics for a given experimental or commercial VCSEL technology, a maximum anticipated bit rate ($BR_{max} = M \cdot f_{3dB}$) and a maximum anticipated energy efficiency ($EDR_{min} = P_{el}/BR_{max}$) may be estimated. Both *EDR* and the bit rate are calculated from the M-factor and therefore the maximum bit rates and *EDR* values at these bit rates change with different interconnect system configurations represented by the respective M-factors. In order to compare the potential of different VCSELs for use in different energy-efficient optical interconnects, i.e. for systems with different M-factors, the bit rates the VCSEL would achieve at a given *EDR* can be compared. Plotting these bit rates at a given fixed *EDR* value versus the M-factor allows us to compare the suitability of VCSELs for use in different optical interconnect systems [8, 11].

In Fig. 1.7 [11] we plot the bit rate at a constant *EDR* of 100 fJ/bit versus the M-factor for the two VCSELs with oxide-aperture diameters of 3 and 5 μm (i.e. for

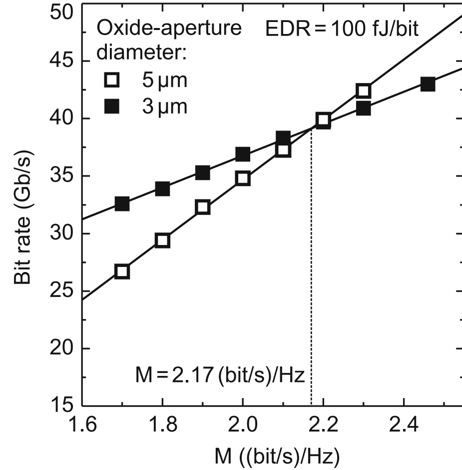


Fig. 1.7 Bit rate versus the M-factor at a calculated energy-to-data ratio (*EDR*) of 100 fJ/bit for two VCSELs with oxide-aperture diameters of 3 and 5 μm, respectively [11]

the same VCSELs as in Fig. 1.4). At an M-factor of 2.17 (bit/s)/Hz both VCSELs achieve exactly the same bit rate at the same *EDR*, whereas at smaller M-factors the 3 μm oxide-aperture diameter VCSEL achieves larger bit rates than the VCSEL with an oxide-aperture diameter of 5 μm. At M-factors larger than 2.17 (bit/s)/Hz the VCSEL with the larger oxide-aperture diameter achieves larger bit rates than the VCSEL with the smaller oxide-aperture diameter. For our large signal modulation setups that we use to perform data transmission experiments with VCSELs, we typically achieve M-factors between 1.4 and 2.0 (bit/s)/Hz. Typically, M-factors larger than 2.0 (bit/s)/Hz are not achieved with simple standard binary coding schemes, but with higher order modulation formats such as 4-PAM and 8-PAM [17], which require more complex and more energy-consuming interconnect configurations. Higher M-factors larger than 2.0 (bit/s)/Hz for the standard nonreturn-to-zero modulation coding scheme have been enabled by the use of driver and receiver circuits that employ equalization [18].

1.5 Energy Efficient Data Transmission Results

In order to verify the impact of the oxide-aperture diameter on the energy efficiency of the large-signal modulation performance of VCSELs, we measure VCSELs with an identical epitaxial design but with different oxide-aperture diameters using the exact same large-signal modulation measurement system configuration for each device [19]. Figure 1.8 shows bit error ratio (*BER*) curves versus the received optical power for VCSELs with different oxide-aperture diameters of an estimated 3.5, 4.0, and 5.0 μm operated in a back-to-back testing configuration at 25 Gb/s. The VCSELs are all measured with the same measurement setup configuration and the same modulation peak-to-peak voltage. The VCSELs are operated at the lowest

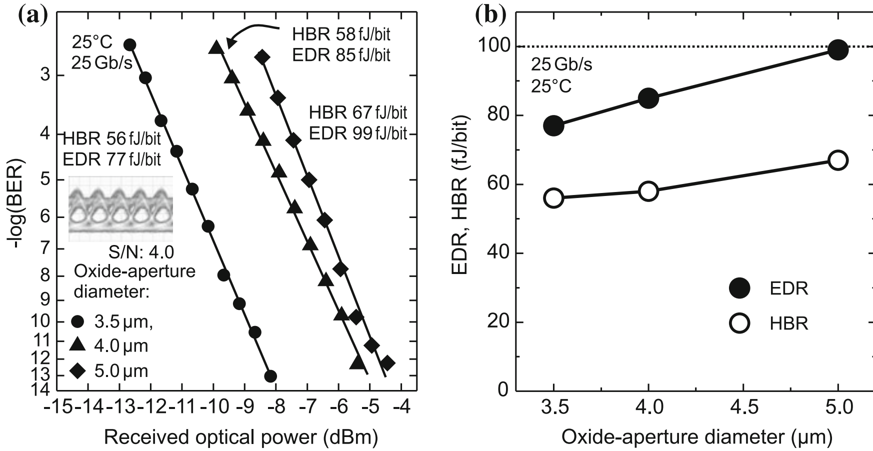


Fig. 1.8 **a** Bit error ratio (*BER*) versus received optical power at 25 Gb/s in a back-to-back testing configuration for VCSELs with oxide-aperture diameters of 3.5, 4.0, and 5.0 µm [17]. **b** Energy-to-data ratio (*EDR*) and heat-to-bit rate ratio (*HBR*) at error-free operation at 25 Gb/s in a back-to-back testing configuration versus the oxide-aperture diameter [11]

possible bias currents that still allow error-free operation, leading to the maximum energy efficiency, *HBR*, and *EDR* of these devices at this given bit rate and for this given measurement setup configuration. A record-low 56 fJ of dissipated energy per bit is demonstrated with the VCSEL having an oxide-aperture of 3.5 µm [19]. Both the *HBR* and *EDR* increase with increasing oxide-aperture diameter [19] as does the received optical power at error-free operation. The measured *EDR* and *HBR* values at 25 Gb/s are shown versus the oxide-aperture diameter in Fig. 1.8b.

The smaller mode-volume of smaller oxide-aperture diameter VCSELs enables error-free data transmission with lower operating power and lower dissipated energy per bit. Because small oxide-aperture diameters also lead to small spectral widths of the optical emissions, e.g. single-mode emission or only a few modes in the optical spectrum, such VCSELs are also well-suited to transmit data across long distances of multimode optical fiber at simultaneously high energy efficiency [20, 21]. At 25 Gb/s we have demonstrated energy-efficient error-free data transmission across up to 1 km of multimode optical fiber with an *HBR* of 100 fJ/bit [22]. The maximum possible fiber transmission distance decreases with increasing bit rate. It is for example, 500 m at 30 Gb/s at an *HBR* of 85 fJ/bit [22]. Using VCSELs with small oxide-aperture diameters it is also possible to achieve energy-efficient operation at relatively high bit rates such as at 40 Gb/s with an *HBR* of only 108 fJ/bit [23]. The measured *BER* versus the received optical power and the accompanying optical eye diagrams at error-free operation are shown in Fig. 1.9a and b, respectively.

Small mode volumes as a result of small oxide-aperture diameters enable extremely energy-efficient operation of VCSELs at low current densities, long multimode fiber distances, and high bit rates. This is due to the higher D-factors

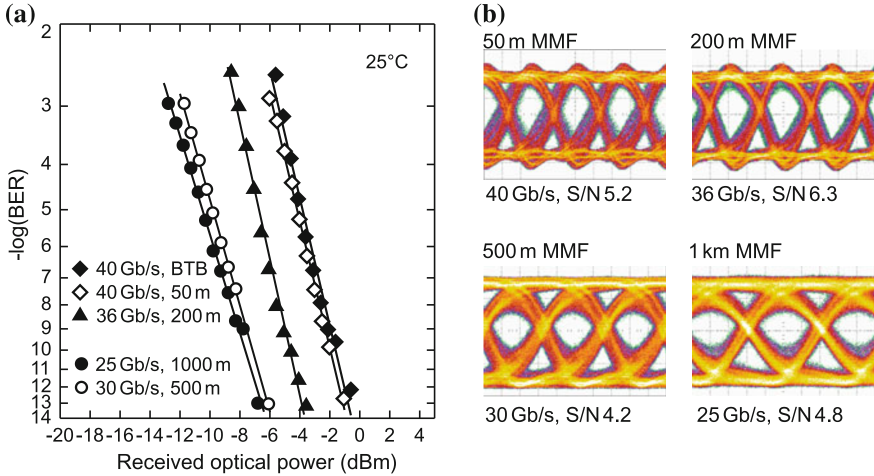


Fig. 1.9 **a** Bit error ratio (*BER*) versus received optical power for data transmission up to 40 Gb/s and across up to 1000 m of multimode optical fiber [11]. **b** Optical eye diagrams at the point of error-free detection for the *BER* curves shown in (**a**) [11]

resulting in larger modulation bandwidths at simultaneously low consumed energy and low operating power. For applications in future short-distance optical computer interconnects that are placed close to the integrated circuit processors, which are a main heat source inside a computer, the VCSELs need to perform best at elevated temperatures of 55 to 85 °C or at even higher temperatures. For a VCSEL that is optimized to operate at its highest bit rates at room temperature, the D-factor typically decreases with increasing external temperature which results in a significant decrease of the error-free bit rate and thus of the energy efficiency.

We have demonstrated that via a large detuning between the VCSEL's peak active region gain wavelength (estimated from the room temperature photoluminescence measurements of epitaxially-grown calibration wafers) and the VCSEL's cavity etalon wavelength, the dynamic performance of the VCSEL can be optimized for operation at elevated temperatures. For our 980-nm VCSELs with a large detuning of −15 nm [24] the D-factor increases monotonically with increasing ambient temperature [25]. The D-factor is plotted versus the ambient temperature in Fig. 1.10a. Due to the −15 nm detuning, the threshold current reaches a minimum value at a temperature of 60 °C. As a result, the energy efficiency of the VCSEL increases for a given bit rate with temperature, i.e. the consumed energy and the dissipated energy per bit both become smaller for elevated ambient temperatures between approximately 55–85 °C. The *EDR* and *HBR* for error-free operation at 38 Gb/s are plotted for our detuned 980 nm VCSEL with an oxide-aperture diameter of 5.5 μm versus the ambient temperature in Fig. 1.10b. In a wide temperature range from 55 to 85 °C the *HBR* is only 177 fJ/bit and highly insensitive to a temperature variation [26]. Thus the increasing D-factor as temperature increases successfully compensates for increased VCSEL thermal losses at higher temperatures.

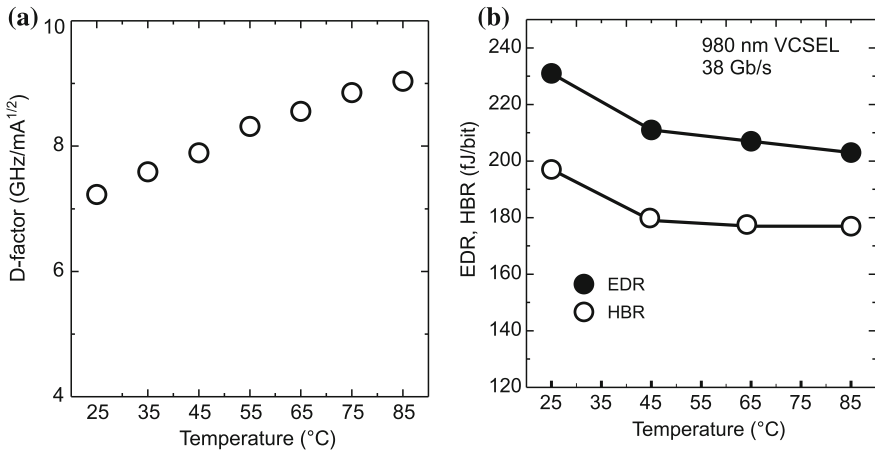


Fig. 1.10 **a** The D-factor versus heat-sink temperature for a 980 nm VCSEL with an oxide-aperture diameter of 5.5 μm and a room temperature peak photoluminescence wavelength-to-etalon wavelength detuning of −15 nm. **b** Energy-to-data ratio (*EDR*) and heat-to-bit rate ratio (*HBR*) at error-free operation at 38 Gb/s versus heat-sink temperature for a 980 nm VCSEL [26]

1.6 Summary

We have demonstrated that VCSELs with smaller oxide-aperture diameters are more energy efficient than similar VCSELs with larger oxide-aperture diameters. The reason is the smaller VCSELs have a larger D-factor and simultaneously they operate at a smaller energy consumption, a result of their smaller mode volume. Especially at low current densities that are required for reliable commercial applications, small oxide-aperture diameter VCSELs are superior to larger oxide-aperture diameter VCSELs with respect to energy efficiency and bandwidth. The small oxide-apertures also lead to a small spectral width of the optical emission allowing the transmission of data across longer distances of up to at least 1000 m of multimode optical fiber. Using VCSELs with small oxide-aperture diameters we have demonstrated that it is possible to transmit error-free data at up to 40 Gb/s across 50 m of multimode optical fiber and at up to 25 Gb/s across up to 1000 m of multimode optical fiber while dissipating less than or close to 100 fJ of heat energy per bit. The temperature dependence of the D-factor can be tuned via the photo-luminescence peak wavelength-to-cavity resonance wavelength detuning. A large detuning leads to an increase of the D-factor with temperature, resulting in a highly temperature insensitive energy efficiency by compensating for increasing thermal losses.

References

1. A. Larsson, M. Geen, J.S. Gustavsson, E.P. Haglund, A. Joel, P. Westbergh, E. Haglund, 30 GHz bandwidth 850 nm VCSEL with sub-100 fJ/bit energy dissipation at 25–50 Gbit/s. Electron Lett. **51**(14), 1096–1098 (2015)
2. E. Haglund, P. Westbergh, J. Gustavsson, E. Haglund, A. Larsson, High-speed VCSELs with strong confinement of optical fields and carriers. J. Lightw. Technol. **34**(2), 269–277 (2016)
3. S.T.M. Fryslie, M.P. Tan, D.F. Siriani, M.T. Johnson, K.D. Choquette, 37-GHz modulation via resonance tuning in single-mode coherent vertical-cavity laser arrays. IEEE Photonics Technol. Lett. **27**(4), 415–418 (2015)
4. S.T.M. Fryslie, and K.D. Choquette, Breakthroughs in photonics 2014: coherent vertical-cavity Surface-emitting laser arrays. IEEE Photonics J. **7**(3), 0700105-1–0700105-5 (2015)
5. H. Dalir, F. Koyama, 29 GHz directly modulated 980 nm vertical-cavity surface emitting lasers with bow-tie shape transverse coupled cavity. Appl. Phys. Lett. **103**(9), 091109 (2013)
6. F. Koyama, H. Dalir, Highly stable operations of transverse-coupled cavity VCSELs with enhanced modulation bandwidth. Electron Lett. **50**(11), 823–824 (2014)
7. P. Moser, W. Hofmann, P. Wolf, J.A. Lott, G. Larisch, A. Payusov, N.N. Ledentsov, D. Bimberg, 81 fJ/bit energy-to-data ratio of 850-nm vertical-cavity surface-emitting lasers for optical interconnects. Appl. Phys. Lett. **98**(23), 231106 (2011)
8. P. Moser, J.A. Lott, G. Larisch, D. Bimberg, Impact of the oxide-aperture diameter on the energy-efficiency, bandwidth, and temperature stability of 980 nm VCSELs. J. Lightw. Technol. **33**(4), 825–831 (2015)
9. C.E. Shannon, Communication in the presence of noise, in *Proceeding of IRE*, vol. 37, no. 1 (1949), pp. 10–21
10. F.R. Yu, X. Zhang, V. Leung, *Green Communications and Networking* (CRC Press, Florida, USA, 2012)
11. P. Moser, *Energy-Efficient VCSELs for Optical Interconnects* (Springer, Berlin, Heidelberg, 2016)
12. L.A. Coldren, S.W. Corzine, Dynamic effects, in *Diode Lasers and Photonic Integrated Circuits*, ed. by K. Chang (Wiley, New York, 1995), pp. 185–213
13. P. Moser, J.A. Lott, D. Bimberg, Energy efficiency of directly modulated oxide-confined high bit rate 850-nm VCSELs for optical interconnects. IEEE J. Sel. Top. Quantum Electron **19**(4), 1702212-1–1702212-12 (2013)
14. T.R. Chen, B. Zhao, L. Eng, Y.H. Zhuang, J. O'Brien, A. Yariv, Very high modulation efficiency of ultralow threshold current single quantum well InGaAs lasers. Electron Lett. **29**(17), 1525–1526 (1993)
15. B.M. Hawkins, R.A. Hawthorne, J.K. Guenter, J.A. Tatum, J.R. Biard, Reliability of various size oxide aperture VCSELs, in *Proceeding of IEEE 52nd Electronic Components and Technology Conference*, 28–31, San Diego, USA, CA (2002), pp. 540–550
16. P. Moser, P. Wolf, G. Larisch, H. Li, J.A. Lott, D. Bimberg, Energy-efficient oxide-confined high-speed VCSELs for optical interconnects, in *Proceeding of SPIE 9001, Vertical-Cavity Surface-Emitting Lasers XVIII*, 7–12 Feb 2014, San Francisco, CA, USA, p. 900103
17. D.M. Kuchta, A.V. Rylyakov, C.L. Schow, J.E. Proesel, C.W. Baks, P. Westbergh, J.S. Gustavsson, A. Larsson, A 50 Gb/s NRZ modulated 850 nm VCSEL transmitter operating error free to 90 °C. J. Lightw. Technol. **33**(4), 802–810 (2015)
18. K. Szczerba, P. Westbergh, M. Karlsson, P.A. Andrekson, A. Larsson, 70 Gbps 4-PAM and 56 Gbps 8-PAM using an 850 nm VCSEL. J. Lightw. Technol. **33**(7), 1395–1401 (2015)
19. P. Moser, J.A. Lott, P. Wolf, G. Larisch, H. Li, N.N. Ledentsov, D. Bimberg, 56 fJ dissipated energy per bit of oxide-confined 850-nm VCSELs operating at 25 Gb/s. Electron Lett. **48**(20), 1292–1294 (2012)
20. P. Moser, J.A. Lott, P. Wolf, G. Larisch, A.S. Payusov, N.N. Ledentsov, W. Hofmann, D. Bimberg, 99 fJ/(bit km) Energy to data-distance ratio at 17 Gb/s across 1 km of multimode

optical fiber with 850-nm single-mode VCSELs. IEEE Photonics Technol. Lett. **24**(1), 19–21 (2012)

21. P. Moser, J.A. Lott, P. Wolf, G. Larisch, A. Payusov, N. Ledentsov, D. Bimberg, Energy-efficient oxide-confined 850 nm VCSELs for long distance multimode fiber optical interconnects. IEEE J. Sel. Top. Quantum Electron 1292–1294 (2012)
22. P. Moser, J.A. Lott, P. Wolf, G. Larisch, H. Li, D. Bimberg, 85-fJ dissipated energy Per Bit at 30 Gb/s across 500-m multimode fiber using 850-nm VCSELs. IEEE Photon Technol. Lett. **25**(16), 1638–1641 (2013)
23. P. Wolf, P. Moser, G. Larisch, H. Li, J.A. Lott, D. Bimberg, Energy efficient 40 Gbit/s transmission with 850 nm VCSELs at 108 fJ/bit dissipated heat. Electron Lett. **49**(10), 666–667 (2013)
24. A. Mutig, *High Speed VCSELs for Optical Interconnects* (Springer, Berlin, Heidelberg, 2011)
25. H. Li, P. Wolf, P. Moser, G. Larisch, A. Mutig, J.A. Lott, D.H. Bimberg, Impact of the quantum well gain-to-cavity etalon wavelength offset on the high temperature performance of high bit rate 980-nm VCSELs. IEEE J. Quantum Electron **50**(8), 613–621 (2014)
26. H. Li, P. Wolf, P. Moser, G. Larisch, A. Mutig, J.A. Lott, D. Bimberg, Energy-efficient and temperature-stable oxide-confined 980 nm VCSELs operating error-free at 38 Gbit/s at 85 °C. Electron Lett. **50**(2), 103–105 (2014)

Chapter 2
High-Speed InP-Based Long-Wavelength VCSELs

Silvia Spiga and Markus C. Amann

Abstract The rapid growth of internet and cloud computing applications drives a huge demand for bandwidth capacity in communication networks, while power consumption, cost, and space density must scale down. This growth leads to an increase in the size of data centers (longer optical links), and of the fibers' channel data rate, rooted in Moore's Law. Until now, multi-mode fibers (MMF) have been largely employed in datacom applications due to the large coupling tolerance. However, the data-carrying capability of MMF decreases with the transmission distance due to pulse broadening resulting from modal and chromatic dispersion. In order to overcome those limits, transceivers based on single mode fiber (SMF) are under development and the first systems are on the market. Vertical-cavity surface-emitting lasers (VCSELs) are the transmitters of choice for short-reach applications due to their low cost, energy efficiency, and small footprint. InP-based VCSELs emitting at long wavelengths (i.e. 1.3 and 1.55 μm) have gained large interest due to their intrinsic lower power consumption (lower band gap) and low losses in silicon waveguides and silica-based optical fibers, which allows longer transmission distances. While short-wavelength GaAs-based VCSELs have achieved small-signal modulation bandwidths up to 30 GHz [1], InP-based VCSELs show inferior modulation capabilities [2, 3]. Up to date, the highest small-signal bandwidth demonstrated on InP-based devices is 22 GHz [3]. The distributed Bragg reflectors (DBRs) commonly used for GaAs-based VCSELs are made of binary and ternary semiconductor compounds, which offer several advantages such as high refractive-index contrast between the layers, good electrical conductivity and low thermal resistivity. The inferiority of semiconductor DBRs lattice matched to InP challenges the modulation bandwidth enhancement of InP-based devices which suffer of poor thermal conductivity, and high lateral spreading resistance. A further challenge is the single-mode laser operation that has motivated the transition from MMF to SMF in datacom systems. In this chapter, the

S. Spiga · M.C. Amann (✉)
Walter Schottky Institut, Technische Universität München, Munich, Germany
e-mail: amann@wsi.tum.de

S. Spiga
e-mail: Silvia.Spiga@wsi.tum.de

© Springer International Publishing AG 2017

G. Eisenstein and D. Bimberg (eds.), *Green Photonics and Electronics*,
NanoScience and Technology, DOI 10.1007/978-3-319-67002-7_2

challenges related to InP-based VCSELs are discussed with focus on active region design, cavity engineering, and current and optical confinement. These arguments apply to all InP-based VCSELs with emission wavelength between 1.3 and 2.0 μm. Stationary and dynamic characteristics are presented for a 1.55 μm VCSEL. Finally, datacom and telecom transmission experiments are presented.

2.1 InP-Based VCSELs

The absorption spectrum of silica-based optical fibers presents two main windows for datacom and telecom with low absorption around 1.3 and 1.55 μm. GaAs-based VCSELs shows excellent stationary and dynamic performances on the short wavelength regime (i.e. 850 and 980 nm) and take full advantage of their well-established technology. Wavelengths up to 1.3 μm have been achieved for GaAs-based VCSELs by means of highly strained GaInAs quantum wells (QWs) [4], GaInNAs QWs [5], and InAs-GaInAs quantum dots [6]. InP-based VCSELs offer, on the other hand, the advantage of growing quantum wells consisting of moderately (compressively) strained quaternary-materials such as AlGaInAs and InGaAsP emitting at wavelengths between 1.3 and 2.0 μm fully covering the optical fibers low-loss windows.

The key rules for designing a high-speed energy-efficient VCSEL are well known. First, the laser active region has to be able of properly confining the carrier and providing high differential gain. Second, the laser's cavity has to be as short as possible to reduce the photon lifetime boosting the laser relaxation resonance frequency and decreasing the intrinsic damping. Finally, current and optical field have to be properly confined. This allows electrical excitation of the optical mode that contributes to carry the transmitted information, and in guiding the relative longitudinal mode in the laser's cavity such that optical losses are minimal.

The active region composition, the cavity design concepts as wells as electrical and optical confinement to achieve high-speed energy-efficient InP-based VCSELs will be discussed in this section.

2.1.1 Active Region

The design of the active region plays a crucial role for stationary and dynamic performances of VCSELs. Different material systems have been proposed and demonstrated. QWs are used in VCSELs to produce a 2-D density of state (DOS) [7], and to allow high strain (ε) [8]. The step-like DOS of QWs enhances the differential gain. The strain reduces the valence-band DOS enabling population inversion to be reached with smaller carrier concentration, and has the effect of decreasing linewidth, chirp and the threshold current, and increasing slope efficiency, output power, and differential gain [9].

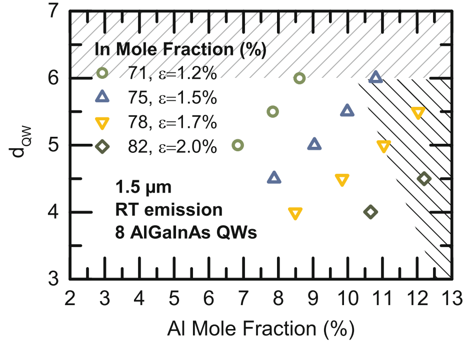


Fig. 2.1 Required QW thickness versus Aluminium mole fraction in the $Al_xGa_yIn_{1-x-y}As$ material system with 7-nm thick $Al_{0.274}Ga_{0.203}In_{0.523}As$ barriers in an active region designed for 1.5 μm emission at room temperature. The shaded area is not accesible for an active region with 8 QWs (see text)

QWs based on quaternary materials offer the advantage of freely designing the emission wavelength and the strain. For the quaternary material system AlGaInAs, Aluminum is used to increase the energy gap while Indium increases the compressive strain in the active region or decreases the energy gap. Unfortunately, a high Aluminum content leads to alloy scattering and lowers the carrier confinement in the active region. On the other hand, an increasing Indium content requires a reduction of the growth temperature and a higher V/III ratio. This means that only wave lengths between 1.3 and 2.2 μm can be achieved with relative high gain and a low photoluminescence linewidth [10].

The quaternary material system AlGaInAs lattice matched to InP for 1.5 μm emission wavelength provides good electron confinement in the active region thanks to its high $\Delta E_c/\Delta E_g$ ratio, compared with the almost equivalent material system GaInAsP. Furthermore, comprising only one group-V element, AlGaInAs grown with solid source MBE enables sharp interfaces and gives the possibility of implementing very high strain in the active region (up to 2%). The required thickness of an 1.5-μm AlGaInAs QW with 7-nm thick $Al_{0.274}Ga_{0.203}In_{0.523}As$ barrier is plotted in Fig. 2.1 versus the Aluminium mole fraction and for compressive strain ranging from 1.2 to 2.0%, generated by increasing the Indium content. Assuming an active region with 8 QWs to enhance the differential gain [11], the lateral shaded area represents thickness-strain combinations which exceed the dislocation-free critical thickness [12]. Under the same assumption, the upper shaded area would lead to an active region thicker than one quarter of the emission wavelength, which means a negligible increase of confinement factor at the cost of high threshold current density. The thickness of the QWs can be optimized within these boundaries.

2.1.2 *Hybrid-Cavity Concepts*

A common approach used to boost the small-signal modulation bandwidth of semiconductor lasers is the reduction of the photon lifetime. On one hand, the

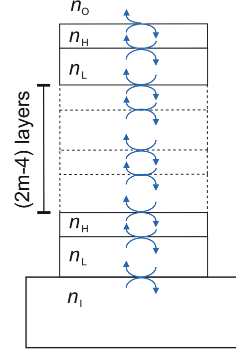


Fig. 2.2 Schematic of a DBR with quarter-wave stacks of layers with pairwise varying refractive index

relaxation resonance frequency of VCSELs is inversely proportional to the square root of the photon lifetime. On the other hand, the damping increases proportionally to the square of the relaxation resonance frequency, and the slope of its growth is proportional to the photon lifetime. The different approaches used to decrease the photon lifetime are the increase of mirror losses [13, 14] or the reduction of the effective cavity length [2, 15]. The last approach is the subject of this subsection. A hybrid dielectric/semiconductor cavity is introduced where a semiconductor cavity is combine with dielectric DBRs to achieve a strong confinement of the field in the active region.

DBRs are quarter-wave stacks of layers with alternating refractive index, as shown schematically in Fig. 2.2. The light is incident from a medium with refractive index n_I and penetrates in a DBR made of m layers with alternating low and high refractive indexes n_L and n_H. The light is than outcoupled to a medium with an index n_O. The *phase penetration depth* is defined as the depth at which the optical field appears to be reflected by a constant-phase mirror. It is common practice to define the *energy penetration depth* as the depth at which the energy density falls to $1/e$ of its initial value. These definitions give similar results for a large number of mirror pairs or for an outcoupling interface with high refractive index ratio. The reflection delay time τ_m of a DBR with $m/2$ pairs at the Bragg frequency f_B is given from the exact relation [16]:

$$\tau_m = \frac{1}{2f_B} \cdot \frac{q}{1-p} \cdot \frac{(1-a^2p^{m-1})(1-p^m)}{(1-q^2a^2p^{2m-2})} \tag{2.1}$$

where

$$q = \frac{n_\mathrm{L}}{n_\mathrm{I}}, p = \frac{n_\mathrm{L}}{n_\mathrm{H}}, a = \frac{n_\mathrm{O}}{n_\mathrm{H}}. \tag{2.2}$$

The optical phase penetration depth L_τ is defined as $nL_\tau = c\tau_m/2$ where n is assumed, for simplicity, to be a common refractive index of the cavity. The optical phase penetration depth is plotted in Fig. 2.3 as a function of the difference between

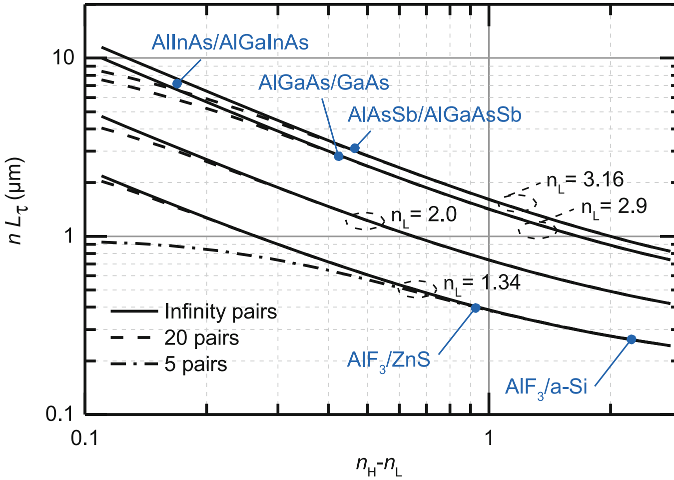


Fig. 2.3 Optical phase penetration depth in a DBR as a function of the difference between the refractive indexes of the two mirror's material for different n_L and for fixed values $n_I = 3.16$ and $n_O = 1.00$ at the Bragg wavelength of 1.55 μm for infinite number of layers (*solid line*), 20 pairs (*dash line*) and 5 pairs (*dash-dot line*). Commonly used material' pairs for DBRs are shown

the refractive indexes of the two mirror's materials and for different refractive indexes n_L. This plot refers to DBRs placed on top of InP and outcoupling to air, and designed at the Bragg wavelength of 1.55 μm. It gives a nice overview of the penetration depth linked to several DBR' material systems. Properly chosen dielectric DBRs, such as AlF_3/Si, have optical penetration depths that are one order of magnitude shorter than semiconductor ones, such as AlInAs/AlGaInAs.

The DBRs commonly used for GaAs-based VCSELs consist of binary and ternary semiconductor compounds such as AlAs and AlGaAs alloys. They offer several advantages such as high refractive-index contrast between the layers and low thermal resistivity. Semiconductor DBRs for InP-based devices are instead based on ternary and quaternary compounds such as AlInAs and AlGaInAs alloys, which have low refractive-index contrast that is long propagation depth, and high thermal resistivity. The latter is due to phonon-phonon interactions and phonon scattering with lattice defects, and increases with the number of foreign atoms added to a host lattice [10]. AlGaAsSb/AlAsSb DBRs latticed-matched to InP offer a higher refractive-index contrast but have poor electrical and thermal conductivities [17]. Different approaches such as InP/air-gaps [18] or dielectric DBRs [2] have been explored to shorten the propagation length of the field in the mirror, but the low thermal conductivity of those DBRs challenges the VCSEL's stationary and dynamic performance. In order to overcome this limitation, wafer-fused AlGaAs/GaAs DBRs have been demonstrated to combine high-refractive index contrast and low thermal resistance [19].

An alternative to DBRs are high-contrast gratings (HCGs). A HCG is a near-wavelength grating made of a high refractive-index material and surrounded

by a low-index material. 1.55-μm InP-based HCG VCSELs feature broadband and high reflectivity for surface-normal incident light with field polarization that can be designed to be either parallel or orthogonal to the elongated elements of the grating [20].

2.1.3 Tunnel-Junction Laser

For energy-efficient and single-mode operation, current and optical field have to be confined such that the fundamental optical mode is electrically excited, and optical losses are minimized. Both current and optical confinement can be achieved in VCSELs by means of current apertures realized by structured buried tunnel junctions (BTJ).

For GaAs-based VCSELs, a current aperture is typically achieved through radial *selective oxidation* of $Al_xGa_{1-x}As$ layers where x is near to 0.98 [21]. The oxidation rate changes by more than two orders of magnitude by varying x from 0.8 to 1 [22]. The large lattice constant mismatch between AlGaAs alloy and InP prevents the integration of this material system on InP-based long wavelength devices. The related material system $Al_xIn_{1-x}As$ is lattice matched to InP for $x = 0.5$, leading to a low oxidation rate. Furthermore, the AlInAs native oxide differs from that reported for AlGaAs due to the composite structure of the oxide related to the low Al content [23]. Another technique used to achieve current confinement in GaAs- and InP-based devices is *ion implantation*, which produces highly resistive regions in the semiconductor cavity. These VCSELs present the advantage of a planar device geometry. However, the lack of inherent optical confinement can lead to varying threshold and modulation limitations [24].

For InP-based VCSELs, tunnel junctions consisting of p^+-AlGaInAs and n^+-GaInAs have been demonstrated [25]. Current and optical confinement are achieved by etching the n^+-layer of the tunnel junction in circular-shaped areas and re-growing with n-InP to match the desired cavity length as shown in Fig. 2.4a. The etched area acts as a reverse biased pn-junction enabling current confinement to the *not* etched area. The band diagrams for the tunnel and blocking junctions are plotted in Fig. 2.4b and c, respectively. This method offers the great advantage of defining the aperture by simple optical lithography but raises the challenges related to re-grow of the structured junction. Re-growing with solid source MBE will *not* flatten the etched structure, which is reproduced at the upper mirror/semiconductor interface. The effect will be an effective fiber-like waveguiding with refractive index contrast Δn_{eff} depending on the height of the overgrown tunnel junction (ΔL):

$$\frac{\Delta n_{eff}}{n_{eff}} = \frac{\Delta L}{L}, \tag{2.3}$$

where L is the cavity length and n_{eff} is the effective refractive index [26]. Corresponding to the diameter of the fiber's core, the diameter of the tunnel junction will

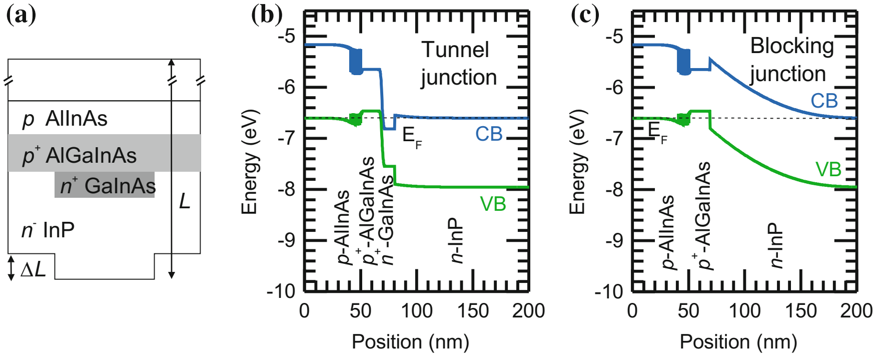


Fig. 2.4 **a** Schematic cross-section of a BTJ. Simulated band diagram **b** of a p$^+$-AlGaInAs/n$^+$-GaInAs tunnel junction, and **c** of a reverse-biased p$^+$-AlGaInAs/n-InP diode

determine how many transverse modes can propagate in the cavity. Single mode operation and high output power can be achieved by choosing this diameter sufficiently small as will be shown in the next section.

One of the main advantages of a tunnel junction is the reduction of the series resistance of VCSELs. The active region of long-wavelength VCSELs is typically embedded between an *n*-InP and a *p*-AlInAs, which, due to their band offset, offer effective carrier confinement to the active region. However, the poor electrical, thermal and optical properties of AlInAs motivate the need for reducing its thickness. This can be achieved by means of a low-resistive tunnel junction allowing to reduce the thickness of the AlInAs layer to less than one quarter of the emission wavelength, and to match the desired total cavity length using low-resistive *n*-InP instead.

2.2 Single-Mode 1.55-µm Short-Cavity VCSELs

Short-reach communication links and networks can employ 850-nm or 980-nm VCSELs, which emit typically in multiple transverse modes. However, signals carried by fundamental transverse modes with narrow linewidths can be transmitted over longer distances by limiting the chromatic dispersion of the fiber. For this reason, access and metropolitan networks are established on single-mode fiber motivating the need for single-mode 1.3- and 1.55-µm VCSELs.

The historical growth of the small-signal modulation bandwidth of short- and long-wavelength VCSELs is shown in Fig. 2.5. While short-wavelength devices have achieved bandwidths up to 30 GHz [1] by 2015, long-wavelength VCSELs show inferior modulation capabilities. The technological issues related to the bandwidth scaling of InP-based devices have been discussed in Sect. 2.1 and are related to the lack of semiconductor DBRs with low thermal resistivity, high

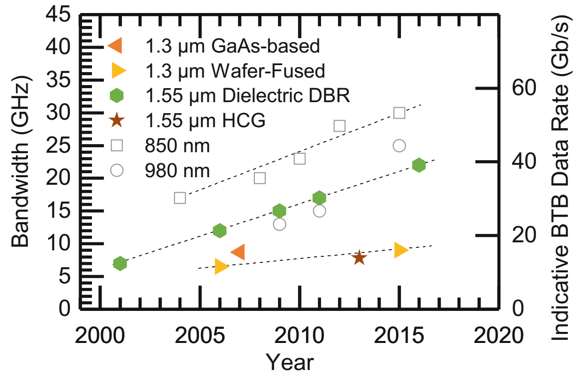


Fig. 2.5 Temporal evolution of the small-signal bandwidth for 1.3- and 1.55-µm VCSEL. For comparison, GaAs-based VCSELs are included (grey)

electrical conductivity and high refractive index contrast. The long-wavelength VCSELs, which have shown the highest modulation capabilities, belong to the family of devices using the hybrid dielectric-semiconductor cavities [3]. While they have the great advantage of a reduced effective cavity length compared to their semiconductor counterpart, they challenge the thermal and electrical management.

In this section, the stationary and dynamic characteristics of 1.55-µm VCSELs with single-mode hybrid dielectric-semiconductor cavities are reviewed.

2.2.1 *Hybrid Dielectric-Semiconductor VCSELs*

A schematic image of an InP-based VCSEL with dielectric DBRs is shown in Fig. 2.6a. This device has a cavity length which is three time the laser's emission wavelength (3-λ cavity) and it is encapsulated in an electrical passivation polymer. The field intensity and refractive index profile along the same cavity is shown in Fig. 2.6b.

The VCSEL's active region consists of eight compressively strained AlGaInAs quantum wells embedded between a *n*-doped InP layer and a *p*-doped AlInAs cladding layer. Current confinement is achieved by a circularly shaped p^+-AlGaInAs/n^+-GaInAs buried tunnel junction. The outcoupling mirror consists of a dielectric AlF_3/ZnS distributed Bragg reflector with 99.3% reflectivity, while the bottom mirror consists of a hybrid metal-dielectric Au-AlF_3/ZnS DBR with reflectivity of 99.9%. Furthermore, passivation with benzocyclobutene (BCB) reduces both chip and contact pad parasitic.

Dielectric DBRs are poor electrical and thermal conductors. Therefore, the heat generated in the VCSEL cavity flows radially across the overgrown InP, around the DBR, and it is drained out by means of the Gold anode which works as an heat sink. This three-dimensional problem can be simplified to a one-dimensional problem by solving the Fourier law for cylindrical coordinates, which lead to a thermal spreading resistance:

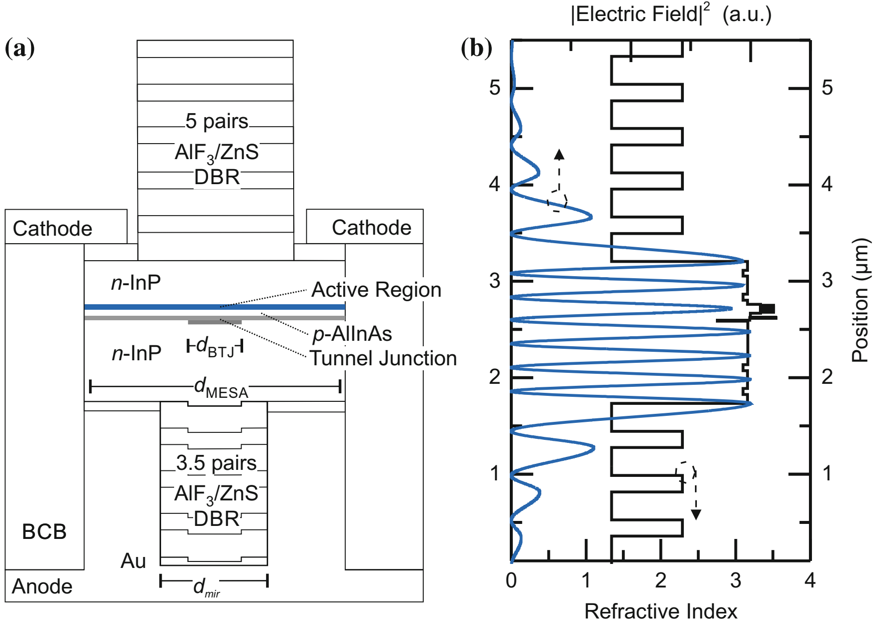


Fig. 2.6 **a** Schematic image of a 3-λ cavity InP-based VCSEL with dielectric DBRs encapsulated inside an electrical passivation polymer. **b** Field intensity and refractive index profile inside a 3-λ 1550-nm VCSEL

$$R_{\mathrm{th}} = \frac{\rho_{\mathrm{th}}}{2\pi L_{\mathrm{th}}} \ln\left(\frac{d_{\mathrm{mir}}}{d_{\mathrm{BTJ}}}\right) \tag{2.4}$$

where ρ_{th} is the thermal resistivity, and L_{th} is the thickness of the heat spreading layer, and the thermal resistance. Analogously, the electrical current has to flow radially across the two DBRs to reach the tunnel junction and the active region experiencing an electrical spreading resistance R_{el}. While small diameters of the DBRs are beneficial for the electrical and thermal resistance of the VCSEL, a lower bound to the diameter is given by the optical beam waist. In other words, the DBR has to be large enough to provide the desired reflectivity to the guided mode.

2.2.2 *Stationary Characteristics*

Optical and electrical stationary characteristics of VCSELs are strongly determined by the diameter of the tunnel junction. Large diameters lead to low serial resistances but allow several transversal modes to propagate within the VCSEL.

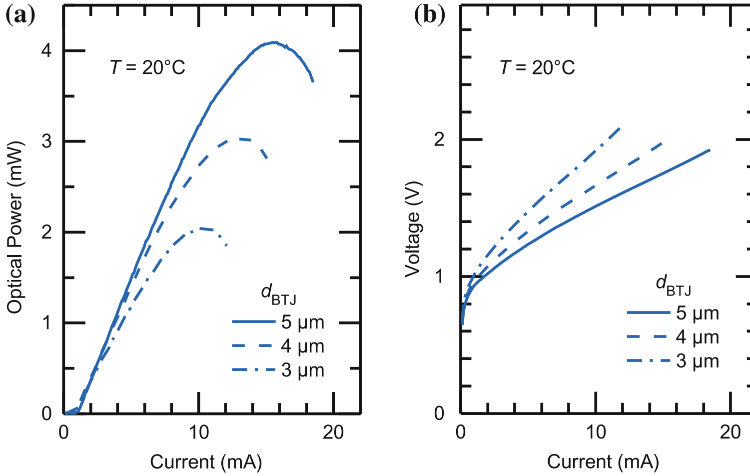


Fig. 2.7 **a** Emitted optical power versus bias current and **b** voltage versus applied current of a 1.55-µm VCSEL with diameter of the tunnel junction of 3, 4, and 5 µm at 20 °C

The output power versus the current for a 1.55-µm VCSEL is shown in Fig. 2.7a for devices with 3-, 4- and 5-µm diameter of the tunnel junction. The effectively pumped volume increases with the diameter of the current aperture leading to higher optical output power. The voltage-current characteristics are shown in Fig. 2.7b showing that the differential resistance decreases with increasing diameter of the tunnel junction. This is due to the increasing serial and spreading resistances which are proportional to $1/d_{\mathrm{BTJ}}^2$ and as $\ln\left(d_{\mathrm{mir}}/d_{\mathrm{BTJ}}\right)$, respectively.

For a VCSEL with aperture diameter of 5 µm, the power-current characteristics are plotted in Fig. 2.8a for different heat-sink temperatures ranging from 5 to 90 °C. Figure 2.8b shows the voltage drop across the VCSEL in the same temperature range displaying only weak temperature dependence with a slight decrease for high temperatures. The threshold current of a semiconductor laser is related to its high-speed performances. This is because the relaxation resonance frequency is proportional to the square root of the current relative to threshold [27]. In other words, lower thresholds are beneficial because they yield relaxation resonance frequencies with a lower current, that is, lower internal temperature. Figure 2.8c shows the threshold current as a function of the heat-sink temperature for three devices with 3-, 4- and 5-µm aperture diameter. The mode-gain offset was chosen to yield minimum threshold current at room temperature. For the same laser, the differential quantum efficiency is plotted as a function of heat-sink temperature in Fig. 2.8d. As can be seen, for smaller BTJs, the differential quantum efficiency decreases. This is due to diffraction effects and to the week confinement of the transversal modes propagating in the longitudinal direction [28].

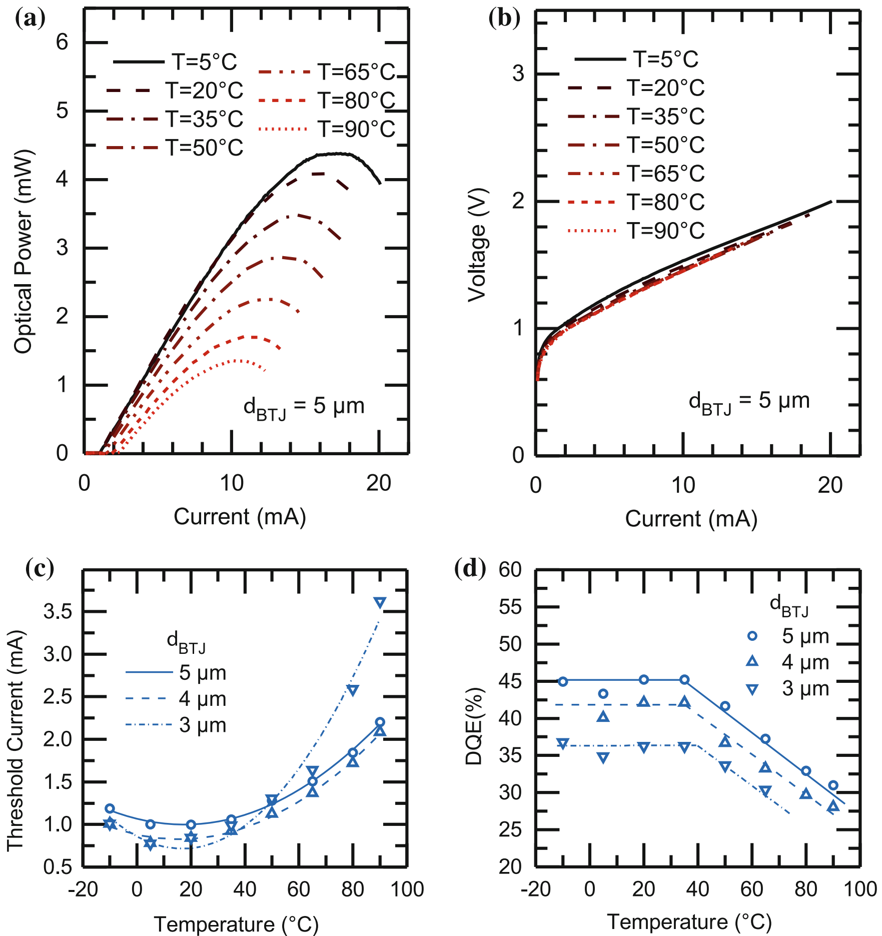


Fig. 2.8 **a** Emitted optical power and **b** voltage as a function of the bias current for a VCSEL with tunnel junction diameter of 5 µm operated at heat sink temperatures between 5 and 90 °C. **c** Threshold current and **d** differential quantum efficiency versus heat sink temperature for three different VCSELs with diameter of the tunnel junction of 3, 4 and 5 µm

Figure 2.9a shows the dissipated power and the wall-plug efficiency versus bias current measured at different heat-sink temperatures. While the dissipated power increases with the square of the applied current, the wall-plug efficiency has a maximum of 39% at 5 mA and at 20 °C. Finally, the room temperature optical spectrum of the same VCSEL is shown in Fig. 2.9b. This shows that, for diameters up to 5 µm, single mode operation with side mode suppression ratios above 40 dB can be achieved.

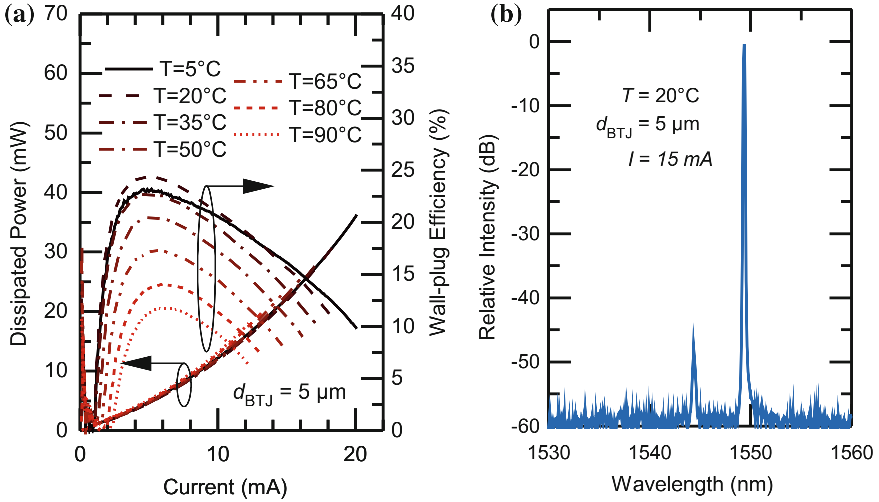


Fig. 2.9 **a** Dissipated power and wall-plug efficiency versus bias current for heat sink temperatures ranging from 5 to 90 °C. **b** Room temperature optical spectrum of a VCSEL at roll-over current with tunnel junction diameter of 5 μm

2.2.3 Dynamic Characteristics

The VCSEL's small-signal modulation response is determined by the combination of intrinsic and parasitic effects [29]. The overall modulation frequency response is given by a three-pole transfer function depending on the relaxation resonance frequency ν_R, the intrinsic damping factor γ and the parasitic cut-off frequency ν_p.

The VCSEL's small-signal modulation response was measured with a vector network analyzer for different bias currents. For an aperture of 5 μm and a 3-λ long semiconductor cavity length, a small-signal modulation bandwidth in excess of 17 GHz is achieved for a bias current of 10.4 mA and a DC electrical-power consumption of 16 mW, respectively, as shown in Fig. 2.10a. The relaxation resonance frequency versus the square root of current relative to threshold is plotted in Fig. 2.10b. The highest value achieved is 19 GHz. In Fig. 2.10c, γ is plotted versus ν_R^2. In the region where self-heating is small, the curve has a linear behavior with a slope K. The K-factor defines the maximum intrinsic capabilities of the laser and is proportional to the photon life time [27].

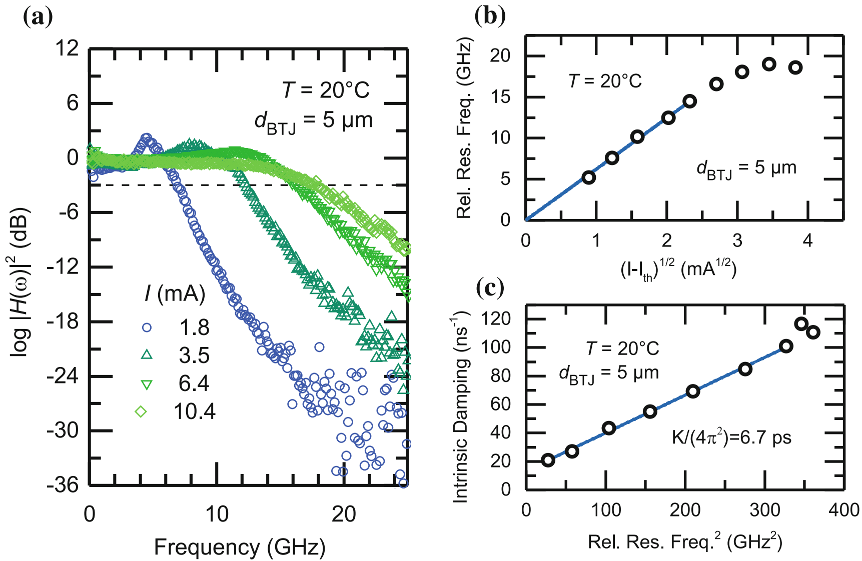


Fig. 2.10 For a 1.55-μm VCSEL with $d_{BTJ} =$ 5 μm operated at room temperature: **a** small-signal modulation response, **b** relaxation resonance frequency versus the square root of current relative to threshold, and **c** intrinsic damping versus squared relaxation resonance frequency

2.3 VCSEL Arrays and Advanced Modulation Formats

Long-wavelength VCSELs find applications in datacom and telecom networks thanks to the low optical attenuation of silica optical fibers. The small-signal modulation bandwidth is a common measure to quantify the high-speed capability of a semiconductor laser and the amount of data that can be transmitted depends on transmission distance, and modulation techniques. Furthermore, forward error correction (FEC) can also be used to boost the transmitted data.

In this section, we review data transmission experiments performed by using 1.5-μm 17-GHz VCSELs whose design and stationary performances are analogue to the one presented in Sect. 2.2. An overview of the large-signal modulation data rates and transmission lengths achieved with different transmission formats and detection schemes is presented in Table 2.1, and it will be discussed in this section.

2.3.1 Data Communication

Optical interconnect for intra data center applications must cover distances of few hundred meters and fulfill the need for low power, low cost and high density [36]. Long-wavelength VCSELs, thanks to the low losses in the SMF, are particularly attractive in large datacenters where links longer than 200 m are required. The

Table 2.1 Large-signal modulation net data rates and transmission lengths achieved with different transmission formats and detection schemes using single 1.5-μm VCSELs with small-signal bandwidth of approximately 17 GHz

VCSEL net bit rate (Gb/s)	SMF Length (km)	Modulation format	Transmitter	Detection	References
10	17.3	NRZ	DAC	Direct	[30]
25	4.2	NRZ	DAC	Direct	[2]
35	BTB	NRZ	DAC	Direct	[2]
40	1	4-PAM	0.13 μm SiGe BiCMOS	Direct	[31]
42	400	4-PAM	DAC	Choerent (20% SD-FEC)	[32]
44	960[a]	3-PAM	DAC	Choerent (20% HD-FEC)	[33]
50	2	NRZ	0.13 μm SiGe BiCMOS	Direct	[34]
56	BTB	NRZ	0.13 μm SiGe BiCMOS	Direct	[34]
79	4	DMT	72 GS/s DAC	Direct (20% HD-FEC)	[35]
88	0.5	DMT	72 GS/s DAC	Direct (20% HD-FEC)	[35]
96	BTB	DMT	72 GS/s DAC	Direct (20% HD-FEC)	[35]

[a]EDFA + DGEF

bit-error rate experiments described in this section refers to VCSELs with emission in the C-band, small-signal bandwidth of approximately 17 GHz and the design presented in Sect. 2.2.

In datacom networks, the need for low-latency and energy-effective links motivates the use of non-return-to-zero (NRZ) modulation. The highest bandwidth demonstrated with this modulation format in back-to-back (BTB) configuration is 56 Gb/s [34]. In this experiment, the electrical signal was generated by a 0.13-μm BiCMOS driver with 2-tap feed-forward equalization (FFE) and neither FEC nor digital-signal-processing (DSP) were used.

Using direct detection and advanced modulation formats such as 4-level pulse-amplitude modulation (4-PAM), net bit rates in excess of 40 Gb/s are transmitted over 1 km of single-mode fiber (SMF). These results were achieved by means of a energy-efficient 0.13 μm SiGe BiCMOS driver with 2-tap FFE architecture [31].

In order to make full use of the bandwidth of the transmission link, discrete multi-tone (DMT) and direct detection is a suitable candidate for short-reach high-density optical links. With this modulation format, 96-Gb/s net bit rate have been demonstrated in BTB configuration [35].

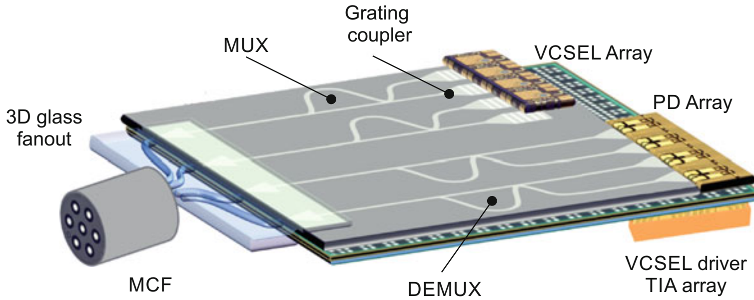


Fig. 2.11 Schematic of a 3D photonic and electronic silicon-on-insulator transceiver based on long-wavelength VCSELs arrays capable of terabits connectivity [37]

With device's data rate limited to a few tens of gigabits, terabit links are achieved by means of aggregate bandwidth. A 3D photonic and electronic silicon-on-insulator (SOI) transceiver based on long-wavelength VCSEL arrays and capable of terabits connectivity is depicted in Fig. 2.11 [37]. In this approach, a VCSEL array emitting at two different wavelengths is heterogeneously integrated to a SOI platform. The emitted light is coupled into a planar silicon waveguide by means of a grating coupler array. Pairs of signals with different wavelengths are multiplexed (MUX) to a single SOI's core and, later, coupled to the core of a multi core fiber (MCF) by means of a glass interface. Optical signals coming to the transceiver are demultiplexed (DEMUX) and read out by an array of photodiodes.

2.3.2 *Telecommunication*

Telecom access networks cover distances of several tens of kilometer. With the introduction of technologies such as fiber-to-the-house (FTTH), long-wavelength VCSELs have achieved increasing relevance in this area [38]. Metro and regional networks require links of the order of several hundreds of kilometers. In addition to the signal attenuation, metro fibers could have high polarization-mode dispersion (PMD) and large variations of chromatic dispersion (CD) compared to the access networks [39].

Single-mode 1.5-μm VCSELs have enabled 10 Gb/s transmission over 17 km through SSMF and using direct detection [30]. Furthermore, digital coherent detection at the receiver allows the recovery of the amplitude and phase of an optical carrier and the compensation by using DSP of CD and PMD. The combination of digital coherent detection and 4-PAM and 3-PAM results in net bit rates in excess of 40 Gb/s and transmission distances typical of metro networks (several hundreds of kilometers). A digital to analog converter (DAC) is used to generate these multi-level PAM signals. The longest distance achieved with this system is 960 km where a 4 × 80 km Erbium-doped-fiber amplifier (EDFA) recirculating

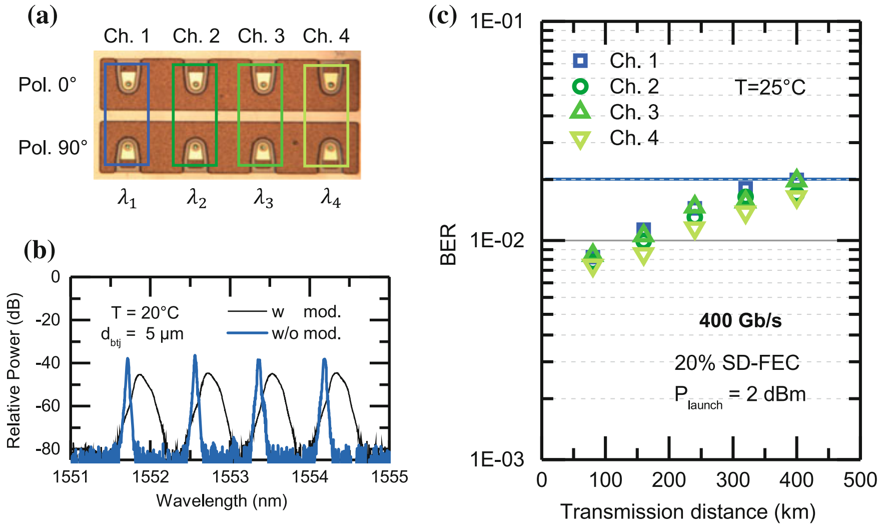


Fig. 2.12 **a** Microscope image of a 2 × 4 VCSEL array. **b** Spectrum of eight VCSELs emitting in four 100-GHz spaced channels transmitting two perpendicularly polarized signals. **c** BER versus transmission distance for the four channels transmission two perpendicularly polarized signals

loops are alternated with a dynamic gain equalization filter (DGEF) to block amplified spontaneous emission, and with an EDFA [33].

In Fig. 2.12a, a microscope image of a 2 × 4 VCSEL array is shown [32]. The array emits at four different wavelengths. The emission spectrum of the eight VCSELs consists of four 100-GHz spaced channels carrying two externally perpendicularly polarized 4-PAM signals as shown in Fig. 2.12b. These signals are detected with digital coherent detection, which enabled polarization DEMUX and compensation of CD, PMD and other impairments using DSP in the electrical domain. In Fig. 2.12c, the bit error rate (BER) versus transmission distance is shown for each of the four channels. A net information bit rate of 333 Gb/s is transmitted over 400-km SSMF assuming 20% overhead soft-decision forward-error-correction (SD-FEC) [32].

2.4 Conclusion

In this chapter, InP-based VCSELs have been studied with focus on the active region design, cavity length engineering, and current and optical confinement. Despite those arguments apply to all InP-based VCSELs with emission wavelength between 1.3 and 2.0 µm, stationary and dynamic characteristics are presented for a 1.55 µm VCSEL. Finally, datacom and telecom transmission experiments have been presented.

Acknowledgements This work was partially supported by the European Commission through the FP7 project MIRAGE (ref. 318228).

References

1. E. Haglund, P. Westbergh, J.S. Gustavsson, E.P. Haglund, A. Larsson, M. Geen, A. Joel, 30 GHz bandwidth 850 nm VCSEL with sub-100 fJ/bit energy dissipation at 25–50 Gbit/s. Electron. Lett. **51**(14), 1096–1098 (2015)
2. M. Müller, W. Hofmann, T. Grundl, M. Horn, P. Wolf, R.D. Nagel, E. Ronneberg, G. Böhm, D. Bimberg, M.-C. Amann, 1550-nm high-speed short-cavity VCSELs. IEEE J. Sel. Top. Quantum Electron. **17**(5), 1158–1166 (2011). doi:10.1109/JSTQE.2011.2109700
3. S. Spiga, D. Schoke, A. Andrejew, M. Müller, G. Boehm, M.-C. Amann, Single-Mode 1.5-μm VCSELs with 22-GHz small-signal bandwidth, in *Optical Fiber Communication Conference*, Anaheim, California, 20 Mar 2016. OSA Technical Digest (online) (Optical Society of America, 2016), p. Tu3D.4. doi:10.1364/OFC.2016.Tu3D.4
4. P. Sundgren, R. Marcks von Wurtemberg, J. Berggren, M. Hammar, M. Ghisoni, V. Oscarsson, E. Odling, J. Malmquist, High-performance 1.3 μm InGaAs vertical cavity surface emitting lasers. Electron. Lett. **39**(15), 1128–1129 (2003). doi:10.1049/el:20030733
5. H. Riechert, A. Ramakrishnan, G. Steinle, Development of InGaAsN-based 1.3 μm VCSELs. Semicond. Sci. Technol. **17**(8), 892 (2002)
6. J.A. Lott, N.N. Ledentsov, V.M. Ustinov, N.A. Maleev, A.E. Zhukov, A.R. Kovsh, M.V. Maximov, B.V. Volovik, Z.I. Alferov, D. Bimberg, InAs-InGaAs quantum dot VCSELs on GaAs substrates emitting at 1.3 μm. Electron. Lett. **36**(16), 1384–1385 (2000). doi:10.1049/el:20000988
7. Y. Arakawa, A. Yariv, Theory of gain, modulation response, and spectral linewidth in AlGaAs quantum well lasers. IEEE J. Quantum Electron. **21**(10), 1666–1674 (1985). doi:10.1109/JQE.1985.1072555
8. I. Suemune, Theoretical study of differential gain in strained quantum well structures. IEEE J. Quantum Electron. **27**(5), 1149–1159 (1991). doi:10.1109/3.83371
9. A.R. Adams, Strained-layer quantum-well lasers. IEEE J. Sel. Top. Quantum Electron. **17**(5), 1364–1373 (2011). doi:10.1109/JSTQE.2011.2108995
10. G. Böhm, R. Shau, R. Meyer, M.C. Amann, M. Ortsiefer, J. Rosskopf, F. Mederer, InP-based VCSEL technology covering the wavelength range from 1.3 to 2.0 μm. Paper presented at the 2002 International Conference on Molecular Beam Epitaxy, 15–20 Sept 2002
11. B. Zhao, T.R. Chen, A. Yariv, The extra differential gain enhancement in multiple-quantum-well lasers. IEEE Photonics Technol. Lett. **4**(2), 124–126 (1992). doi:10.1109/68.122336
12. M. Gendry, V. Drouot, C. Santinelli, G. Hollinger, Critical thicknesses of highly strained InGaAs layers grown on InP by molecular beam epitaxy. Appl. Phys. Lett. **60**(18), 2249–2251 (1992). doi:10.1063/1.107045
13. D. Ellafi, V. Iakovlev, A. Sirbu, S. Grigore, Z. Mickovic, A. Caliman, A. Mereuta, E. Kapon, Effect of cavity lifetime variation on the static and dynamic properties of 1.3-μm wafer-fused VCSELs. IEEE J. Sel. Top. Quantum Electron. **21**(6), 1–9 (2015). doi:10.1109/JSTQE.2015.2412495
14. P. Westbergh, J.S. Gustavsson, B. Kögel, A. Haglund, A. Larsson, Impact of photon lifetime on high-speed VCSEL performance. IEEE J. Sel. Top. Quantum Electron. **17**(6), 1603–1613 (2011). doi:10.1109/JSTQE.2011.2114642
15. K.Y. Lau, A. Yariv, Ultra-high speed semiconductor lasers. IEEE J. Quantum Electron. **21**(2), 121–138 (1985). doi:10.1109/JQE.1985.1072624

16. D.I. Babic, S.W. Corzine, Analytic expressions for the reflection delay, penetration depth, and absorptance of quarter-wave dielectric mirrors. IEEE J. Quantum Electron. **28**(2), 514–524 (1992). doi:10.1109/3.123281
17. S. Nakagawa, E. Hall, G. Almuneau, J.K. Kim, D.A. Buell, H. Kroemer, L.A. Coldren, 1.55-µm InP-lattice-matched VCSELs with AlGaAsSb-AlAsSb DBRs. IEEE J. Sel. Top. Quantum Electron. **7**(2), 224–230 (2001). doi:10.1109/2944.954134
18. L. Chao-Kun, D.P. Bour, Z. Jintian, W.H. Perez, M.H. Leary, A. Tandon, S.W. Corzine, M.R. T. Tan, High temperature continuous-wave operation of 1.3- and 1.55-µm VCSELs with InP/ air-gap DBRs. IEEE J. Sel. Top. Quantum Electron. **9**(5), 1415–1421 (2003). doi:10.1109/ JSTQE.2003.820924
19. A. Syrbu, A. Mircea, A. Mereuta, A. Caliman, C.A. Berseth, G. Suruceanu, V. Iakovlev, M. Achtenhagen, A. Rudra, E. Kapon, 1.5-mW single-mode operation of wafer-fused 1550-nm VCSELs. IEEE Photonics Technol. Lett. **16**(5), 1230–1232 (2004). doi:10.1109/LPT.2004. 826099
20. R. Yi, Y. Weijian, C. Chase, M.C.Y. Huang, D.D.P. Worland, S. Khaleghi, M.R. Chitgarha, M. Ziyadi, A.E. Willner, C.J. Chang-Hasnain, Long-wavelength VCSEL using high-contrast grating. IEEE J. Sel. Top. Quantum Electron. **19**(4), 1701311–1701311 (2013). doi:10.1109/ JSTQE.2013.2246780
21. K.D. Choquette, K.M. Geib, H.C. Chui, H.Q. Hou, R. Hull, Selective oxidation of buried AlGaAs for fabrication of vertical-cavity lasers. MRS Online Proc. Libr. Arch. **421**, 53 (1996). doi:10.1557/PROC-421-53
22. K.D. Choquette, R.P. Schneider Jr., K.L. Lear, K.M. Geib, Low threshold voltage vertical-cavity lasers fabricated by selective oxidation. Electron. Lett. **30**(24), 2043–2044 (1994). doi:10.1049/el:19941421
23. P. Petit, P. Legay, G. Le Roux, G. Patriarche, G. Post, M. Quillec, Controlled steam oxidation of AlInAs for microelectronics and optoelectronics applications. J. Electron. Mater. **26**(12), L32–L35 (1997). doi:10.1007/s11664-997-0065-0
24. C.W. Wilmsen, H. Temkin, Vertical-Cavity Surface-Emitting Lasers: Design, Fabrication, Characterization, and Applications, vol. 24 (Cambridge University Press, 2001)
25. M. Ortsiefer, R. Shau, G. Böhm, F. Köhler, G. Abstreiter, M. Christian-Amann, Low-resistance InGa(Al)As tunnel junctions for long wavelength vertical-cavity Surface-emitting Lasers. Jpn. J. Appl. Phys. **39**(4R), 1727 (2000)
26. G. Hadley, K. Lear, M. Warren, K. Choquette, J. Scott, S. Corzine, Comprehensive numerical modeling of vertical-cavity surface-emitting lasers. IEEE J. Quantum Electron. **32**(4), 607–616 (1996)
27. L.A. Coldren, S.W. Corzine, M.L. Mashanovitch, Diode Lasers and Photonic Integrated Circuits, vol. 218 (Wiley, 2012)
28. R. Burton, M. Stern, P. Kendall, P. Robson, Modelling of diffraction in pillar vertical cavity surface-emitting lasers with embedded Bragg layers. Opt. Quant. Electron. **28**(11), 1677–1684 (1996)
29. R.S. Tucker, High-speed modulation of semiconductor lasers. IEEE Trans. Electron Devices **32**(12), 2572–2584 (1985)
30. M.C. Amann, E. Wong, M. Müller, Energy-efficient high-speed short-cavity VCSELs. Paper presented at the Optical Fiber Communication Conference and Exposition (OFC/NFOEC), 2012 and the National Fiber Optic Engineers Conference, 4–8 Mar 2012
31. W. Soenen, R. Vaernewyck, Y. Xin, S. Spiga, M.C. Amann, K.S. Kaur, P. Bakopoulos, J. Bauwelinck, 40 Gb/s PAM-4 transmitter IC for long-wavelength VCSEL links. IEEE Photonics Technol. Lett. **27**(4), 344–347 (2015). doi:10.1109/LPT.2014.2372041
32. C. Xie, S. Spiga, P. Dong, P. Winzer, M. Bergmann, B. Kögel, C. Neumeyr, M.C. Amann, 400-Gb/s PDM-4PAM WDM system using a monolithic 2 × 4 VCSEL array and coherent detection. J. Lightwave Technol. **33**(3), 670–677 (2015). doi:10.1109/JLT.2014.2363017
33. C. Xie, P. Dong, P. Winzer, C. Gréus, M. Ortsiefer, C. Neumeyr, S. Spiga, M. Müller, M.C. Amann, 960-km SSMF transmission of 105.7-Gb/s PDM 3-PAM using directly modulated

VCSELs and coherent detection. Opt. Express **21**(9), 11585–11589 (2013). doi:10.1364/OE.21.011585

34. D.M. Kuchta, F.E. Doany, L. Schares, C. Neumeyr, A. Daly, B. Kögel, J. Rosskopf, M. Ortsiefer, Error-free 56 Gb/s NRZ modulation of a 1530 nm VCSEL link, in *2015 European Conference on Optical Communication (ECOC)*, 27 Sept 2015–1 Oct 2015, pp. 1–3. doi:10.1109/ECOC.2015.7341677
35. C. Xie, P. Dong, S. Randel, D. Pilori, P.J. Winzer, S. Spiga, B. Kögel, C. Neumeyr, M. Amann, Single-VCSEL 100-Gb/s short-reach system using discrete multi-tone modulation and direct detection, in *Optical Fiber Communication Conference*, Los Angeles, California, 22 Mar 2015. OSA Technical Digest (online) (Optical Society of America, 2015), p. Tu2H.2. doi:10.1364/OFC.2015.Tu2H.2
36. A. Vahdat, H. Liu, X. Zhao, C. Johnson, The emerging optical data center, in *Optical Fiber Communication Conference* (Optical Society of America, 2011), p. OTuH2
37. D. Apostolopoulos, P. Bakopoulos, D. Kalavrouziotis, G. Giannoulis, G. Kanakis, N. Iliadis, C. Spatharakis, J. Bauwelinck, H. Avramopoulos, Photonic integration enabling new multiplexing concepts in optical board-to-board and rack-to-rack interconnects, in *SPIE OPTO* (International Society for Optics and Photonics, 2014), pp. 89910D-89910D-89915
38. L.G. Kazovsky, W.-T. Shaw, D. Gutierrez, N. Cheng, S.-W. Wong, Next-generation optical access networks. J. Lightwave Technol. **25**(11), 3428–3442 (2007)
39. M.H. Eiselt, B.T. Teipen, Requirements for 100 Gb/s Metro Networks. Paper presented at the Optical Fiber Communication Conference and National Fiber Optic Engineers Conference, San Diego, California, 22 Mar 2009

Chapter 3
Quantum-Dot Semiconductor Optical Amplifiers for Energy-Efficient Optical Communication

Holger Schmeckebier and Dieter Bimberg

Abstract Quantum-dot (QD) based semiconductor optical amplifiers (SOAs) are key components for a large number of different applications in particular for all-optical communication networks. They are superior to classical semiconductor amplifiers in many important respects. Multi-wavelength amplification and signal processing at symbol rates larger than 40 GBd and operation in advanced modulation formats is needed in these networks. An introduction into the basics of QD SOAs as well as their key parameters is given at the beginning of this chapter. A novel concept for direct phase modulated signal generation is presented, unique for QD based SOAs. Error-free 25 GBd differential-phase shift keying (DPSK) signal is demonstrated, based there upon. The unique QD properties, i.e. decoupled gain dynamics of the various bound QD states, allows amplifying signals in dual-communication-band configuration both for small and large wavelength differences. Error- and distortion-free amplification of bidirectional 40 GBd on-off keying (OOK) signals, exhibiting a spectral separation of more than 91 nm is presented. Finally, all-optical wavelength conversion (AOWC) of phase-coded signals using four-wave mixing is shown. A guideline for the optimization of the conversion efficiency is given. Eventually error-free 40 GBd differential (quadrature) phase-shift keying (D(Q)PSK) AOWC is demonstrated.

H. Schmeckebier (✉) · D. Bimberg
Department of Solid-State Physics, Technische Universität Berlin,
Hardenbergstr. 36, 10623 Berlin, Germany
e-mail: h.schmeckebier@tu-berlin.de

D. Bimberg
King Abdulaziz University Jeddah, Jeddah, Saudi Arabia
e-mail: bimberg@physik.tu-berlin.de

© Springer International Publishing AG 2017

G. Eisenstein and D. Bimberg (eds.), *Green Photonics and Electronics*,
NanoScience and Technology, DOI 10.1007/978-3-319-67002-7_3

3.1 Introduction

The global IP traffic has increased tremendously in the last two decades, amounting presently to 60 EB per month in 2013 [1, 2]. A continuing annual increase of 23% per year is predicted, which will result in 168 EB per month in 2019. Metropolitan and wide area networks (MANs and WANs) presently carry each about 50% of the total traffic. MAN traffic however, is forecast to grow significantly faster than long-haul traffic [2]. Consequently, the number, diversity and complexity of photonic devices, presenting the physical layer of the networks will increase rapidly. Innovative solutions are required to control power consumption and cost of the networks.

Exclusively single-mode fibers (SMF) are used in the communication networks. The C-band around 1.55 μm is currently the most commonly used wavelength band, because here fibers show the lowest attenuation [3, 4]. The O-band around 1.3 μm exhibits a twice as large attenuation in the fiber, but features a zero crossing of the dispersion. Hence, this band is particularly interesting for low-cost MANs and access networks (ANs) as costly dispersion compensating techniques can be avoided.

State-of-the-art ANs and MANs use up to 40 wavelength channels, each carrying presently a data stream of typically 10 Gb/s, increasing to 40 Gb/s in the near future [5–8]. Next-generation optical MANs and ANs must serve a significantly increased number of customers and increased data rates per costumer. A larger number of customers can be accommodated by combining the currently used time-division multiplexing with wavelength-division multiplexing approaches [5, 7, 8]. Increased data rates per customer can be realized by boosting the current widely used intensity modulation, on-off keying (OOK), and by implementing advanced modulation formats, encoding multiple bits per transmitted symbol. Next-generation optical networks, such as converged MANs and reach extended ANs, will cover longer distances, and accommodate larger splitting ratios (number of customers) [5, 6, 9, 10]. Optical amplifiers are essential for both types of networks to compensate the additional losses caused by the extended reach and larger splitting ratios. Such amplifiers have to support multiple modulation formats, particularly intensity- and phase-coded formats, as well as multi-wavelength channel amplification with low channel crosstalk. Low energy consumption and low cost will be equally decisive parameters as maximum bit rate in the future, to keep not only investment down, but also end of life cost and take care of environmental concerns.

Various approaches for optical amplifiers have been developed in the past for optical communication networks, such as rare earth doped fiber amplifiers, Raman amplifiers and semiconductor optical amplifiers (SOA). Fiber amplifiers are widely used in WANs at 1.55 μm as they offer large gain and excellent noise performance. Their spectral gain bandwidth is however strongly limited and the investment costs are huge. The limited gain bandwidth can be overcome by a still more costly combination of fiber and Raman-amplifiers [3]. SOAs are low-cost mass products, showing reduced power consumption, smaller footprint, ease of integration in

photonic integrated circuits, and broad gain spectra, but at the expense of a slightly increased noise [11]. The decisive advantages of SOAs outweigh the poorer noise performance in case of networks spanning medium distances like MANs and ANs. Thus, SOAs are superior to fiber amplifiers widely used for linear amplification applications [6, 9, 12–16]. In addition, increasing network integration and modulation speed demand for optical signal processing, such as conversion of information between individual wavelength channels (all-optical wavelength conversion (AOWC)) [17].

The performance of any SOA depends on the nature of the gain material, and its carrier dynamics. In contrast to bulk semiconductors, semiconductor nanostructures exhibit a confinement of the charge carriers [18, 19]. The carrier confinement strongly influences the optical, electronic and thermal properties of the gain medium. In comparison to conventional bulk or quantum-well (QW) SOAs, quantum-dot (QD)-based SOAs demonstrate a number of unique properties [20], i.e. much fast carrier dynamics or decoupling of gain and phase dynamics.

The fast gain dynamics of QD SOAs has been demonstrated to enable single and multi-channel amplification of intensity-coded signals with symbol rates up to 80 GBd [21–26]. Also nonlinear effects, like cross-gain modulation (XGM), cross-phase modulation (XPM) and four-wave mixing (FWM), benefit from the fast QD gain dynamics even beyond saturation input power levels. Subsequently, these nonlinear effects have been used successfully to demonstrate single and multi-channel all-optical wavelength conversion (AOWC) of intensity-coded signals with symbol rates up to 320 GBd [22, 24, 27–34]. Beyond the common on-off keying (OOK) format, future networks will use advanced modulations formats, which are mainly based on phase or phase and intensity modulation. Thus, not only the gain dynamics but also the phase dynamics of QD SOA are decisive. XGM and XPM, however, are inappropriate to convert the wavelength of phase-coded signals in contrast to FWM.

This chapter summarizes the development of novel energy efficient QD SOA concepts based on the unique properties of GaAs-based QDs as well as wavelength conversion of high-bit rate phase-coded signals. Section 3.2 introduces the basic parameters of optical amplifiers, particularly those of QD SOAs, and like their dynamics. The design and static characteristics of such are presented. In Sect. 3.3 a novel concept based on the decoupling of gain and phase dynamics in a saturated QD SOA is discussed. Directly modulated phase-coded signal generation is introduced and demonstrated. Section 3.4 presents a novel concept for a dual-band optical amplifier. It exploits the individuality of QD energy states to amplify data signals. Section 3.5 is dedicated to nonlinear wavelength conversion in QD SOA, in particular the conversion of phase-coded signals. It starts with an introduction to FWM und subsequently shows how to optimize the static FWM performance. Finally, wavelength conversion of phase-coded signals is presented.

Parts of the results presented below were published in [34].

3.2 Basics of Quantum-Dot Semiconductor Optical Amplifiers

In this section basic parameters of optical amplifiers starting with their cw performance are defined. Subsequently, the gain, phase and carrier dynamics of SOA and the advantages of QD gain media are reviewed. Finally, cw characteristics of actual QD SOAs are demonstrated, important for the subsequent chapters.

3.2.1 Parameters of SOAs

3.2.1.1 Gain

The gain G is measured as a function of a signal input [35] generated by a polarized narrow bandwidth source with the optical power P_{in}. G is then defined as

$$G = \frac{P_{out} - P_{ASE}}{P_{in}},$$

with P_{out}: signal output power, P_{ASE}: amplified spontaneous emission (ASE) output power interpolated at the signal wavelength.

For fiber-coupled devices, the gain is commonly called fiber-to-fiber gain measured from the in-fiber input and in-fiber output power. These are the gain values of practical importance. The gain excluding fiber-coupling losses is called chip gain. Chip gain is commonly used to compare different types of devices excluding coupling losses If not stated differently, all gain and power levels given in this chapter are fiber-to-fiber gain values, being important for system applications.

3.2.1.2 Gain Saturation—Saturation Power

For small input power the gain is independent of its level. For a sufficiently large input power, the gain starts to depend on the input signal, called gain saturation or gain compression, defining the maximum extractable output power. Gain saturation is defined as the input (output) power at which the unsaturated or linear gain is reduced by 3 dB. The onset of gain saturation divides the amplifier operation into two regimes. The linear or small-signal regime is used for single and multi-channel amplification of data signals with low signal degradation and cross-talk between the channels. In contrast, cross-talk between the channels appears in the nonlinear regime, which can be utilized for signal processing. Typically, linear amplification is used up to the 3 dB saturation power, whereas efficient signal processing is usually employed for much larger input power levels.

3.2.1.3 Gain Bandwidth

The linear spectral gain peak and gain bandwidth characterize the accessible wavelength range for amplification of signals. The gain bandwidth is defined as the spectral range in which the gain is reduced by less than 3 dB with respect to the peak gain. Since the ASE represents the small-signal amplification of an internally generated signal, both the spectral gain peak and gain bandwidth can be estimated from the ASE spectra. The 3 dB ASE typically slightly underestimates the 3 dB gain bandwidth [34].

3.2.1.4 Polarization Dependent Gain

Two properties define the polarization dependence of SOAs. First, the typically asymmetric and tilted SOA waveguide structure supports TE polarized light [36]. Secondly, the gain media itself can exhibit a polarization dependent gain [36]. An effective polarization dependence of the gain can be suppressed by opposing polarization dependencies of both properties. The QD SOAs used here exhibit a predominant TE polarized gain. In general QD gain material can be designed to be polarization independent [22].

3.2.1.5 Noise Figure

The total output power of the SOA is given by the sum of the amplified input power and the ASE power. Hence, the ASE represents a noise floor degrading the amplified signal quality with respect to the input signal, reducing the optical signal-to-noise ratio (OSNR) of a data signal. As a figure of merit, the noise figure is defined as the OSNR of the input signal minus the OSNR of the output signal, both in logarithmic units. The noise figure NF can be calculated from gain measurements by [34, 37]:

$$\mathrm{NF} = 10\log\left(\frac{1}{\mathrm{G}} + \frac{2\mathrm{P}_{\mathrm{ASE,p}}\cdot\lambda}{\mathrm{G}\cdot\mathrm{h}\cdot\mathrm{c}\cdot\mathrm{B}_{0,\lambda}}\right) \quad \text{and} \quad \mathrm{P}_{\mathrm{ASE,p}} = \mathrm{P}_{\mathrm{ASE}}\frac{1}{1+\frac{\mathrm{P}_{\mathrm{ASE,r}}}{\mathrm{P}_{\mathrm{ASE,p}}}}$$

with $P_{ASE,p}$ ($P_{ASE,}$ r): power level of ASE co-(rectangular-)polarized to the signal, h: Planck constant; c: vacuum speed of light, $B_{0,\lambda}$: 3 dB filter bandwidth of the spectrometer, λ: wavelength of the signal. The minimum achievable noise figure within the linear gain regime (usually at large gain values) is NF = 3 dB.

3.2.2 Dynamics of Conventional and QD SOAs

This section follows closely descriptions given e.g. in [38, 39]. In the linear gain regime the output power is mainly limited by the number of the input photons (constant gain), whereas P_{out} is mainly limited by the number of charge carriers in the nonlinear regime. Thus, the carrier dynamics in the nonlinear gain regime determines the dynamics of the amplification for modulated signals. The response of a gain material to a modulated signal depends on the time scale of the modulated signal and the carrier dynamics. If the carrier density is also time dependent e.g. by modulation of the electrical pump current or by a strongly modulated optical input signal, the gain becomes time dependent as well.

Assuming the carrier dynamics to be orders of magnitude slower than the temporal amplitude changes of the optical input signal (slow amplifier), the gain material cannot respond to the modulated signal in time. For the amplifier, the modulated signal appears like an unmodulated signal with an averaged input power. Consequently, the gain remains time independent but is reduced to a steady state determined by a temporal average over the carrier depletion (during amplification) and injection (pumping). Hence, slow amplifiers exhibit input power independent pattern-effect-free amplification and multi-wavelength channel amplification without a XGM induced crosstalk between the channels. But of course, the average gain of each channel is influenced by the others.

Amplifiers exhibiting a much faster dynamics (fast amplifiers), at least of the order of the signal changes, can respond to the signal changes and processing of the signals becomes feasible. Patterning-free amplification and multi-wavelength channel amplification are usually feasible, as long as the amplifiers is operated in the linear gain regime. Commonly, SOAs belong to this amplifier type. The gain remains time independent in the linear gain regime, but commonly is time dependent in the nonlinear gain regime. Fast amplifiers are thus also suitability for processing of optical data signals.

Inter-band processes as well as intra-band processes contribute to the carrier dynamics and thus to the temporal gain dynamics. The different processes are experimentally accessible by various techniques, pump-probe configurations being the most common ones. A sufficiently short (usually a few hundred femtoseconds) and intense (nonlinear gain regime) optical pump pulse depletes a specific part of the density of states. Hence, the pulse pushes the gain medium out of the steady state condition. The appearing gain and phase change as well as the recovery of the system back to the steady state condition are probed using a sufficiently short and weak (linear gain regime) probe pulse. This pulse is temporally delayed with respect to the perturbing pump pulse [40].

A sketch of a pump-probe curve is shown in Fig. 3.1a and c for conventional and QD SOAs, respectively. The corresponding changes of the carrier densities are depicted in Fig. 3.1b and d.

In conventional SOAs, the amplification of the pump pulse injected at time $t = 0$ causes a gain compression. The carrier density is depleted in an energy interval

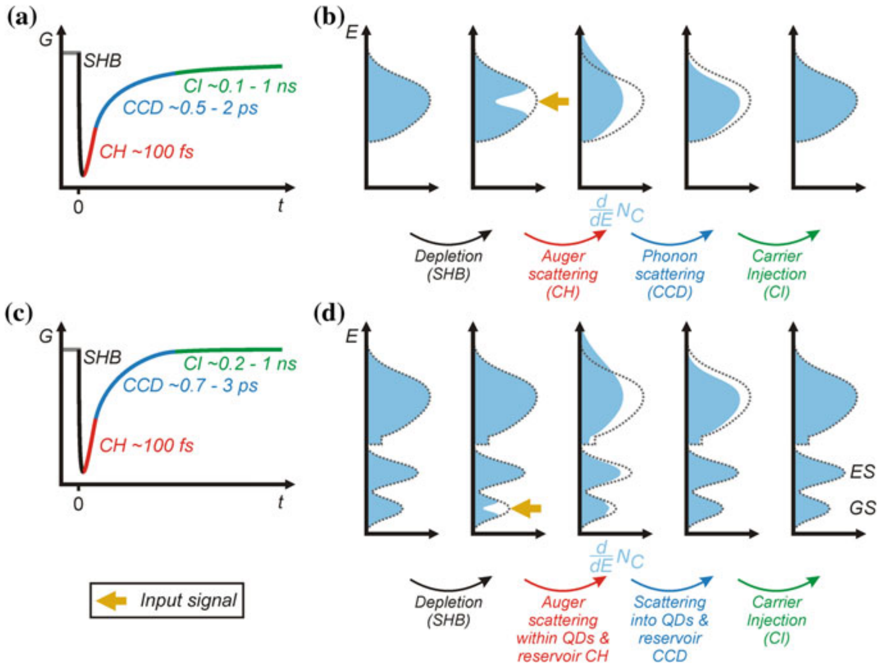


Fig. 3.1 Sketch of the pump-probe curve for **a** conventional and **c** QD gain media. The corresponding change of the carrier density d/dE N(E) is sketched in (**b**) and (**d**). For the sake of simplicity, only a joined electron-hole density for conduction and valance band is plotted. Reprint of [34]

determined by the photon energy of the optical pulse, called spectral hole burning (SHB). The spectral hole closes on a time scale of about 100 fs due to intra-band Auger scattering resulting in a partial recovery of the gain. The resulting carrier distribution equals a Fermi-distribution with a temperature larger than the lattice temperature and is thus called carrier heating (CH). The carrier distribution relaxes by LO phonon scattering within a few picoseconds, called carrier cool down (CCD). Finally, inter-band carrier injection (CI) refills the carrier density to its origin on a nanosecond time scale. Gain recovery is thus controlled by three main time constants and requires for conventional gain media up to 1 ns. This time can be reduced to the tens picosecond range e.g. by increased carrier injection. The time dependent gain and carrier density results also in a change of the refractive index and accordingly into phase changes on a comparable time scale. In case of a repeating optical signal, carrier depletion and CI are repeated resulting in a dynamic equilibrium called carrier-density pulsation (CDP).

Gain recovery dynamics is advantageous in QD gain media. QDs exhibit discrete bound states, a ground state (GS) and typically excited states (ES) [19, 41, 42]. The QD ensemble exhibits an inhomogeneously broadened optical emission resulting in a large spectral gain bandwidth of up to 120 nm [43, 44]. The QD states

supply the gain, but the carrier reservoir being at much larger energy in this separate confinement heterostructure (SCH) control the index of refraction of the amplified light in the waveguide in case of a highly populated reservoir.

Similar to conventional gain media, SHB occurs in the inhomogeneously broadened QD GS carrier density due to amplification of the pump pulse. Refilling of the QD GS is due to Auger scattering from the reservoir on a time scale down to 100 fs [45–47]. Subsequently, CCD of the reservoir carrier distribution takes place on the time scale of a few picoseconds [45–47]. The QD GS gain recovery is complete before CI refills the reservoir typically on a longer time scale. As long as the carrier reservoir is strongly populated still after QD state depletion, the QD gain recovery is only limited by the closure of the SHB as well as by the intra-dot relaxation time for larger depletion and/or durations [48]. The gain of QDs emitting at a different center wavelength as compared to the depleted QDs is not influenced if the reservoir remains strongly populated. A perturbation between the QDs can be mediated only via the carrier reservoir [49]. If the carrier reservoir is not strongly populated, the QD gain recovery is controlled by the carrier density of the reservoir and thus follows the reservoir dynamics limited by CDP, similar to conventional gain media [39].

The phase dynamics is determined by the carrier density change within the complete SCH (Kramers-Kronig relation). In principle each group of states, i.e. QD GS, ESs and the reservoir exhibit carrier density changes and thus contribute to the phase dynamics. The number of carriers in the reservoir is orders of magnitude larger than that in the QDs and the gain recovery of the QDs states is fast. Therefore only mid- and long-term phase dynamics are dominated by the reservoir carrier dynamics. Hence, the separation of the energy states into gain providing QD states and a carrier reservoir, controlling the index of refraction at much larger energy results I for the case of a strongly populated reservoir in a decoupling of gain and phase dynamics [50] and presumably a lower α-factor [23, 51, 52].

3.2.3 Design and Static Characteristics of QD SOAs

3.2.3.1 Design

Our QD SOAs based on AlGaAs/GaAs heterostructures are grown by molecular beam epitaxy (MBE). The active region contains ten stacks of InGaAs/GaAs self-organized QDs each overgrown by an InGaAs QW (DWELL). The emission wavelength is at 1.31 μm, the center of the O-band. Suitable GaAs barrier layers between the QD layers provide for strain relaxation. The optical mode is vertically guided by a symmetric waveguide, which is formed by sandwiching the 420–440 nm thick GaAs core region between 1.5 μm thick AlGaAs cladding layers.

The lateral confinement of the light as well as a current guiding is realized by tailoring the SOA structure using a 4 μm thick ridge-waveguide structure. Deeply-etched ridge waveguide structures are formed by etching through the active

region, whereas shallow-etched ridge waveguide structures stop about 170 nm above the active region. The complete structure is covered by silicon nitride for electrical isolation. In addition, the deeply-etched structure is planarized using Benzocyclobutene (BCB) reducing the parasitic capacitances being important for direct modulation. Both structures receive a tops p-contract and backside n-contact. The deeply-etched structure receive in addition a topside n-contact enabling the electrically contact by a HF probe head. The shallow-etched ridges offer comparatively lower waveguide losses. The deeply-etched waveguides offer a stronger confinement of the optical mode and the capability for high-speed current modulation. The shallow (deeply)-etched ridges are tilted by 8° (6.8°) with respect to the facet normal and the facets are anti-reflection coated to suppress the onset of lasing even at high current densities.

The devices are glued on a copper-mounts and mounted on a temperature controlled heat think (room temperature (20–25 °C)). Fiber-coupling is realized by positioning tapered fibers using a feedback control system. The fiber-coupling losses are typically 4–4.5 dB per side.

3.2.4 QD SOA Sample Series

Six different QD SOAs are used for the experiments presented in this chapter. Results for the QD SOA No. 5 will be presented exemplary now in detail. The waveguide of the 5 mm long QD SOA is shallow-edged. The diode exhibits a series resistance of 0.53 Ω. The total in-fiber ASE intensity reaches 12 mW at a current of 1 A and is still not saturating. The QD GS ASE emission saturates at a current of around 350 mA, whereas the ES ASE emission is still unsaturated at a current of 1 A. The GS ASE peak wavelength is 1306 nm. Increasing the current to 500 mA reduces the peak by about 0.7 dB and shifts the wavelength to 1310 nm. For currents up to 1 A, the QD GS ASE emission suffers from device heating, resulting in a reduction of up to 5 dB and a wavelength red-shift to 1320 nm.

The device exhibits a linear fiber-to-fiber gain as large as 25.5 dB at a current of 400 mA and a wavelength of 1310 nm. The corresponding 3 dB saturation input (output) power level is –11.5 dBm (11 dBm). The corresponding fiber-to-fiber noise figure is about 10.5 dB (Fig. 3.2).

Table 3.1 provides an overview of the static characteristics of all six QD SOAs. Increasing the length of the devices increases the gain up to 25.5 dB (34.5 dB chip gain) at the expense of a spectral bandwidth decreasing from 34 to 21 nm. The saturation input and output power are decreased and increased, respectively, accordingly to the gain increase. All devices exhibit a polarization dependence. Spectrally resolved measurements of the polarized ASE of QD SOA No. 3 showed a TE-TM polarization ratio of up to 19 dB for QD GS emission and about 7 dB for the ES emission [34].

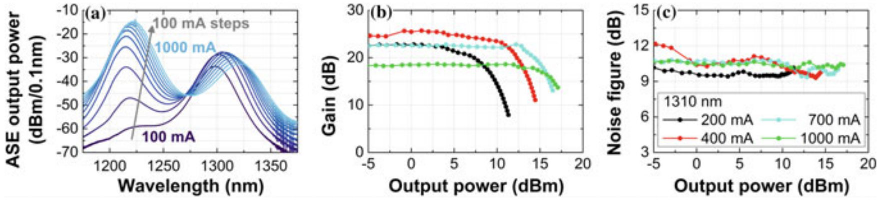


Fig. 3.2 Characteristics of QD SOA No. 5. **a** Spectrally resolved in-fiber ASE in dependence of the current. **b** Fiber-to-fiber gain and **c** noise figure as a function of the in-fiber output power, both measured at a wavelength of 1310 nm for different currents ((**b**) and (**c**) share the same legend)

Table 3.1 List of characteristics of the QD SOAs used for experiments in this chapter. QD SOA No. 1 (marked with *) is deeply-etched, whereas all others are shallow-etched. J is current density at which the maximum gain is obtained for an input wavelength λ_G. All gain G, noise figure NF, 3 dB saturation input power P_{in}^{sat} and output power P_{out}^{sat} values are fiber-to-fiber and in-fiber values, respectively. The fiber coupling losses vary from device to device and with the current between 4 and 4.5 dB per facet. The chip gain, the chip noise figure and the chip saturation power levels can be obtained by increasing the given in-fiber values by 8–9 dB, 4–4.5 dB, and 4–4.5 dB, respectively

QD SOA		ASE		Amplification using QD GS					
No.	Length (mm)	λ_{peak} (nm)	FWHM (nm)	λ_G (nm)	G (dB)	NF (dB)	P_{in}^{sat} (dBm)	P_{out}^{sat} (dBm)	J (kA/cm^2)
1*	2	1296	33.0	1296	8.4	11.8	4.1	9.5	1.9
2	3	1284	34.0	1290	14.6	9.4	3.3	14.9	3.4
3	3	1299	28.0	1305	13.5	8.6	1.0	11.5	3.3
4	4	1304	23.2	1310	22.1	9.5	−6.3	12.8	3.1
5	5	1307	21.0	1310	25.5	10.5	−11.5	11.0	2.9

3.3 Phase Modulation of QD SOAs

3.3.1 Introduction of the Concept

Directly modulated emitters are extensively used in low cost optical networks. They simplify their architecture, and reduce the overall investment and maintenance costs. As the number of wavelength channels are rapidly increasing, especially in fiber-to-the home (FTTH) systems, a more flexible network design is preferable to further reduce the maintenance costs.

Realizing color-free up- and down-stream operation in passive optical networks presents one option [5, 53–55]. The up-stream wavelength is defined here by the central office instead of the customer. Different approaches based on lasers, amplifiers and modulators have been demonstrated [53–56]. A simple approach is based on directly modulated SOA demonstrated for OOK in [57, 58]. Phase-coded signals, such as differential-phase-shift keying (DPSK), offer reduced optical signal-to-noise ratio (OSNR) requirements in comparison to OOK signals.

Directly modulated generation of phase-coded signals has the potential to simplify the network architecture, improve the reliability, reduce the cost and possibly the energy consumption in comparison to alternative color-free configurations. Amongst them are OOK-based directly modulated SOAs, or combinations of standard Mach-Zehnder modulators or phase modulators with amplifiers, required to overcome the modulator losses. DPSK generation up to 16 GBd has been demonstrated using DFB-lasers [59, 60]. 10 GBd OOK generation has been demonstrated for directly modulated SOAs [57, 58, 61].

QD based SOAs are ideal for realizing direct phase modulation [62] due to the unique decoupling of phase and gain dynamics as discussed in Sect. 3.2.2. Large drive currents lead to saturated QD emission and a strongly populated reservoir. A modulation of the current will then result in a change of the reservoir carrier density, and thus the index of refraction, but has no or negligible influence on the occupation of the QD states, and thus the gain. The change of the refractive index can be probed by an optical wave. Hence, the phase of an optical input signal can be modulated by changing the carrier density without changing the amplitude.

3.3.2 Prove of the Concept

A cw external cavity laser (ECL) emitting at a wavelength of 1296 nm is injected into the QD SOA with its polarization aligned to the TE axis. The QD SOA is driven by a DC current of 125 mA, which is superimposed by an electrical non-return-to zero (NRZ) pseudo-random-binary sequence $2^7 - 1$. Hence, the optical output of the QD SOA is either OOK or DPSK modulated or both, depending on the operating conditions. The optical output signal is wavelength filtered to prevent ASE loading of the receiver amplifiers.

The properties of the output are tested in a BtB configuration. The multi-format receiver contains an optical pre-amplifier stage and subsequently two different receiver types. An optical switch is used to detect the signal with one of two receivers. A direct detection receiver is used to characterize the remaining amplitude modulation by eye-pattern measurements. A differential receiver comprising a delay interferometer (DI) and a photodiode is used to demodulate the phase-modulated signal. The DI's free-spectral range equals the symbol rate. The signal quality is evaluated by bit-error ratio (BER) measurements as well as eye-pattern measurements.

The device under test is the 2 mm long QD SOA No. 1 (see Table 3.1). The ASE peak is observed at a DC current of close to 150 mA at a wavelength of 1296 nm. At this wavelength, the ASE varies by only 0.6 dB when changing the current from 100 to 200 mA. This reduction is predominantly caused by device heating induced shift of the ASE peak wavelength. The ASE variation is expected to be even smaller for a fast modulation of the current, because the modulation speed in optical communication systems is much faster than acoustic phonon emission in semiconductors. At a current of 150 mA, the device shows a

small-signal fiber-to-fiber gain of 8.5 dB as well as a 3 dB saturation input power of 4.1 dBm.

The optimum modulation point for the generation of phase-coded signals, i.e. DC current and modulation amplitude, is given by a trade-off between different effects:

- A low DC current results in unsaturated GS emission and consequently in significant amplitude modulation. A large DC current causes device heating induced gain reduction, and hence a lower signal quality at the receiver.
- A small modulation amplitude causes only a small variation of the carrier density and thus a small phase shift. Large modulation amplitudes increase the phase shift but can push the device out of the saturated regime resulting into amplitude modulation of the optical output signal.

Suppressed amplitude modulation but significant phase modulation is expected from the DC results for a current range of 100–200 mA.

Much below 100 mA and a modulation rate of 6 GBd, a disturbed OOK eye diagram was measured. For currents above 100 mA no eye could be observed, but still an amplitude modulation as depicted in Fig. 3.3c for a rate of 10 GBd. A DC current of 125 mA and a peak-to-peak voltage of 3 V were identified to yield optimum results, demonstrated by the open eye at the DI output port 1. No eye could be observed simultaneously at the DI port 2. Changing the phase of DI by 180° inverts the observed behavior of the ports. Similar results were obtained for all tested combinations of current, modulation amplitude and modulation rate.

A simple simulation of the DI was performed to analyze the observations. The phase and amplitude of an electromagnetic wave is modulated by a data signal without any bandwidth limitation. The modulated signal is either directly detected using a square-law detector or interferes with its one symbol delayed copy and both DI output ports are equipped with square-law detectors. The bandwidth of real detectors is simulated by stepwise integration over a temporal interval (corresponding to the symbol rate), which simply filters out the fast carrier oscillation. The results are shown in Fig. 3.4.

Starting with no amplitude modulation (left part of the figure), the directly detected signal shows no modulation whereas both DI output ports exhibit clear

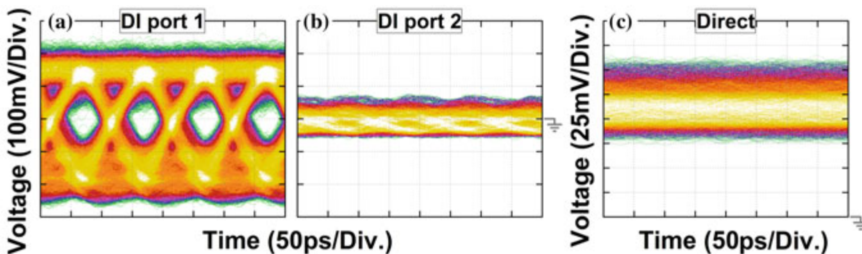


Fig. 3.3 Eye diagrams of the 10 GBd directly modulated phase-coded signal generation recorded at **a** DI port 1, **b** DI port 2 and **c** the direct detection pass. The receiver input power was set to the BER threshold of 10^{-9}. Reprint of [34]

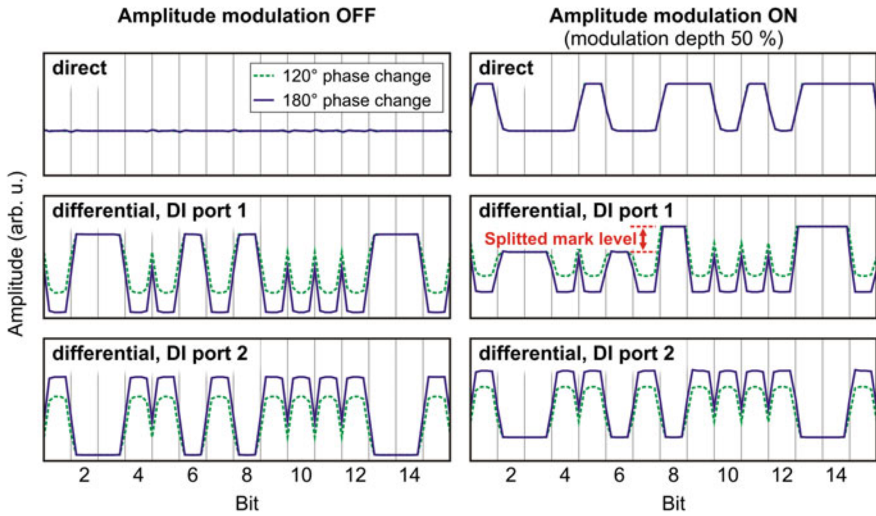


Fig. 3.4 Simulated detection of a DPSK signal using a direct detection receiver and a differential detection receiver. The differential receiver is simulated in single-ended detection for DI port 1 and port 2. The DPSK signal is simulated with a 120° and 180° phase change, both (*left*) without and (*right*) with a simultaneous amplitude modulation. In the presence of an amplitude modulation, DI port 1 exhibits two mark bit levels, whereas port 2 shows only one mark level. All figures share the same legend. Reprint of [34]

space and mark levels. The extinction ratio is given by the phase change and reaches its maximum at 180°. Taking into account an additional amplitude modulation (right part of the figure), a different behavior of the DI output ports is observed. Port 2 experiences only a reduction of the extinction ratio between mark and space level, whereas port 1 experiences in addition a splitting of the mark level.

Comparing the eye diagrams for symbol rates of 10 and 25 GBd, depicted in Figs. 3.3 and 3.5, the remaining amplitude modulation is damped with increasing symbol rate as indicated by the reduced amplitude swing. In consequence, the demodulated DI output shows a significantly improved eye diagram 25 GBd, demonstrated by the reduced space level variation. The experimentally observed behavior presumably results from a residual amplitude modulation of the QD SOA. This could be reduced by decreasing the modulation amplitude which would also reduce the phase change and thus decrease the extinction ratio. The differential detection processes only phase changes occurring between two subsequent symbols. According to Sect. 3.2.2, the effective time constant of the reservoir includes all carrier relaxation processes. This effective time constant is slower than the time constant of the current RF-signal induced adiabatic chirp in the reservoir. Consequently, the differential detection processes only the fast contributions to the total

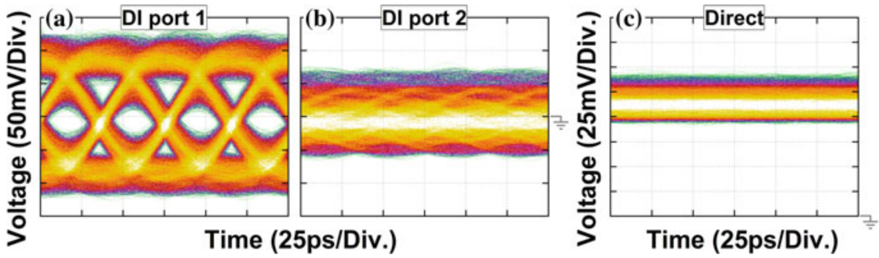


Fig. 3.5 Eye diagrams of the 25 GBd directly modulated phase-coded signal generation recorded at **a** DI port 1, **b** DI port 2 and **c** the direct detection pass. The receiver input power was set to the BER threshold of 10^{-9}. Reprint of [34]

phase change within the reservoir. The remaining amplitude modulation especially at lower symbol rates could cause problems even for large symbol rates, like 25 GBd, when using larger PRBS length.

BER measurements were performed for a fixed QD SOA optical input power of −5 dBm and for symbol rates from 10 to 25 GBd. The BER curves are evaluated in terms of receiver input power penalty at a BER of 10^{-9} (error-free) and 10^{-3} with respect to the lowest symbol rate of 10 GBd. Doubling the symbol rate for an ideal transmitter-receiver setup would result in a 3 dB receiver input power penalty as the ideal receiver requires a constant received energy per symbol for comparable BERs. The penalty for a real non-ideal transmitter-receiver system is usually larger due to bandwidth limitations of the transmitter and/or receiver. The results are shown in Fig. 3.6a.

Error-free generation of a phase-coded signal was observed by us up to 25 GBd, presumably limited by a non-optimized RF design. The receiver input power penalty is about 6.6 dB, thus 2.7 dB larger than the reference for a BER of 10^{-9} and 10^{-3}, respectively. The penalty follows the reference up to a symbol rate between 15 GBd and 20 GBd, presumably corresponding to the small-signal modulation bandwidth of the QD SOA.

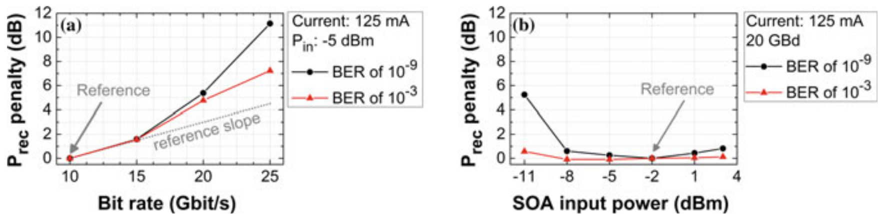


Fig. 3.6 Receiver input power penalty versus **a** the bit rate and **b** the QD SOA optical input power level. The optical input power was fixed to –5 dBm while scanning the symbol rate and the symbol rate was fixed to 20 GBd while scanning the optical input power level. The QD SOA was driven in all case with a current of 125 mA. The penalty was determined at a BER of 10^{-9} and 10^{-3} with respect to the curve marked with “Reference”. Reprint of [34]

At a symbol rate of 20 GBd, the dependence on the input power level is investigated by changing the input power levels within a range of −11 to 3 dBm. The BER curves are evaluated in terms of receiver input power penalty with respect to the optimal input power of −2 dBm. The results are shown in Fig. 3.6b. A penalty below 1 dB can be observed for QD SOA input power levels from –8 to 3 dBm for both BER thresholds. A further reduction of the input power to –11 dBm leads only to a slight increase of the penalty at a BER of 10^{-3} whereas the penalty is larger than 5 dB for a BER of 10^{-9} caused by an error-floor. Hence, error-free phase-coded signal generation is demonstrated for an input power range of more than 14 dB and symbol rates of 25 GBd, demonstrating the suitability of directly modulated QD SOAs for low-cost color-free networks.

3.4 Concept of Dual-Communication-Band Amplifiers

The data up and down stream in current MANs and ANs are typically realized in two different optical communication bands to avoid cross-talk. Hence, these approaches will require one SOA per band. A reach extender can be implemented based on a wavelength splitter, separating and recombining the two bands with the amplifiers in-between.

Using a dual-band amplifier instead will considerably simplify the network architectures as filters are not required. The power budget will be increased since the losses are reduced due to the lower number of inline components. In addition, such amplifiers allow a reduction of the energy consumption. The power consumption of a single device is presumable lower than that of two devices and only one thermoelectric cooler is required.

Dual-band amplifiers have to provide the capability of simultaneously amplifying counter-propagating up-stream and down-stream data signals from different wavelength bands whereas each band carries multiple wavelength channels. This section will present the concept of a dual-band bidirectional SOA based on QDs [63].

3.4.1 Introduction of the Concept

The gain of QD ensembles is inhomogeneously broadened due to the size and composition distribution of the QDs, which are non-interacting on a fast time scale. The non-interacting nature of the bound charge carriers in zero-dimensional systems is in marked contrast to the carriers interacting on a fs time scale in any higher dimensional system like quantum wells. Amplification via the GS of QDs was found to be very advantageous for single- and multi-wavelength amplification and for various modulation formats with symbol rates up to 80 GBd [21–25, 64–66]. The dual-band QD SOAs presented below use for the first time both, the QD GS

and ES gain to amplify data signals representing now up and down stream and thus the two bands. This concept can be extended such that the inhomogeneously broadened GS and ES ensembles present the basis for multi-wavelength-channel amplification within both communication bands.

In order to use both QD states simultaneously, details of the inter-state cross-gain modulation (XGM) and the cross-phase modulation (XPM) will be of primary importance for realizing this concept. The dynamics of QD gain media depend on the charge-carrier dynamics of the different energy states and the external optical field. The amplitude- and/or phase-modulated optical field induces changes of the charge-carrier occupation of the different energy levels. Thermal equilibrium population of the bound states can be reestablished in principle by scattering down from the carrier reservoir or by transitions between bound QD states.

Amplification experiments performed on the QD GS demonstrate a fast carrier refilling of the GS and thus a low channel crosstalk even in the nonlinear gain regime as long as the carrier reservoir is highly populated [24]. Based on pump-probe experiments and simulations [67, 68], similar results are expected for multi-channel ES amplification. The XGM between wavelength channels addressing different groups of QD states is mediated only via the reservoir and can be neglected as long as the reservoir remains strongly populated. XPM in QD SOAs on the other hand is controlled only by the carrier reservoir but not by the QD states. Hence, the XPM between the QD states is mediated via the reservoir and are thus comparable to XPM in single- and multi-channel single-band configuration [24]. The only remaining question discussed in the following is the importance of XGM between the states of individual QDs.

3.4.2 *Proof of Concept*

Next we present the proof of concept of an ultra-broadband, bidirectional dual-band QD SOA for OOK modulated signals using the 3 mm long QD SOA No. 3 (see Table 3.1). This modulation format is expected to cause the largest possible crosstalk similar to multi-wavelength channel single-band configurations. Two 40 GBd NRZ OOK (PRBS $2^7 - 1$) signals with a wavelength spacing of 91.5 nm, a GS signal wavelength at 1305 nm and an ES signal wavelength of 1213.5 nm, represent the two bands. The signals are counter-propagating. The output signals are easily separated by coarse-wavelength division multiplexing filters from the counter propagating input signals. The BER and eye diagram of the GS is measured using an optical pre-amplifier based receiver. BER and eye diagram measurements of the ES use a direct detection receiver (details see [34, 63]).

Three different operating regimes have been investigated. The ES ASE is unsaturated, close to saturation and saturated for currents of 400 mA, 700 mA and 1000 mA, respectively. In contrast, the GS ASE is saturated for all three currents. Due to device heating, the GS ASE peak power decreases and is red shifted by

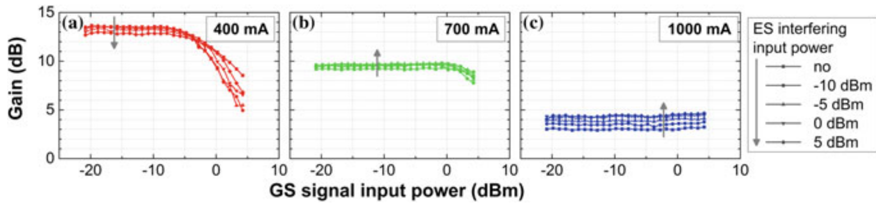


Fig. 3.7 Gain of the GS signal versus input power for three different QD SOA currents (**a**–**c**). The gain is shown for different QD SOA input power levels of the interfering ES signal listed in the legend. The *arrow* in each graph shows the change of power according to the arrow in the legend

0.035 nm/mA on average within this current range. In contrast, the ES ASE peak power first increases with increasing current from 400 to 700 mA and decreases similar to the GS ASE peak at larger currents. On average, the ES peak is red shifted by about 0.026 nm/mA.

The gain characteristics of both signals in presence and/or absence of the counter-propagating signal (called interfering signal) is shown in Figs. 3.7 and 3.8 for three different currents and for different input power levels. An ES signal changes the linear GS gain by less than ±1.2 dB for all currents. Thus one essential requirement for the simultaneous amplification of GS and ES signals is fulfilled.

The influence of the GS signal on the gain of the ES is more complex. For a current of 400 mA the ES gain reduction caused by a GS signal exceeds 10 dB for the largest input power. This effect is typical for an unsaturated amplifier and is similar to multi-channel single-band configurations. The ES gain performance is strongly improved for larger currents where the ES is saturated as well. At 700 mA the linear gain reduction is small, with values below 1.5 dB for GS input power levels up to 0 dBm. Only at a very large GS input power level of 5 dBm, again the ES gain drops up to 7.5 dB. Finally, at a current of 1000 mA, the ES gain is nearly independent of the GS input signal power. At these large currents the carrier reservoir is highly populated, resulting in a very efficient simultaneous refilling of both states, independent of the optical input power. But indeed, it is also supported by the reduced GS gain. Taking into account XGM, saturation input power as well

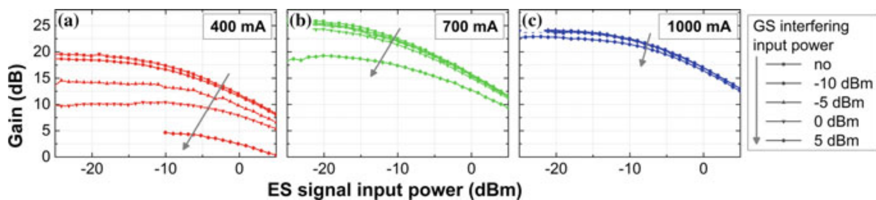


Fig. 3.8 Gain of the ES signal versus input power for three different QD SOA currents (**a**–**c**). The gain is shown for different QD SOA input power levels of the interfering GS signal listed by the legend. The *arrow* in each graph shows the change of power according to the arrow in the legend

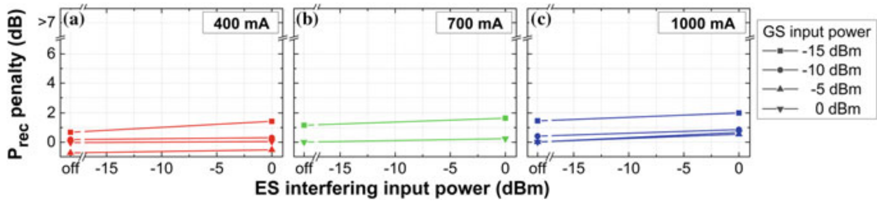


Fig. 3.9 Reveiver input power P_{rec} penalty of the 40 GBd OOK GS at a BER of 10^{-9} versus the QD SOA interfering ES input power. The QD SOA currents are **a** 400 mA, **b** 700 mA and **c** 1000 mA. The results in the absence of the interfering ES signal are plotted at a x-axis value labeled with "off"

as the gain values, a current of about 700 mA is expected to be the optimum current for the dual-band QD SOAs investigated here.

BER measurements have been performed in dependence on the receiver input power level P_{rec} without (BtB) and with the QD SOA. The BER curves are analyzed in terms of the P_{rec} penalty at a BER of 10^{-9} with respect to the corresponding BtB measurement.

Figure 3.9 shows the penalties of the GS signal as a function of the input power of the ES signal. The penalty is nearly independent of the QD SOA current and the GS input power in absence of an ES signal. A negligibly small penalty below 1 dB is found for all tested input power levels, except for the lowest one of −15 dBm. Here, the penalty is slightly increased to about 1.5 dB. The presence of an ES signal leads at most to a minor increase of the penalty by 0.7 dB, despite a very large ES input power of 0 dBm. 0 dBm is more than 5 dB above the 3 dB saturation input power of the ES signal. Still, the GS penalty increase is small even for a small GS input power of −15 dBm. Typical eye diagrams are depicted in Fig. 3.10, showing only minor patterning effects for the largest tested GS input power of 0 dBm. Comparing the eye diagrams in the absence and presence of the ES signal, no additional signal degradation or patterning can be observed.

Hence, the GS penalty change in the presence of an ES signal is small, below 2 dB, compared to the BtB configuration for all tested currents. QD SOA based distortion-free amplification of 40 GBd OOK GS signal nearly independent of ES signals, currents and the GS input power levels is thus demonstrated.

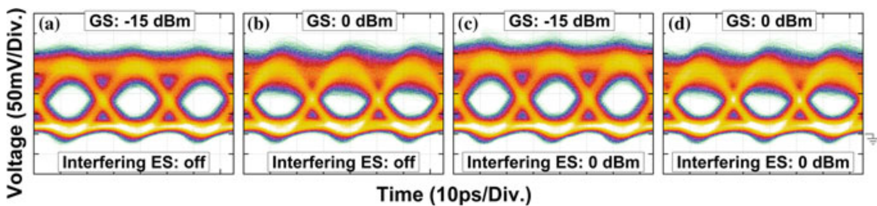


Fig. 3.10 Eye diagrams for the GS signal measured at a BER of 10^{-9} **a**, **b** in the absence and **c**, **d** in the presence of an ES signal. The GS input power and ES power is given in each diagram. The QD SOA current is set to 400 mA in all cases. Reprint of [34]

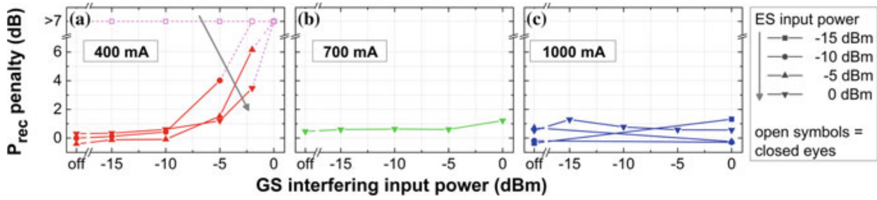


Fig. 3.11 Reveiver input power P_{rec} penalty of the 40 GBd OOK ES at a BER of 10^{-9} versus the QD SOA GS input power. The QD SOA currents are **a** 400 mA, **b** 700 mA and **c** 1000 mA. The results in the absence of the GS signal are plotted at a x-axis value labeled with "off". Measurements showing closed eye diagram are plotted with penalties labeled by ">7". The connections to the measured penalties are *dashed* and the color is brighter to improve the readability as well as to symbolize the limits

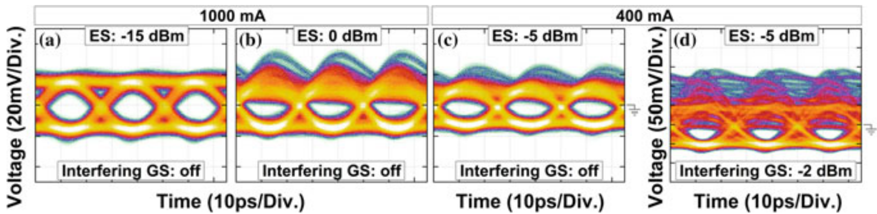


Fig. 3.12 Eye diagrams for the ES signal measured at a BER of 10^{-9} **a**–**c** in the absence and **d** in the presence of an interfering GS signal. The ES input power and the interfering input power is given in each diagram. The QD SOA current is set to **a**, **b** 1000 mA and **c**, **d** 400 mA

Figure 3.11 shows the penalties of the ES signal as a function of the input power of the GS signal. In the absence of the GS signal, the penalties are negligible at a current of 400 mA and ES input power between −10 and 0 dBm. For even smaller ES input power of −15 dBm no open eye was observed, caused probably by gain and noise limitations of the QD SOA at this unsaturated current in combination with the direct detection receiver, having a comparatively smaller sensitivity than the GS receiver. This limitation disappears at larger currents like 1000 mA, where the penalties are minuscule still for an input power of −15. ES input power levels of −5 dBm and 0 dBm are much larger than the 3 dB saturation input power, resulting in significant pattern effects at all currents present in the eye diagrams in Fig. 3.12c.

In the presence of a GS signal, the behavior becomes more complex. At a current of 400 mA, the ES penalty depends on the input power of both signals. The penalty for low GS input power levels of −15 and −10 dBm is comparable to the absence of a GS signal. In contrast, the penalty is strongly increasing for input power levels of −5 and −2 dBm due to the large XGM visible by the strong patterning in the eye diagrams, see Fig. 3.12d. This finally results in no useful BER results for a GS input power of 0 dBm. For increasing ES power the penalty increase is reduced, caused by the increasing GS power. Hence, at 400 mA a simultaneous amplification of both signals is only feasible for low GS input power levels up to −10 dBm

as expected from the gain measurements. This limitation is resolved with increasing current. At 700 mA, the ES penalty is nearly independent of the interfering GS input power and a large ES input power of 0 dBm and completely independent at 1000 mA, even for different ES input power levels from −15 dBm up to 0 dBm. The largest observed penalty change is found for the lowest ES input power of −15 dBm in combination with the largest interfering GS input power of 0 dBm. This combination represents the worst case scenario for the amplification, because the XGM probed by the ES is the strongest. Still for this combination however, the receiver power penalty is below 1.4 dB. The ES eye diagrams in the presence of a GS input power of 0 dBm are comparable to the eye diagrams in the absence of the GS signal at 1000 mA (Fig. 3.12a and b). Therefore, 40 GBd OOK ES signal amplification is independent of the presence of a GS signal as long as the QD SOA is driven at a current saturating the ES. Distortion-free amplification of the 40 GBd OOK ES signal is ensured up to ES input power levels of −5 dBm, which is larger than the 3 dB saturation input power.

Based on the above results a QD SOA current of around 700 mA is determined here as optimum point of operation. Modeling of the experiments predicts still larger ranges for the GS and ES input power and QD SOA currents, and discusses carrier dynamics and cross-gain effects [34, 49, 63]. The experimentally found optimum current of 700 mA could be verified by the model. In addition, input power levels of up to −7 dBm and −13 dBm for the GS and ES signal, respectively, are predicted as the optimum.

For the present optimal current of 700 mA, the fiber-to-fiber gain values of the GS and ES signals are 10 dB and 26 dB, respectively. In particular the GS gain could be increased by a longer device and by improved thermal managing. The signal quality evaluation demonstrates for both signals a penalty well below 2 dB for channel input power levels up to 0 dBm, independent of an interfering signal. The functionality of an ultra-broadband bidirectional dual-band amplifier using simultaneously QD GS and ES for amplification is thus demonstrated. Such an amplifier has the potential to drastically simplify the network architecture and reduce the power consumption also resulting in reduced investment and maintenance costs also for reach extended ANs and converged MANs.

3.5 Signal Processing—Wavelength Conversion

Fast processing optical data, such as switching and wavelength conversion, present the key for the implementation of all-optical AN and MAN networks. Signal processing in the optical rather than in the electrical domain offers numerous advantages like avoiding multiple optical-electrical conversions, enabling simpler systems, enhanced bandwidth and transparency to modulation formats [34, 39]. The most important requirements for nonlinear components performing these tasks are low cost, low power consumption and small footprint [8].

All-optical wavelength conversion (AOWC) describes the conversion of information from one wavelength channel to another. Optical demultiplexing [69, 70] and logical operations [71, 72] are based on AOWC. AOWC plays an equally important role for large capacity, highly branched reach extended networks to avoid signal blocking issues [73–75] in the course of the large density multi-wavelength channel transmission.

Various schemes of AOWC have been developed and realized in the past decades, [73, 76–78]. The vast majority of them are based on the interaction of two or more signals in a nonlinear device, such as highly nonlinear fibers, periodically poled lithium niobate, silicon waveguides, lasers and SOAs [17, 76, 79, 80]. SOAs combine the advantage of a small footprint, moderate to low optical input power requirements, large conversion efficiencies due to the simultaneous availability of gain, as well as the ease of integration with other components.

This section reviews [25, 32, 39, 81, 82] and starts with an introduction to non-linear effects in QD SOAs, particular four-wave mixing (FWM). Subsequently, roads to optimize the static performance of FWM in QD SOAs is presented. Finally, wavelength conversion of phase-coded signals is demonstrated.

3.5.1 Non-linearities of SOA Gain Media

Three different nonlinear effects in semiconductor gain materials are commonly employed for wavelength conversion: cross-gain modulation (XGM), cross-phase modulation (XPM) and four-wave mixing (FWM).

XGM is based on gain saturation by an amplitude modulated pump signal and the resulting gain compression. The resulting gain modulation can be probed by a weak continuous wave probe signal whose wavelength differs from the pump wavelength. The amplitude modulation of the pump signal is thus converted to the probe signal with an inverted bit logic. Since a change of the carrier density in the active material also changes the refractive index, XPM occurs simultaneously to XGM and vice versa. Hence, XPM in SOAs can be used to transfer information from the amplitude modulated pump signal to the probe signal using a phase sensitive setup [73, 76, 83]. XGM and XPM thus transfer the information from one input signal to a second input signal. In contrast FWM-based wavelength conversion generates new frequency components. The beating of two optical input signals in a gain medium generates a gain and index grating, which is modulated by the difference frequency of the two input signals as shown in Fig. 3.13a. The amplitude and/or phase modulation of the input signals is transferred to the new frequency components via modulation of the gain and index grating.

XGM, XPM and FWM have been investigated and compared in detail for various types of SOA gain materials in [39, 73]. XGM and XPM offer the advantage of a large conversion range, and especially for XGM simple configurations. XGM-based wavelength conversion has been demonstrated in SOAs for

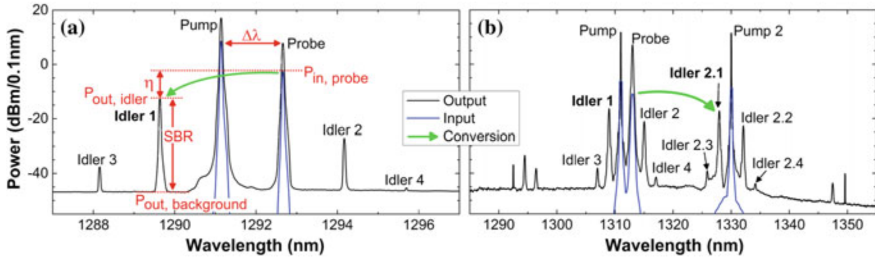


Fig. 3.13 Typical input and output spectra measured for FWM **a** single-pump down-conversion and **b** dual-pump up-conversion in a QD SOA. New frequency components are called "idler x". The main parameters of FWM are detuning Δλ, conversion efficiency η and signal-to-background ratio (SBR)

OOK modulated signals up to 320 Gbs [27, 30, 84]. However, the inability to convert phase-coded signals and the limited extinction ratios are drawbacks.

FWM-based conversion offers a very large bandwidth, high extinction ratio, amplitude and phase sensitivity, inversion of the spectral phase, as well as a large conversion range, if a multiple-pump configuration is used [39, 73, 85]. The conversion efficiency and optical signal-to-noise ratio (OSNR) are often mentioned to be limited, but can be improved to positive conversion efficiencies and OSNRs larger than 25 dB [85]. FWM in bulk, quantum-well QW and QD SOAs has been demonstrated for various modulation formats and data rates up to 320 GBd [86–91].

The subsequent sections focus on FWM-based wavelength conversion in QD SOAs for phase-coded signals. An investigation of similar structures with OOK signals has been presented in [25, 32, 39].

3.5.2 *Four-Wave Mixing in QD SOAs*

Beating of the pump and probe signal in again medium causes a modulation of the carrier density with the beating frequency Δω. This modulation generates a dynamic gain and index grating being responsible for the generation of new frequency components. If the probe and/or pump signal carries an amplitude or phase modulation, the grating is modulated similarly and consequently the beat signals. In Sect. 3.2.2 we showed the carrier dynamics in semiconductor gain media being mainly determined by SHB, CDP and CH. Partly degenerate FWM is thus controlled mainly by these processes for detuning up to a few THz contributing independently. The third-order susceptibility can be expressed as a sum of the contributions of the individual processes [92, 93]. Other processes, like two photon absorption and the Kerr effect are important for larger detuning [94].

FWM-based AOWC transfers the information from the probe signal to the target idler at a different wavelength. Multiple idlers are generated by the nonlinear

interaction of the corresponding signals, resulting among others in difference-frequency components. FWM in general is based on the interaction of three fields generating a fourth frequency component. For fully degenerate FWM, all three frequency components are identical the fourth one. For partly degenerated FWM described here, two frequencies are identical (the pump signal, larger intensity) whereas the third frequency is different (the probe signal). The exemplary spectrum in Fig. 3.13a shows four FWM idlers. Idler 1 and idler 2 are caused by mixing the pump and probe signals. Idler 3 respectively idler 4 are caused by mixing pump and probe in addition with idler 1 respectively idler 2. The output power of the idlers depends among others on the power of the signals mixing with each other. Based on the simple model described in [95], the idler 1 (idler 2) output power is proportional to the cube of the gain, to the square of the pump (probe) input power and linear to the probe (pump) input power, but decreases with the frequency detuning of pump and probe signals [34]. Additionally, the sign of the α-factor (the linewidth enhancement factor) of the individual processes determine how the corresponding third-order susceptibilities contributions are adding up. In most common configurations, the probe signal carries the information and exhibits a smaller amplitude than the pump signal. Idler 1 is then more intense than idler 2 and thus typically used as the wavelength converted signal.

Experimental and theoretical investigations demonstrated that conventional gain materials, such as bulk and QW, exhibit a strongly asymmetric FWM characteristic with respect to the sign of the frequency detuning [92, 96, 97]. For positive detuning ($\Delta\omega > 0$) all three components of the third-order susceptibilities have the same sign and add up. For negative detuning ($\Delta\omega < 0$), the third-order susceptibilities of CDP and CH interfere destructively with the third-order susceptibilities of SHB within the bandwidth of SHB. Consequently, the FWM products are weaker in power for negative than for positive detuning.

Gain materials with small or zero α-factor, like QDs, allow a more symmetric FWM conversion with respect to the sign of the detuning [41, 51]. Hence, the existence of a huge carrier reservoir enables the fast carrier dynamics and therefore a reduction of the influence of CDP. Additionally, QD gain materials exhibit a spectrally broader and improved conversion efficiency due to the fast time constants and broader gain spectra [41, 51]. FWM based conversion for detuning energies larger than the homogenous linewidth of the QD states has to be mediated via the carrier reservoir. The gain depletion of a specific QD sub-ensemble caused by the input pump and probe signal recovers via Auger scattering from the reservoir to the QD states. The reservoir CH is modulated with the detuning frequency $\Delta\omega$ and thus the carrier capture into all QD sub-ensembles is modulated.

3.5.2.1 Definition of Parameters

FWM in active materials is characterized by two main parameters, the FWM conversion efficiency η and the signal-to-background ratio (SBR) [94]. Both are depicted in Fig. 3.13a and depend on the frequency detuning of the input signals.

The frequency detuning is defined as the difference between the frequency of the pump and the probe signal $\Delta\omega = \omega_{pump} - \omega_{probe}$. Idler 1 is than generated at a frequency $\omega_{idler\ 1} = \omega_{pump} + \Delta\omega$. A positive and negative frequency detuning refers to a frequency up-conversion (towards larger frequency) and down-conversion (towards smaller frequency), respectively. The wavelength detuning is defined accordingly as $\Delta\lambda = \lambda_{PROBE} - \lambda_{PUMP}$. Hence, a positive wavelength detuning refers similar to the positive frequency detuning to a frequency up-conversion representing a wavelength down-conversion.

The FWM conversion efficiency η and the SBR of the FWM-converted idler are defined as:

$$\eta[dB] = 10\log_{10}\frac{P_{out,\,idler}[mW]}{P_{in,\,probe}[mW]}, \quad SBR[dB] = 10\log_{10}\frac{P_{out,\,idler}[mW]}{P_{in,\,background}[mW]},$$

with the $P_{out,\ background}$: power of the background noise measured at the wavelength of the converted idler within a certain optical bandwidth [94].

3.5.2.2 Dual-Pump FWM

The spectral detuning range in which single-pump FWM-based wavelength conversion exhibits a sufficient conversion efficiency is limited. Typically, the FWM efficiency is reduced at least by 5–15 dB per 5 nm increase of the detuning. An improvement of the conversion range can be achieved by using a second pump signal [32, 85, 89, 98]. The beating of the first pump and probe signal generates a gain and index grating modulated by Δω via the modulation of the carrier distribution in the active material as described above. The additionally injected second pump signal probes the phase and gain gratings. Hence, additional FWM idlers (called idler 2.x) are generated around the second pump with a frequency spacing of Δω to first order with respect to this second pump, see Fig. 3.13b. Similar to the single-pump scheme, idler 2.1 is the strongest and will be considered here. The weak group of peaks visible at both edges of the spectrum are generated by direct mixing of pump 2 and pump 1 as well as the probe.

The spectral position of pump 2 can be changed independently of the single-pump configuration as long as idler 2.1 is not overlapping with one of the single-pump idlers. The spectral position of idler 2.1 relative to pump 2 depends on the single-pump configuration. It will remain on the lower wavelength side of pump 2 for a down-conversion configuration of the single-pump configuration, independent of the down- or up-conversion configuration of pump 2 with respect to pump 1. The minimum conversion span in the dual-pump scheme is therefore typically slightly asymmetric per default for a given single-pump configuration.

3.5.3 Optimization of Static Four-Wave Mixing in QD SOAs

The fiber-to-fiber FWM efficiency and SBR in QD SOAs as a function of the most important parameters is discussed in what follows. In all configurations, the polarization of the optical input signal is TE, to ensure maximum gain.

3.5.3.1 Detuning

FWM results for a single-pump configuration of the 3 mm long QD SOA No. 2 (see Table 3.1) are shown in Fig. 3.14. The continuous wave pump wavelength is set close to the ASE peak, and the continuous wave probe wavelength is scanned. The pump and probe input power levels are set to 11 dBm and 2 dBm, respectively.

The converted idler 1 exceeds the noise floor within a detuning range of more than 50 nm for a current of 650 mA. The wavelength up-conversion (positive detuning) is less than 5 dB more efficient, i.e. very symmetric. Increasing the current beyond SOA gain saturation leads to disappearance of the asymmetry at large detuning. In addition, larger current also increases the conversion efficiency. The maximum efficiency is observed to be −9 dB for a detuning of +0.5 nm and a current of 650 mA. This corresponds to a chip conversion efficiency of 0 dB.

The SBR shows a slight asymmetric behavior, similar to the efficiency. At a large operating current, the asymmetry also disappears at large detuning. A maximum SBR of 32 dB is found for a detuning of +0.5 nm and an operating current of 650 mA. For this current, the SBR exceeds 20 dB within a detuning range of 13.5 nm from −4.5 to +9 nm. The conversion efficiency is larger than −26 dB within this range. Hence, the conversion efficiency and SBR in QD SOAs are more symmetric for both up and down conversion [39, 41, 99, 100], demonstrating an important advantage of QD SOAs.

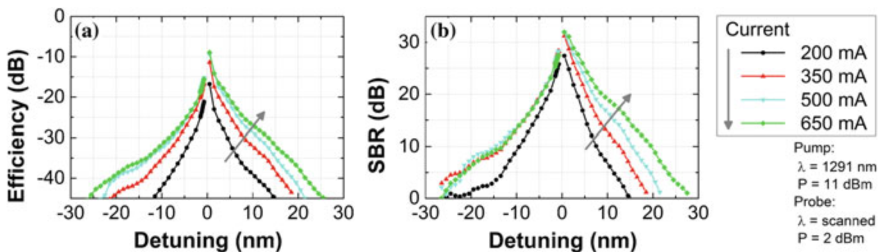


Fig. 3.14 FWM **a** efficiency and **b** SBR of FWM idler 1 as a function of the detuning for different QD SOA currents. The wavelength and input power of pump and probe signal are given in the figure. Both figures share the same legend. The *arrow* shows the change of current according to the arrow in the legend. Reprint of [34]

3.5.3.2 Input Power

FWM conversion efficiency and SBR are shown in Fig. 3.15 in dependence of in-fiber pump and probe input power levels for a detuning of +1 nm and an operating current of 650 mA (QD SOA No. 2).

Both, efficiency and SBR, follow the power increase first linearly and start to saturate at a power of around 1 dBm. The efficiency is close to saturation and reaches a value of about −8.5 dB at a pump power of 11 dBm. Increasing the probe power by 10 dB results in an efficiency change of less than 1.5 dB. The probe power has thus only a small influence on the conversion efficiency. In contrast, the SBR is significantly increased by about 4.6 dB for a power increase by a factor of 5. Larger input power improves both, the conversion efficiency and SBR. For moderate and large input probe power of −3 and 2 dBm, the SBR exceeds 20 dB for pump input power larger than 3 dBm and 0 dBm, respectively. Under these conditions the conversion efficiency is larger than −15 dB and −21 dB, respectively.

3.5.3.3 Gain via QD SOA Length

The FWM conversion efficiency and SBR strongly depend on the QD SOA gain and thus the length of the SOA. Linear gain values of 15.6 dB, 23.3 dB and 26.6 dB were obtained for the 3 mm, 4 mm and 5 mm long QD SOAs No. 3, 4 and 5, respectively, all driven at a current density of 1.67 kA/cm^2. The improvements of the FWM efficiency and SBR with increasing SOA length are shown in Fig. 3.16. The pump wavelength is chosen for each SOA at its ASE peak. The in-fiber pump and probe power levels are 2 and −8 dBm in all configurations. Hence, the SOAs with larger lengths are driven in deeper gain saturation due to the reduced saturation input power levels (compare Table 3.1).

The efficiency is increased at least by 2.2 and 3.5 dB within a 20 nm detuning range, comparing the 3 mm long SOA with the 4 and 5 mm long one. The

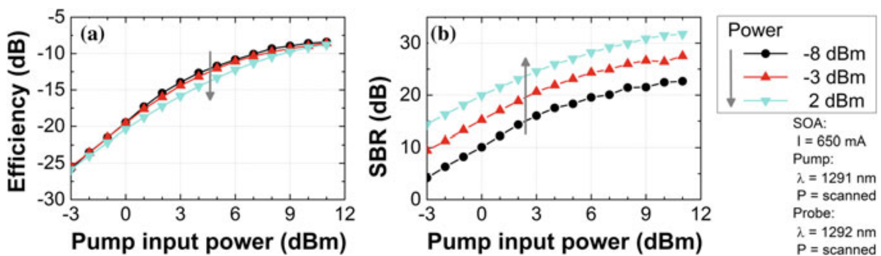


Fig. 3.15 FWM down-conversion **a** conversion efficiency and SBR as a function of the optical pump input power for three different probe power levels. The pump and probe wavelengths as well as SOA current are shown in the figure. Both figures share the same legend. The arrows show the change of probe power Reprint of [34]

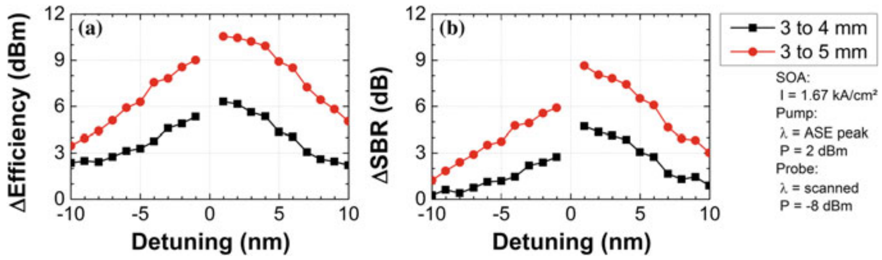


Fig. 3.16 FWM **a** efficiency and **b** SBR improvement for increasing SOA length in dependence of the detuning. The wavelength and input power of pump and probe signal as well as the SOA current density are given in the figure. Both figures share the same legend. Reprint of [34]

asymmetry of the up and down-conversion efficiency increases with increasing SOA length. A peak efficiency improvement of 10.5 dB is obtained for the 5 mm long SOA for a wavelength down-conversion of 2 nm (detuning +1 nm).

A slight increase of the SBR is observed for large detuning, probably due to the gain bandwidth narrowing caused by ASE self-saturation of longer devices. However, the 5 mm long SOA offers an SBR improvement of more than 3 dB within a detuning range of 16.9 nm from −6.9 to +10 nm.

The FWM conversion efficiency and SBR significantly benefit from both, the larger gain and deeper saturation with increasing SOA length and for given input power levels. Hence, the 5 mm long SOA is most suitable for wavelength conversion with high efficiency and SBR.

3.5.3.4 Detuning Dependence for Dual-Pump Configuration

The detuning range in which the SBR exceeds 20 dB using the single-pump scheme is limited to about 15 nm (e.g. from −5 to +10 nm). Hence, the static wavelength conversion is limited to less than −10 to +20 nm. A dual-pump scheme provides an extension of this range.

Dual-pump FWM is investigated using 5 mm long QD SOA No. 5 (see Table 3.1) driven by a current of 900 mA. The pump 1 and probe wavelengths are set to 1311 and 1313 nm, which are on the short wavelength side of the ASE peak observed at 1319 nm. The wavelength of pump 2 is scanned. The pump 1, pump 2 and probe input power levels are 15 dBm, 13 dBm and 10 dBm, respectively. Idler 2.1 is generated with a 2 nm detuning on the short wavelength side of pump 2 and follows its wavelength scan. Meanwhile, idler 1 remains at a 4 nm down-converted wavelength of 1309 nm since the wavelengths of pump 1 and probe are fixed. The FWM efficiency, SBR and peak power of idler 2.1 and idler 1 are depicted in Fig. 3.17 in dependence of the wavelength of idler 2.1. The figure exhibits in addition a detuning [34] $\overline{\Delta\lambda} = 0.5\left(\lambda_{\text{probe}} - \lambda_{\text{idler 2.1}}\right)$.

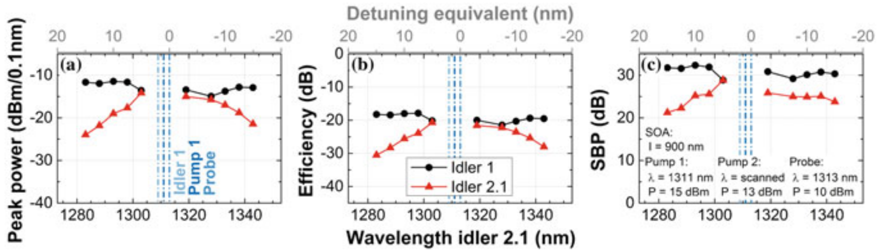


Fig. 3.17 FWM **a** peak power, **b** efficiency and **c** SBR of idler 1 and idler 2.1 in dependence of the wavelength of idler 2.1. The fixed wavelength position of pump 1, probe and idler 1 are marked with the vertical lines (*dash-dot*) in the center of each graph. A detuning equivalent given by half the probe-idler 2.1 wavelength difference is plotted as the top x-axis in all three graphs to simplify the comparison to the single-pump detuning dependence e.g. shown in Fig. 3.14 (*Note* the axis direction is inverted). All figures share the same legend. Reprint of [34]

The efficiency and SBR of idler 1 are nearly unaffected by the position of pump 2. Only at very small positive detuning, both values show a slight decrease possibly due to some cross-talk. This could be induced by direct interaction among the two pumps or the large number of additional FWM products and hence the reduced gain.

The idler 2.1 up and down-conversion efficiency is comparable for the smallest tested detuning equivalent of 5 nm. Both, the up and down-conversion efficiency, are reduced with increasing detuning, but the up-conversion remains at first more efficient than the down-conversion. The larger gain on the up-conversion side for smaller detuning indicated by the ASE peak can explain this result. Both conversion efficiencies show only a difference of 2.5 dB for the largest tested detuning, equivalent to ±15 nm. The dual-pump up and down conversion is thus much more symmetric than in a single-pump configuration, but is not significantly improved for down-conversion. In contrast, the idler 2.1 SBR is significantly improved for both, up and down-conversion, and is in addition more symmetric than for the single-pump case. The SBR of idler 2.1 exceeds 20 dB in the entire detuning equivalent range of 30 nm (from −15 to +15 nm). The conversion efficiency exceeds −30 dB in nearly the entire detuning equivalent range. The converted idler 2.1 peak power exceeds −30 dBm (−24 dBm) required for an error-free detection of differential (quadrature) phase-shift keying (D(Q)PSK) signals. Hence, a dual-pump scheme enables conversion distances across the entire spectral gain bandwidth, i.e. a wavelength range of up to 60 nm around pump 1.

3.5.4 FWM of D(Q)PSK Signals

Wavelength conversion of OOK signals at large symbol rates has been achieved using FWM in QD SOAs [25, 39]. Future optical networks will require additionally

modulation format transparent wavelength converters, since these networks will use both, amplitude and phase-coded signals. FWM-based wavelength conversion of phase-coded signals will require the generation of a sufficiently fast phase grating similar to the fast gain grating used for OOK modulated signals. Conversion of up to DPSK and QPSK modulated signals up to 40 GBd and 25 GBd, respectively, using QD SOAs has been demonstrated in [90, 101, 102]. Here we demonstrate the conversion of phase-coded signals at large symbol rates of 40 GBd for DPSK and DQPSK for single- and dual-pump configurations.

The D(Q)PSK transmitter generates a 40 GBd D(Q)PSK probe signal with a wavelength of 1313 nm using a modulator cascade. A dual-drive Mach-Zehnder modulator generates first a DPSK signal and a subsequent phase modulator transforms it into a D(Q)PSK signal. The modulators are driven with an electrical 40 Gbs PRBS $2^7 - 1$ bit sequence, limited by the modulator characteristics. The pump 1 is set to a wavelength of 1311 nm. Probe, pump 1 and pump 2 are combined and injected to the QD SOA with their polarization aligned to the TE axis. The power level of each of the three signals, as well as the wavelength of pump 2, are adjustable. A filter following the SOA is used to separate the signal to be measured from others (compare Fig. 3.13). The blocked signals are suppressed by more than 30 dB with respect to the signal to be measured. The 40 GBd D(Q)PSK signal is detected with a pre-amplified differential detection receiver based on a delay interferometer and a balanced detector. The signal quality is evaluated by BER measurements without offline processing and compared to the back-to-back (BtB) configuration, provided by bypassing the device under test. For DQPSK, the BER was measured twice, with a phase difference of 90° of the interferometer. The retrieved errors are added up. Setup details can be found in [34].

The device under test is the 5 mm long QD SOA No. 5 (see Table 3.1). The SOA is driven with a current of 900 mA and the ASE peak appears at a wavelength of around 1319 nm. At a wavelength of 1310 nm, the linear gain as well as a 3 dB saturation input are 21.5 dB and −3.6 dBm, respectively.

3.5.4.1 Single-Pump FWM

A 4 nm wavelength down-conversion to 1309 nm is considered here for a single-pump configuration, i.e. in absence of the pump 2. The receiver power required for a BER of 10^{-9} is −30.4 dBm (−21.6 dBm) for the D(Q)PSK modulated signal.

At a BER threshold of 10^{-9}, the wavelength converted DPSK signal shows a slightly improved receiver power −30.6 dB presumably caused by a marginal change of the receiver sensitivity upon the wavelength shift of 4 nm. Hence, the QD SOA enables an error-free wavelength down-conversion of a 40 GBd DPSK signal by 4 nm without any signal degradation.

The wavelength converted DQPSK signal shows a receiver power penalty below 1.6, i.e. a larger receiver input power is required for a BER threshold of 10^{-9}, which could be caused by the modulator cascade characteristics [34]. The 40 GBd D(Q)

PSK signal is converted by 4 nm from 1313 to 1309 nm with a receiver power penalty well below 2 dB.

3.5.4.2 Dual-Pump FWM

In contrast to single-pump FWM, dual-pump wavelength conversion is possible for a huge spectral range up to 60 nm. In such experiments, the wavelength dependence of the setup, of the transmitter and the receiver, must be taken into account. BER measurements show only the wavelength dependence of the coupled transmitter-receiver system, but not the dependence of the individual parts. The latter would be required because the wavelength converter leads to different wavelength of receiver and transmitter. Hence, one cannot distinguish between penalty changes caused by the wavelength converter and the system. To get an idea of the wavelength dependence of the system, two other BtB configurations are recorded, by changing the probe wavelength to 1288 nm and 1323 nm, respectively. The receiver input power levels required for BER thresholds of 10^{-9} and 10^{-3} are listed in Table 3.2. Figure 3.18 depicts the penalties for 40 GBd D(Q)PSK wavelength conversion as a function of the wavelength of idler 2.1.

An error-free 40 GBd DPSK wavelength conversion is observed for a wavelength range of 45 nm from 1283 to 1328 nm. A conversion up to 1343 nm, as demonstrated for the static wavelength conversion is limited by the wavelength dependent receiver sensitivity due to the gain bandwidth of the fiber amplifiers. The penalties for the wavelength down-converted signal obtained at a BER of 10^{-9} range from −0.6 to +1.8 dB, using the 1313 nm BtB measurement as reference. In contrast, the penalties for the wavelength up-converted signal ranges from +0.8 to +6.1 dB. Using the references localized within the corresponding wavelength range, the penalties increase by 1.2 dB and decrease by 1.6 dB for the down and up-conversion, respectively. Hence, the wavelength dependence of the transmitter-receiver system, particularly the sensitivity of the receiver at large wavelengths, slightly limit the conversion range.

40 GBd DPSK wavelength conversion is achieved within a range of 40 nm (1283–1323 nm) with a penalty below 3 dB for both BER thresholds. This range can be extended to 45 nm for lower penalty requirements. The receiver sensitivity limits the long wavelength side representing wavelength up-conversion. Based on the static FWM properties, in particular the peak power and SBR, down-conversion

Table 3.2 List of modulation format dependent receiver input power of the differential detection receiver measured in BtB configuration for different wavelengths

Modulation type	BER threshold	Receiver input power		
		1288 nm	1313 nm	1323 nm
		dBm	dBm	dBm
DPSK	10^{-9} 10^{-3})	−31.6 (−37.7)	−30.4 (−36.6)	−28.8 (−34.4)
DQPSK	10^{-9} (10^{-3})	−23.6 (−32.3)	−21.6 (−30.4)	−21.2 (−29.4)

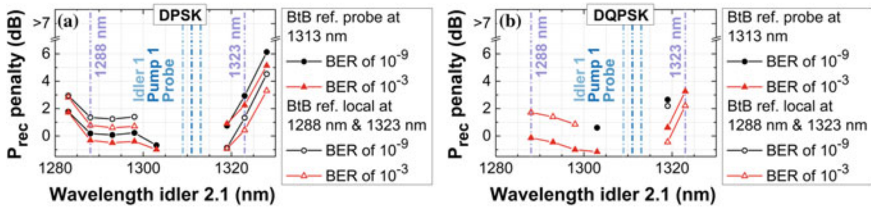


Fig. 3.18 Receiver power penalty versus the wavelength of the converted idler 2.1 for **a** the 40 GBd DPSK and **b** the 40 GBd DQPSK modulated signal. The penalty is evaluated with the probe BtB measurement (wavelength at 1313 nm) as a reference across the entire converted wavelength span at two different BER thresholds (*closed symbols*). In addition, the penalty is also evaluated for the short and long wavelength region using the 1288 nm and 1313 nm BtB measurement as a reference, respectively (*open symbols*). The fixed wavelength position of pump 1, probe and idler 1 as well as the alternative BtB signals are marked by the vertical lines (*dash-dot*). Reprint of [34]

within the entire static conversion range of 60 nm seems to be feasible, using a receiver with an improved spectral bandwidth. The large conversion range of at least 30 nm would be symmetric for up and down conversion.

The detection of the converted 40 GBd DQPSK signal suffers from the low receiver sensitivity described above for DPSK. Its impact on the DQPSK signal is even worse, since the power budget is significantly smaller than DPSK. However, an error-free BER of 10^{-9} can be reached for wavelength up and down-conversion to 1319 nm and 1303 nm, respectively, only limited by the available receiver power. The penalties are below 3 dB.

The 40 GBd DQPSK signal can be converted within a range of 35 nm for a BER of 10^{-3}. This range is again limited on the large wavelength side by the receiver sensitivity. The observed penalties vary within a range of −1.2 to 3.3 dB and −0.4 to 2.2 dB, using the 1313 nm BtB and the localized BtB measurements as references, respectively. The penalties of the converted DPSK and DQPSK show a comparable performance at a BER of 10^{-3}. The large spectral range observed for DPSK signals is limited in case of the DQPSK by both, the receiver sensitivity and the reduction of the output peak power of idler 2.1 with increasing conversion distance.

Still DQPSK wavelength conversion is achieved at 40 GBd within a range of 16 nm (1319–1303 nm) and 35 nm (1288–1323 nm) with a penalty below 3 dB for BER of 10^{-9} and 10^{-3}. An improved receiver sensitivity could enable DQPSK wavelength conversion across a wavelength range as large as that obtained for the DPSK signal, since the static SBR should be sufficient.

The fast gain dynamics of QD SOAs even in deep saturation ensures a sufficiently fast modulation of the gain and refractive index grating. Thus, wavelength conversion of high symbol rate OOK signals is feasible [25, 39, 85]. We demonstrated here sufficiently fast modulation of the phase grating enabling the conversion of high symbol rate phase-coded signals, i.e. 40 GBd D(Q)PSK signals.

3.6 Summary

QD based SOAs are enabling devices for cost and energy efficient up and down stream multiple wavelength division multiplexing in access and metropolitan area networks. Slow interaction between carriers bound to separate ensembles of quantum dots or in ground and excited states, together with fast gain recovery of QDs by down scattering from carrier reservoirs at much higher energies are unique for quantum dots. These unique properties enable multi-channel dual-band amplification free of cross talk.

Equally important for applications is the decoupling of the physical origin of gain and phase in amplifiers. The gain is governed by the QD population, which easily can be driven in inversion, whereas the phase is controlled by the carrier reservoir. The ultrafast gain dynamics of QD SOAs even in deep saturation enables fast modulation of the gain and refractive index. Thus the generation of phase coded signals by direct modulation of QD SOAs is possible up to very large frequencies.

Much beyond that, nonlinear applications like fast conversion of large symbol rate phase-coded signals, i.e. 40 GBd D(Q)PSK signals, are enabled by the specific nature of QDs. The wavelength conversion ranges are presently only limited by the wavelength dependent receiver sensitivity and can be increased to at least 60 nm (35 nm) for D(Q)PSK signals. Hence, QD SOAs enable wavelength conversion of 40 GBd and more phase-coded signals.

QD SOAs will become game breaking devices for linear applications as well as for energy efficient all optical wavelength conversion, required for signal processing in modulation format transparent future ANs and MANs.

Acknowledgements The authors like to thank the following persons for the contribution to the results presented in this chapter: D. Arsenijević[a], R. Bonk[c], C.-A. Bunge[e], G. Fiol[a], G. Eisenstein[d], S. König[c], C. Meuer[a,b], C. Schmidt-Langhorst[b], C. Schubert[b], M. Stubenrauch[a], A. Zeghuzi[a]

[a]At this time with Department of Solid-State Physics, Technische Universitaet Berlin (TUB)

[b]At this time with Heinrich Hertz Institute (HHI), Fraunhofer Institute for Telecommunications

[c]At this time with Karlsruhe Institute of Technology (KIT)

[d]At this time with Technion

[e]At this time with Hochschule fuer Telekommunikation Leipzig (HfTL)

References

1. Cisco Systems Inc., Cisco visual networking index: forecast and methodology, 2012–2017, in *WHITE PAPER* (2013)
2. Cisco Systems Inc., Cisco visual networking index: forecast and methodology, 2014–2019, in *WHITE PAPER* (2015)
3. J.M. Senior, M.Y. Jamro, *Optical Fiber Communications: Principles and Practice*, 3rd edn (Financial Times/Prentice Hall, 2009)
4. Corning Inc., Data sheet: corning SMF-28e+ photonic optical fiber (2010)

5. P. Vetter, Next generation optical access technologies, in *European Conference and Exhibition on Optical Communication (ECOC)* (2012), pp. 1–42
6. K.C. Reichmann, P.P. Iannone, C. Brinton, J. Nakagawa, T. Cusick, M. Kimber, C. Doerr, L.L. Buhl, M. Cappuzzo, E.Y. Chen, L. Gomez, J. Johnson, A.M. Kanan, J. Lentz, Y.F. Chang, B. Pálsdóttir, T. Tokle, L. Spiekman, A symmetric-rate, extended-reach 40 Gb/s CWDM-TDMA PON with downstream and upstream SOA-Raman amplification. J. Light. Technol. **30**(4), 479–485 (2012)
7. D. Nesset, NG-PON2 technology and standards. J. Light. Technol. **33**(5), 1136–1143 (2015)
8. R. Bonk, Linear and nonlinear semiconductor optical amplifiers for next-generation optical networks, Doctoral Thesis, KIT Scientific Publishing (2013)
9. N. Suzuki, S. Yoshima, J. Nakagawa, Extended reach bidirectional optical amplified GE-PON with a high 46 dB span-budget for 64 far-end user ONUs, in *Joint International Conference on the Optical Internet (COIN) and Australian Conference on Optical Fibre Technology (ACOFT)* (2007), pp. 1–3
10. T. Pfeiffer, Converged heterogeneous optical metro-access networks, in *European Conference and Exhibition on Optical Communication (ECOC)* (2010), p. Tu.5.B.1
11. D.R. Zimmerman, L.H. Spiekman, Amplifiers for the masses: EDFA, EDWA, and SOA amplets for metro and access applications. J. Light. Technol. **22**(1), 63–70 (2004)
12. L. Spiekman, D. Piehler, P. Iannone, K. Reichmann, H.-H. Lee, Semiconductor optical amplifiers for FTTx, in *International Conference on Transparent Optical Networks (ICTON)*, vol. 2 (2007), p. Mo.D2.4
13. P.P. Iannone, H.H. Lee, K.C. Reichmann, X. Zhou, M. Du, B. Palsdottir, K. Feder, P. Westbrook, K. Brar, J. Mann, L. Spiekman, Four extended-reach TDM PONs sharing a bidirectional hybrid CWDM amplifier. J. Light. Technol. **26**(1), 138–143 (2008)
14. P.P. Iannone, K.C. Reichmann, C.Brinton, J. Nakagawa, T. Cusick, E.M. Kimber, C. Doerr, L.L. Buhl, M. Cappuzzo, E.Y. Chen, L. Gomez, J. Johnson, A.M. Kanan, J. Lentz, Y.F. Chang, B. Pálsdóttir, T. Tokle, L. Spiekman, Bi-directionally amplified extended reach 40 Gb/s CWDM-TDM PON with burst-mode upstream transmission, in *Optical Fiber Communication Conference (OFC) and National Fiber Optic Engineers Conference (NFOEC)* (2011), p. PDPD6
15. H. Schmuck, Breitband-Zugangsnetze der nächsten Generation Verbundprojekt: CONDOR —Converged Heterogeneous Metro/Access Infrastructure (Project Report) (2013)
16. K. Taguchi, H. Nakamura, K. Asaka, T. Mizuno, Y. Hashizume, T. Yamada, M. Ito, H. Takahashi, S. Kimura, and N. Yoshimoto, 40-km reach symmetric 40-Gbit/s λ-tunable WDM/TDM-PON using synchronized gain-clamping SOA, in *Optical Fiber Communication Conference (OFC) and National Fiber Optic Engineers Conference (NFOEC)* (2013), p. OW4D.6
17. A.E. Willner, O.F. Yilmaz, J. Wang, X. Wu, A. Bogoni, L. Zhang, S.R. Nuccio, Optically efficient nonlinear signal processing. IEEE J. Sel. Top. Quantum Electron. **17**(2), 320–332 (2011)
18. M. Asada, Y. Miyamoto, Y. Suematsu, Gain and the threshold of three-dimensional quantum-box lasers. IEEE J. Quantum Electron. **22**(9), 1915–1921 (1986)
19. D. Bimberg, M. Grundmann, N.N. Ledentsov, *Quantum Dot Heterostructures*, 1st edn. (Wiley, 1998)
20. D. Bimberg, Quantum dot based nanophotonics and nanoelectronics. Electron. Lett. **44**(3), 168–170 (2008)
21. M. Sugawara, N. Hatori, M. Ishida, H. Ebe, Y. Arakawa, T. Akiyama, K. Otsubo, T. Yamamoto, Y. Nakata, Recent progress in self-assembled quantum-dot optical devices for optical telecommunication: temperature-insensitive 10 Gb/s directly modulated lasers and 40 Gb/s signal-regenerative amplifiers. J. Phys. D Appl. Phys. **38**(13), 2126–2134 (2005)
22. T. Akiyama, M. Sugawara, Y. Arakawa, Quantum-dot semiconductor optical amplifiers. Proc. IEEE **95**(9), 1757–1766 (2007)

23. G. Contestabile, A. Maruta, S. Sekiguchi, K. Morito, M. Sugawara, K.-I. Kitayama, Regenerative amplification by using self-phase modulation in a quantum-dot SOA. IEEE Photonics Technol. Lett. **22**(7), 492–494 (2010)
24. C. Schmidt-Langhorst, C. Meuer, A. Galperin, H. Schmeckebier, R. Ludwig, D. Puris, D. Bimberg, K. Petermann, C. Schubert, 80 Gb/s multi-wavelength booster amplification in an InGaAs/GaAs quantum-dot semiconductor optical amplifier, in *European Conference and Exhibition on Optical Communication (ECOC)* (2010), p. Mo.1.F.6
25. H. Schmeckebier, C. Meuer, D. Bimberg, C. Schmidt-Langhorst, A. Galperin, C. Schubert, Quantum dot semiconductor optical amplifiers at 1.3 μm for applications in all-optical communication networks. Semicond. Sci. Technol. **26**(1), 14009 (2011)
26. R. Bonk, T. Vallaitis, J. Guetlein, C. Meuer, H. Schmeckebier, D. Bimberg, C. Koos, W. Freude, J. Leuthold, The input power dynamic range of a semiconductor optical amplifier and its relevance for access network applications. IEEE Photonics J. **3**(6), 1039–1053 (2011)
27. G. Contestabile, A. Maruta, S. Sekiguchi, K. Morito, M. Sugawara, K. Kitayama, 160 Gb/s cross gain modulation in quantum dot SOA at 1550 nm, in *European Conference and Exhibition on Optical Communication (ECOC)* (2009) pp. 1–2
28. G. Contestabile, A. Maruta, S. Sekiguchi, K. Morito, M. Sugawara, K.-I. Kitayama, Cross-gain modulation in quantum-dot SOA at 1550 nm. IEEE J. Quantum Electron. **46**(12), 1696–1703 (2010)
29. G. Contestabile, A. Maruta, S. Sekiguchi, K. Morito, M. Sugawara, K.-I. Kitayama, All-optical wavelength multicasting in a QD-SOA. IEEE J. Quantum Electron. **47**(4), 541–547 (2011)
30. C. Meuer, C. Schmidt-Langhorst, R. Bonk, H. Schmeckebier, D. Arsenijević, G. Fiol, A. Galperin, J. Leuthold, C. Schubert, D. Bimberg, 80 Gb/s wavelength conversion using a quantum-dot semiconductor optical amplifier and optical filtering. Opt. Express **19**(6), 5134–5142 (2011)
31. M. Matsuura, O. Raz, F. Gomez-Agis, N. Calabretta, H.J.S. Dorren, 320-Gb/s wavelength conversion based on cross-gain modulation in a quantum-dot SOA, in *European Conference and Exhibition on Optical Communication (ECOC)* (2011), p. Mo.1.A.1
32. C. Meuer, C. Schmidt-Langhorst, H. Schmeckebier, G. Fiol, D. Arsenijević, C. Schubert, D. Bimberg, 40 Gb/s wavelength conversion via four-wave mixing in a quantum-dot semiconductor optical amplifier. Opt. Express **19**(4), 3788–3798 (2011)
33. M. Matsuura, O. Raz, F. Gomez-Agis, N. Calabretta, H.J.S. Dorren, 320–40-Gb/s optical demultiplexing by means of optical filtering of chirped signal using a quantum-dot SOA, in *Optical Fiber Communication Conference (OFC) and National Fiber Optic Engineers Conference (NFOEC)* (2012), p. OTh3H.4
34. H. Schmeckebier, *Quantum-Dot-Based Semiconductor Optical Amplifiers for O-Band Optical Communication* (Springer Science and Business Media, 2016)
35. Internatinal norm IEC 61290-1-1:1998, optical spectrum analyzer
36. H. Kawaguchi, Semiconductor lasers and optical amplifiers for switching and signal processing, in *Handbook of Laser Technology and Applications: Laser Design and Laser Systems*, vol. 2, 1st edn., ed. by C.E. Webb, J.D. C. Jones (CRC Press, 2004)
37. R. Bonk, T. Vallaitis, W. Freude, J. Leuthold, R.V. Penty, A. Borghesani, I.F. Lealman, Linear semiconductor optical amplifiers, in *Fibre Optic Communication: Key Devices*, 1st edn., ed. by H. Venghaus, N. Grote (Springer Science & Business Media, 2012), p. 682
38. J. Mørk, M.L. Nielsen, T.W. Berg, The dynamics of semiconductor optical amplifiers: modeling and applications. Opt. Photonics News **14**(7), 42–48 (2003)
39. C. Meuer, GaAs-based quantum-dot semiconductor optical amplifiers at 1.3 μm for all-optical networks, Doctoral Thesis, Mensch & Buch (Verlag) (2011)
40. J. Shah, *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures*, vol. 11, 1st edn. (Springer Science & Business Media, 2013)
41. M. Sugawara, H. Ebe, N. Hatori, M. Ishida, Y. Arakawa, T. Akiyama, K. Otsubo, Y. Nakata, Theory of optical signal amplification and processing by quantum-dot semiconductor optical amplifiers. Phys. Rev. B **69**(23), 235332 (2004)

42. D. Bimberg (ed.), *Semiconductor Nanostructures*, 2nd edn. (Springer, 2008)
43. L.H. Li, M. Rossetti, A. Fiore, L. Occhi, C. Velez, Wide emission spectrum from superluminescent diodes with chirped quantum dot multilayers. Electron. Lett. **41**(1), 41–43 (2005)
44. T. Akiyama, M. Ekawa, M. Sugawara, K. Kawaguchi, A. Kuramata, H. Ebe, Y. Arakawa, An ultrawide-band semiconductor optical amplifier having an extremely high penalty-free output power of 23 dBm achieved with quantum dots. IEEE Photonics Technol. Lett. **17**(8), 1614–1616 (2005)
45. P. Borri, W. Langbein, J.M. Hvam, F. Heinrichsdorff, M.H. Mao, D. Bimberg, Spectral hole-burning and carrier-heating dynamics in InGaAs quantum-dot amplifiers. IEEE J. Sel. Top. Quantum Electron. **6**(3), 544–551 (2000)
46. T.W. Berg, S. Bischoff, I. Magnusdottir, J. Mørk, Ultrafast gain recovery and modulation limitations in self-assembled quantum-dot devices. IEEE Photonics Technol. Lett. **13**(6), 541–543 (2001)
47. A.V. Uskov, T.W. Berg, J. Mørk, Theory of pulse-train amplification without patterning effects in quantum-dot semiconductor optical amplifiers. IEEE J. Quantum Electron. **40**(3), 306–320 (2004)
48. A.V. Uskov, E.P. O'Reilly, M. Laemmlin, N.N. Ledentsov, D. Bimberg, On gain saturation in quantum dot semiconductor optical amplifiers. Opt. Commun. **248**(1–3), 211–219 (2005)
49. B. Lingnau, Nonlinear and nonequilibrium dynamics of quantum-dot optoelectronic devices, Doctoral Thesis, Technische Universität Berlin (2015)
50. A.V. Uskov, E.P. O'Reilly, R.J. Manning, R.P. Webb, D. Cotter, M. Laemmlin, N.N. Ledentsov, D. Bimberg, On ultrafast optical switching based on quantum-dot semiconductor optical amplifiers in nonlinear interferometers. IEEE Photonics Technol. Lett. **16**(5), 1265–1267 (2004)
51. T. Akiyama, H. Kuwatsuka, N. Hatori, Y. Nakata, H. Ebe, M. Sugawara, Symmetric highly efficient (~0 dB) wavelength conversion based on four-wave mixing in quantum dot optical amplifiers. IEEE Photonics Technol. Lett. **14**(8), 1139–1141 (2002)
52. T. Vallaitis, C. Koos, R. Bonk, W. Freude, M. Laemmlin, C. Meuer, D. Bimberg, J. Leuthold, Slow and fast dynamics of gain and phase in a quantum dot semiconductor optical amplifier. Opt. Express **16**(1), 170–178 (2008)
53. S.-M. Lee, S.-G. Mun, M.-H. Kim, C.-H. Lee, Demonstration of a long-reach DWDM-PON for consolidation of metro and access networks. J. Light. Technol. **25**(1), 271–276 (2007)
54. S.-J. Park, Y.-B. Choi, J.-M. Oh, S.-G. Koo, D. Lee, An evolution scenario of a broadband access network using R-SOA-based WDM-PON technologies. J. Light. Technol. **25**(11), 3479–3487 (2007)
55. Y. Tian, Q. Chang, Y. Su, A WDM passive optical network enabling multicasting with color-free ONUs. Opt. Express **16**(14), 10434–10439 (2008)
56. S.-G. Mun, J.-H. Moon, H.-K. Lee, J.-Y. Kim, C.-H. Lee, A WDM-PON with a 40 Gb/s (32 x 1.25 Gb/s) capacity based on wavelength-locked Fabry-Perot laser diodes. Opt. Express **16** (15), 11361–11368 (2008)
57. M. Omella, I. Papagiannakis, B. Schrenk, D. Klonidis, J.A. Lázaro, A.N. Birbas, J. Kikidis, J. Prat, I. Tomkos, 10 Gb/s full-duplex bidirectional transmission with RSOA-based ONU using detuned optical filtering and decision feedback equalization. Opt. Express **17**(7), 5008–5013 (2009)
58. B. Schrenk, G. De Valicourt, M. Omella, J.A. Lazaro, R. Brenot, J. Prat, Direct 10-Gb/s modulation of a single-section RSOA in PONs with high optical budget. IEEE Photonics Technol. Lett. **22**(6), 392–394 (2010)
59. R.S. Vodhanel, A.F. Elrefaie, M.Z. Iqbal, R.E. Wagner, J. Gimlett, S. Tsuji, Performance of directly modulated DFB lasers in 10-Gb/s ASK, FSK, and DPSK lightwave systems. J. Light. Technol. **8**(9), 1379–1386 (1990)
60. A.S. Karar, Y. Gao, K.P. Zhong, J.H. Ke, J.C. Cartledge, Generation of DPSK signals using a directly modulated passive feedback laser, in *European Conference and Exhibition on Optical Communication (ECOC)* (2012), p. Tu.4.A.1

61. H. Kim, Transmission of 10-Gb/s directly modulated RSOA signals in single-fiber loopback WDM PONs. IEEE Photonics Technol. Lett. **23**(14), 965–967 (2011)
62. A. Zeghuzi, H. Schmeckebier, M. Stubenrauch, C. Meuer, C. Schubert, C.-A. Bunge, D. Bimberg, 25 Gbit/s differential phase-shift-keying signal generation using directly modulated quantum-dot semiconductor optical amplifiers. Appl. Phys. Lett. **106**(21), 213501 (2015)
63. H. Schmeckebier, B. Lingnau, S. Koenig, K. Lüdge, C. Meuer, A. Zeghuzi, D. Arsenijević, M. Stubenrauch, R. Bonk, C. Koos, C. Schubert, T. Pfeiffer, D. Bimberg, Ultra-broadband bidirectional dual-band quantum-dot semiconductor optical amplifier, in *Optical Fiber Communication Conference (OFC)* (2015), p. Tu3I.7
64. T. Vallaitis, R. Bonk, J. Guetlein, D. Hillerkuss, J. Li, R. Brenot, F. Lelarge, G.-H. Duan, W. Freude, J. Leuthold, Quantum dot SOA input power dynamic range improvement for differential-phase encoded signals. Opt. Express **18**(6), 6270–6276 (2010)
65. R. Bonk, G. Huber, T. Vallaitis, S. Koenig, R. Schmogrow, D. Hillerkuss, R. Brenot, F. Lelarge, G.-H. Duan, S. Sygletos, C. Koos, W. Freude, J. Leuthold, Linear semiconductor optical amplifiers for amplification of advanced modulation formats. Opt. Express **20**(9), 9657–9672 (2012)
66. S. Lange, G. Contestabile, Y. Yoshida, K.-I. Kitayama, Phase-transparent amplification of 16 QAM signals in a QD-SOA. IEEE Photonics Technol. Lett. **25**(24), 2486–2489 (2013)
67. S. Schneider, P. Borri, W. Langbein, U. Woggon, R.L. Sellin, D. Ouyang, D. Bimberg, Excited-state gain dynamics in InGaAs quantum-dot amplifiers. IEEE Photonics Technol. Lett. **17**(10), 2014–2016 (2005)
68. I. O'Driscoll, T. Piwonski, J. Houlihan, G. Huyet, R.J. Manning, B. Corbett, Phase dynamics of InAs/GaAs quantum dot semiconductor optical amplifiers. Appl. Phys. Lett. **91**(26), 263506 (2007)
69. K. Uchiyama, S. Kawanishi, M. Saruwatari, 100-Gb/s multiple-channel output all-optical OTDM demultiplexing using multichannel four-wave mixing in a semiconductor optical amplifier. IEEE Photonics Technol. Lett. **10**(6), 890–892 (1998)
70. I. Shake, H. Takara, K. Uchiyama, I. Ogawa, T. Kitoh, T. Kitagawa, M. Okamoto, K. Magari, Y. Suzuki, T. Morioka, 160 Gbit/s full optical time-division demultiplexing using FWM of SOA-array integrated on PLC. Electron. Lett. **38**(1), 37–38 (2002)
71. A. Sharaiha, H.W. Li, F. Marchese, J. Le Bihan, All-optical logic NOR gate using a semiconductor laser amplifier. Electron. Lett. **33**(4), 323–325 (1997)
72. H. Sun, Q. Wang, H. Dong, N.K. Dutta, XOR performance of a quantum dot semiconductor optical amplifier based Mach-Zehnder interferometer. Opt. Express **13**(6), 1892–1899 (2005)
73. S.J.B. Yoo, Wavelength conversion technologies for WDM network applications. J. Light. Technol. **14**(6), 955–966 (1996)
74. H. Zang, J.P. Jue, B. Mukherjee, A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks. Opt. Netw. Mag. **1**, 47–60 (2000)
75. S. Sygletos, I. Tomkos, J. Leuthold, Technological challenges on the road toward transparent networking. J. Opt. Netw. **7**(4), 321–350 (2008)
76. C. Schubert, R. Ludwig, H.-G. Weber, High-speed optical signal processing using semiconductor optical amplifiers, in *Ultrahigh-speed optical transmission technology*, 1st edn., ed. by H.-G. Weber, M. Nakazawa (Springer, Heidelberg, 2007), p. 482
77. T. Yamamoto, M. Nakazawa, Ultrafast OTDM transmission using fiber devices for pulse compression, shaping, and demultiplexing, in *Ultrahigh-speed optical transmission technology*, 1st edn., ed. by H.-G. Weber, M. Nakazawa (Springer, Heidelberg, 2007), p. 482
78. H. Ishikawa (ed.), *Ultrafast All-Optical Signal Processing Devices*, 1st edn. (Wiley, 2008)
79. J. Leuthold, C. Koos, W. Freude, Nonlinear silicon photonics. Nat. Photonics **4**(8), 535–544 (2010)
80. D. Cotter, R.J. Manning, K.J. Blow, A.D. Ellis, A.E. Kelly, D. Nesset, I.D. Phillips, A. J. Poustie, D.C. Rogers, Nonlinear optics for high-speed digital information processing. Science **286**(5444), 1523–1528 (1999)

81. D. Bimberg, C. Meuer, G. Fiol, H. Schmeckebier, D. Arsenijević, Four-wave mixing in 1.3 μm quantum-dot semiconductor optical amplifiers, in *International Conference on Transparent Optical Networks (ICTON)* (2010), p. We.D4.2
82. C. Meuer, H. Schmeckebier, C. Schmidt-Langhorst, G. Fiol, D. Arsenijević, C. Schubert, G. Eisenstein, D. Bimberg, Wavelength conversion of 40-Gb/s NRZ DPSK signals within a 45-nm range using a 1.3 μm quantum-dot semiconductor optical amplifier, in *European Conference and Exhibition on Optical Communication (ECOC)* (2011), p. We.10.P1.26
83. J. Leuthold, C.H. Joyner, B. Mikkelsen, G. Raybon, J.L. Pleumeekers, B.I. Miller, K. Dreyer, C.A. Burrus, 100 Gbit/s all-optical wavelength conversion with integrated SOA delayed-interference configuration. Electron. Lett. **36**(13), 1129–1130 (2000)
84. Y. Liu, E. Tangdiongga, Z. Li, H. de Waardt, A.M.J. Koonen, G.D. Khoe, H.J.S. Dorren, X. Shu, I. Bennion, Error-free 320 Gb/s SOA-based wavelength conversion using optical filtering, in *Optical Fiber Communication Conference (OFC)* (2006), p. PDP28
85. G. Contestabile, A. Maruta, K.-I. Kitayama, Four wave mixing in quantum dot semiconductor optical amplifiers. IEEE J. Quantum Electron. **50**(5), 379–389 (2014)
86. A.E. Kelly, A.D. Ellis, D. Nesset, R. Kashyap, D.G. Moodie, 100 Gbit/s wavelength conversion using FWM in an MQW semiconductor optical amplifier. Electron. Lett. **34**(20), 1955–1956 (1998)
87. G. Contestabile, L. Banchi, M. Presi, E. Ciaramella, Investigation of transparency of FWM in SOA to advanced modulation formats involving intensity, phase, and polarization multiplexing. J. Light. Technol. **27**(19), 4256–4261 (2009)
88. M. Matsuura, O. Raz, F. Gomez-Agis, N. Calabretta, H.J.S. Dorren, 320 Gbit/s wavelength conversion using four-wave mixing in quantum-dot semiconductor optical amplifiers. Opt. Lett. **36**(15), 2910–2912 (2011)
89. H. Schmeckebier, C. Meuer, D. Arsenijević, G. Fiol, C. Schmidt-Langhorst, C. Schubert, G. Eisenstein, D. Bimberg, Wide-range wavelength conversion of 40-Gb/s NRZ-DPSK signals using a 1.3-μm quantum-dot semiconductor optical amplifier. IEEE Photonics Technol. Lett. **24**(13), 1163–1165 (2012)
90. G. Contestabile, Y. Yoshida, A. Maruta, K.-I. Kitayama, Coherent wavelength conversion in a quantum dot SOA. IEEE Photonics Technol. Lett. **25**(9), 791–794 (2013)
91. L. Krzczanowicz, M.J. Connelly, 40 Gb/s NRZ-DQPSK data all-optical wavelength conversion using four wave mixing in a bulk SOA. IEEE Photonics Technol. Lett. **25**(24), 2439–2441 (2013)
92. K. Kikuchi, M. Kakui, C.-E. Zah, T.-P. Lee, Observation of highly nondegenerate four-wave mixing in 1.5 μm traveling-wave semiconductor optical amplifiers and estimation of nonlinear gain coefficient. IEEE J. Quantum Electron. **28**(1), 151–156 (1992)
93. H. Kuwatsuka, T. Simoyama, H. Ishikawa, Enhancement of third-order nonlinear optical susceptibilities in compressively strained quantum wells under the population inversion condition. IEEE J. Quantum Electron. **35**(12), 1817–1825 (1999)
94. S. Diez, Why and how to study four-wave mixing?, in *Photonic Devices for Telecommunications*, 1st edn., ed. by G. Guekos (Springer Science & Business Media, 1999), p. 404
95. K. Obermann, A. Mecozzi, J. Mørk, Theory of four-wave mixing, in *Photonic Devices for Telecommunications*, 1st edn., ed. by G. Guekos (Springer Science & Business Media, 1999), p. 404
96. H. Kuwatsuka, Wavelength conversion devices, in *Ultrafast All-Optical Signal Processing Devices*, 1st edn., ed. by H. Ishikawa (Wiley, 2008), p. 258
97. S. Diez, K. Obermann, S. Scotti, D. Marcenac, F. Girardin, Related topics, in *Photonic Devices for Telecommunications*, 1st edn., ed. by G. Guekos (Springer Science & Business Media, 1999), p. 404
98. G. Grosskopf, R. Ludwig, H.G. Weber, 140 Mbit/s DPSK transmission using an all-optical frequency convertor with a 4000 GHz conversion range. Electron. Lett. **24**(17), 1106–1107 (1988)

99. T. Akiyama, O. Wada, H. Kuwatsuka, T. Simoyama, Y. Nakata, K. Mukai, M. Sugawara, H. Ishikawa, Nonlinear processes responsible for nondegenerate four-wave mixing in quantum-dot optical amplifiers. Appl. Phys. Lett. **77**(12), 1753 (2000)
100. G. Contestabile, Y. Yoshida, A. Maruta, K. Kitayama, 100 nm-bandwidth positive-efficiency wavelength conversion for m-PSK and m-QAM signals in QD-SOA, in *Optical Fiber Communication Conference (OFC) and National Fiber Optic Engineers Conference (NFOEC)* (2013), p. OTh1C.6
101. M. Matsuura, N. Kishi, 40-Gbit/s RZ-DPSK wavelength conversion using four-wave mixing in a quantum dot SOA, in *Optical Fiber Communication Conference (OFC) and National Fiber Optic Engineers Conference (NFOEC)* (2011), p. JThA024
102. M. Matsuura, N. Calabretta, O. Raz, H.J.S. Dorren, Multichannel wavelength conversion of 50-Gbit/s NRZ-DQPSK signals using a quantum-dot semiconductor optical amplifier. Opt. Express **19**(26), B560–B566 (2011)

Chapter 4
Quantum-Dot Mode-Locked Lasers: Sources for Tunable Optical and Electrical Pulse Combs

Dejan Arsenijević and Dieter Bimberg

Abstract In this chapter optical and electrical properties of quantum-dot mode-locked semiconductor lasers as well as applications based on these devices are discussed. Section 4.1 gives a short overview of different pulse generation and mode-locking techniques, with the main focus on passive mode locking, as well as details on the laser design and advanced features of quantum-dot devices. Timing-jitter reduction and frequency-tuning techniques (hybrid mode locking, optical injection and optical self-feedback) are compared in Sect. 4.2. Section 4.3 is devoted to applications of mode-locked lasers in photonic terahertz signal generation and optical data communication systems.

4.1 Quantum-Dot Mode-Locked Lasers

Short optical pulses from semiconductor lasers can be generated in a variety of ways. The simplest approaches are based on switching the gain (gain-switching) or the resonator quality (Q-switching) of a laser, which is initially emitting in the continuous-wave (CW) regime. Pulse widths down to single-digit picoseconds were reported [1, 2]. Generation of much shorter pulses at frequencies far beyond the intrinsic laser bandwidth is possible using mode-locking (ML). Here, the phases of the longitudinal modes of the laser are locked. A pulse train at a repetition rate corresponding to the inverse of the cavity round-trip time is emitted. Three ML techniques exist for semiconductor lasers. Active ML is based on direct modulation of the drive current at a frequency equal to the mode spacing or at their

---

D. Arsenijević (✉) · D. Bimberg
Department of Solid-State Physics, Technische Universität Berlin, Hardenbergstr. 36, 10623 Berlin, Germany
e-mail: dejan@arsenijevic.eu

D. Bimberg
e-mail: bimberg@physik.tu-berlin.de

D. Bimberg
King Abdulaziz University, Jeddah, Saudi Arabia

© Springer International Publishing AG 2017

G. Eisenstein and D. Bimberg (eds.), *Green Photonics and Electronics*,
NanoScience and Technology, DOI 10.1007/978-3-319-67002-7_4

sub-harmonics. For large frequencies, suitable external modulation sources do not exist. Passively ML lasers (PMLLs) incorporate an additional element in the cavity, a saturable absorber, which, as well as the gain section, is direct current (DC) biased. No external radio-frequency (RF) reference source is needed. PML presents the basis of stable and compact devices generating ultra-short pulses of a few hundreds of femtoseconds at several hundreds of GHz. Hybrid ML (HML), combines active and PML. The reverse bias of the saturable absorber is additionally modulated here by a RF signal. Strongly reduced timing jitter of the optical comb and external synchronization to other devices result.

MLLs are strategic devices for a large variety of applications. For THz photonics [3] and radio-over-fiber systems [4], MLL can simplify the source/transmitter architecture. In high bit-rate optical communication networks [5], MLLs are utilized as optical clocks and, in combination with modulators, as transmitters for optical time-division multiplexing (OTDM) systems. Other applications in communications comprise multi-carrier generation [6] and all-optical sampling [7].

PML in a quantum-dot (QD) monolithic two-section MLL has been first reported by Huang et al. [8]. Progress has been made since then focusing on narrow pulse width and high fundamental repetition rates ranging from 5 to 80 GHz [9, 10]. QDs as gain media for MLLs feature several decisive advantages [11] as compared to quantum-well (QW) lasers. The inhomogeneous distribution of QD sizes and shapes in the self-organized Stranski-Krastanov growth mode leads to inhomogeneously broadened gain spectra and thus, if all modes are locked, to very narrow pulses. Transparency current densities as low as 6 A/cm^2 per QD layer were reported [12]. A low threshold current density and internal quantum efficiencies exceeding 90% result in a low amount of dissipated heat. Most of the QD MLLs are operating slightly above the threshold, thus less chirped pulses are present. Temperature insensitive threshold currents were reported for p-doped devices [13] up to 80 °C. Saturable absorbers based on QDs exhibit fast recovery on the order of 700 fs under large reverse bias [14], enabling the generation of sub-picosecond pulses at large frequencies. The timing jitter of MLLs is a result of random fluctuations of the photon density, the gain and the effective refractive index caused by amplified spontaneous emission (ASE). Compared to QWs, QD MLLs exhibit lower levels of ASE [15], which directly reduces the jitter.

4.1.1 Device Structures

The experimental investigations presented in this chapter are carried out using two-section AlGaAs/GaAs QD MLLs with a monolithically integrated saturable absorber. The laser structure is grown by molecular beam epitaxy. In order to ensure sufficient gain, the active region consists of 5–15 InAs QD layers. We focus here on 10 QD layers. The emission wavelength of 1310 nm, the zero-dispersion point of a standard single-mode fiber (SMF-28), is achieved by overgrowth of the QDs with an InGaAs layer, referred to as dot-in-a-well structure. P-doping of GaAs

spacers between QD layers prefills the hole states, speeds up the recovery of the ground-state population and provides temperature insensitivity of the threshold current [16]. Details on the growth and the laser structure can be found in [17, 18]. The wafers are processed into ridge-waveguide edge emitters with stripe widths of typically 1–5 μm, here 4 μm. Strong index guiding of the optical mode and suppression of current spreading is achieved by etching through the active region. Combined with a narrow stripe width, transverse single-mode emission and high coupling efficiencies into a fiber are provided. Polymer interlayers minimize parasitic capacitances such that the absorber section can be effectively modulated by an external RF signal. The fundamental repetition frequency f of the MLL, given by $f = c/2nL$, where c is the speed of light in vacuum and n the effective refractive index, is equal to the spacing of the longitudinal modes, which in turn is defined by the resonator length L. Thus, a cavity of 1 mm yields a repetition frequency of ~40 GHz. The ridge width and particularly the length ratio of the gain to the absorber section (typically ranging from 3:1 to 10:1) are crucial design parameters of monolithic MLLs defining the optical and electrical properties of the pulse train as well as the bias range where complete mode locking occurs [19]. The two sections are formed by etching a ~20 μm gap in the metallization and the upper highly doped contact layer, ensuring electrical isolation of >10 kΩ. The MLL chips, reported here, contain a 900 μm long gain section and a 100 μm saturable absorber. The laser facet at the absorber side is coated for high reflectivity of 99%, the gain-section facet is as cleaved (32%). Data transmission experiments (see Sect. 4.3.2) require stable long-term operation of the transmitter. Therefore, the MLL chips are packaged into modules including a thermo-electric cooler and a lensed SMF-28 coupling [20]. In order to allow the modules to be operated in the HML mode, a narrowband bias-tee is integrated as well as an impedance matching network in the RF path to the absorber. Reflection coefficient levels of −20 dB at the desired frequency is evidence of a strongly improved impedance matching to 50 Ω termination. Impedance matching enables the modules to operate in the HML regime using moderate RF powers. Additionally, the RF extraction efficiency, which is the most critical issue in generating THz signals by means of a MLL, is decisively increased (see Sect. 4.3.1).

4.1.2 Passive Mode-Locking

PML in semiconductor lasers eliminates the need for an external modulation source. The pulse generation is based on the nonlinear absorption characteristics of the saturable absorber, leading from primary random to a defined phase relation between the longitudinal modes. Both, absorption and gain dynamics influence the pulse train. In steady state, unsaturated losses exceed the gain. However, once an optical pulse enters the absorber, its leading edge is absorbed, i.e. the pulse is shortened, and the absorber saturates. Loss saturation occurs faster than gain saturation, which means that the absorber saturation energy is less than that of the gain

medium. Both recover to the initial high net-loss state after a certain time, thereby building a net-gain window which determines the pulse characteristics. Saturable absorbers in semiconductor lasers are termed as *slow* [21]. The trailing edge of the pulse might be only partially attenuated during its transit through the absorber section. Beside other causes like spectral bandwidth limitations, waveguide dispersion, self-phase modulation or spontaneous emission noise, the pulse broadening process is attributed for the main part to gain saturation in the gain section [22]. The importance of saturation energies and absorber recovery times are evident. The ratio of the gain to the absorber section saturation energies, i.e. the key pulse forming parameter, increases for QWs [22] and even further for QDs [23] as compared to bulk active material, due to the reduction in differential gain with increasing carrier density owing to the step-like (QWs) and discrete (QDs) density-of-states function, respectively. A fast recovery of the saturable absorber to the high net-loss level before arrival of the subsequent pulse preventing noise, is mandatory for stable ML operation, especially at large repetition frequencies. The recovery time can be decisively reduced by applying a large reverse bias to the absorber, which enhances the carrier sweep rate out of the active region, as long as the photon energy is still in resonance to the band gap in these large electrical fields. Ultra-fast recovery times of 700 fs under a bias voltage of −10 V are reported for 1.31 μm QD-based absorbers, attributed to direct carrier tunneling out of the QD ground state [14].

The color-coded 2D map in Fig. 4.1 shows a typical optical pulse-width dependence on bias conditions for a 40 GHz QD PMLL with a gain-to-absorber length ratio of 9:1. The full-width at half-maximum (FWHM) of the deconvoluted pulse width, derived from autocorrelation measurements assuming a Gaussian pulse shape, ranges from 2 to 9 ps. Depending on the bias point, different operating regimes are observed. In general, QD MLL exhibit complete ML across much wider bias ranges as compared to QW devices [24]. Here, complete ML extends over gain currents and absorber biases of 25–95 mA and −0.5 to −10 V, respectively. Basically, ML is established at the onset of lasing. For relative short absorbers and increasing current, after zero intensity, pure Q-switching and

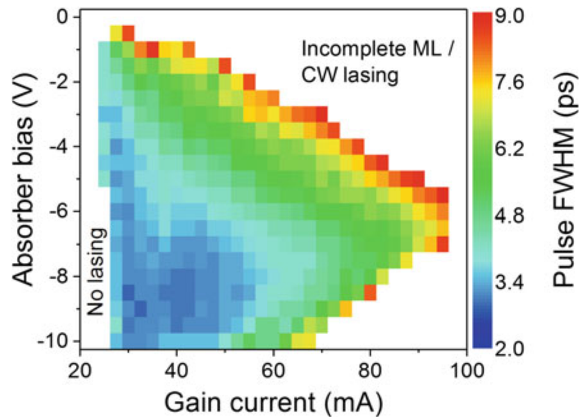


Fig. 4.1 Pulse FWHM dependence on gain current and absorber voltage at room temperature. The deconvoluted pulse width is derived from autocorrelation measurements assuming a Gaussian pulse shape

Q-switched ML can occur prior to fundamental ML, incomplete ML (pulses with CW background) and finally CW lasing at large currents [25]. The Q-switching mode is restricted to a very narrow range of 1–2 mA. For longer absorbers, Q-switching disappears completely. Moreover, wider ML regions, narrow sub-ps pulses and larger optical powers are feasible, owing to increased saturable absorption, reduced pulse-broadening in the gain section and larger gain saturation energy, respectively [19, 23]. For a constant current value and increasing reverse bias up to −8.5 V, the pulse width decreases exponentially, in agreement with the exponential dependence of the absorber recovery time on the voltage [26]. At absorber voltages exceeding −8.5 V, the pulse width increases again due to a redshift of the absorption spectrum caused by the quantum-confined Stark effect [27].

The product of the temporal and spectral FWHM is termed the time-bandwidth product. A given spectral bandwidth enables the generation of pulses exhibiting a certain minimum temporal width. The more modes are locked, the narrower the resulting pulse is. However, the number of phase-locked longitudinal modes depends on the MLL design and the carrier injection level. Fourier-limited optical pulses from two-section QD MLLs have been reported at different repetition frequencies: 10 ps at 18 GHz [28], 3 ps at 50 GHz and 0.78 ps [29] at 24 GHz [26]. A detailed analysis of the true pulse shape, its phase and chirp, basing upon frequency-resolved optical-gating measurements, can be found in [30].

4.2 Jitter Reduction and Frequency Tuning

The properties of pulse sources are not sufficiently described by the pulse duration and the pulse shape, but depend equally on their timing jitter and amplitude noise. Pulse trains generated by PMLLs exhibit relatively large timing jitter, as well as a large optical linewidth and thus do not meet requirements for high-speed systems or optical sampling, making them initially ineffective for such applications. The introduction of phase-noise reduction techniques is required.

The precision of the laser cavity definition is defined by the accuracy of the cleavage technique. A standard method, diamond scribing followed by mechanical break, depends on the experience of the operator and results in typical length variations of minimum ±5 µm. For a monolithic 40 GHz MLL chip, the corresponding frequency deviation is then ±200 MHz. Universal techniques allowing the tuning of the device repetition rate thus become essential to match the MLL frequency to the desired one.

Different approaches exist to control the frequency and improve at the same time the timing jitter and the optical linewidth of MLLs. In this section, HML, optical injection and optical self-feedback (OFB) are discussed.

Conventionally, the phase-noise characteristics of pulse sources are specified by the timing jitter as the figure-of-merit. The optical output of a MLL, which is detected using a large-bandwidth photodetector and an electrical spectrum analyzer,

finally results in a frequency-dependent electrical current. RF spectra and single-sideband (SSB) phase-noise traces can be measured. The integrated root-mean-square (rms) timing jitter is calculated by integrating the SSB phase-noise power spectral density, according to von der Linde's method [31]:

$$\sigma_i^{rms}(f_l, f_u) = \frac{1}{2\pi n f_c} \sqrt{2 \int_{f_l}^{f_u} L(f_\Delta) df_\Delta}, \tag{4.1}$$

where $L(f_\Delta)$ denotes the SSB phase-noise spectral density at the frequency offset f_Δ with respect to the carrier frequency f_c, n is the number of harmonics at which the phase noise is measured, while f_l and f_u are the lower and upper frequency boundaries across which $L(f_\Delta)$ is integrated. The integrated rms timing jitter is often quoted as a single number, without giving the integration interval. Different integration widths and boundaries are used in reports, ranging from 100 Hz to 5 GHz, which makes a comparison of different device reports difficult. According to the ITU-T recommendation for telecom applications, the specified integration range for 40 GHz pulse sources is from 16 to 320 MHz [32]. In this chapter, the integrated timing jitter is calculated using integration borders of 10 kHz to 1 GHz. Depending on the particular device used in each subsection and its operating conditions, the timing jitter values of the PMLLs differ, but are always in the range of 3.8–9 ps.

Von der Linde's method is well suited for active or hybrid ML, where an external reference source is used and the carrier is shaped as a narrow line on a pedestal. However, the noise characteristics of PMLL differ fundamentally [33], as PML is dominated by broadband spontaneous emission and operates without precise timing of the net-gain window. Consequently, broad RF linewidths and non-stationary timing-jitter fluctuations [34] are present. For large fluctuations and integrated-jitter values exceeding single-digit picoseconds, the measurement system is not able to track the carrier frequency and amplitude, and thus the measurements fail. Due to the absence of a restoring force in PMLLs, the timing of each pulse is correlated to that of the previous one. The issue of timing-jitter fluctuations is expressed by diffusion theories for Gaussian processes, leading to a model, where the pulse-to-pulse rms timing jitter can be determined from the SSB phase-noise spectral density [35]. In another approach [36], the RF linewidth of the first harmonic is linked to the pulse-to-pulse rms timing jitter, which is calculated using:

$$\sigma_{p-p}^{rms} = \mathrm{T}\sqrt{\frac{\Delta\upsilon_{RF} N T}{2\pi}}, \tag{4.2}$$

where $\Delta\upsilon_{RF}$, N and T denote the FWHM of the Lorentzian fit of the carrier line in the electrical spectrum, the number of periods between the two compared pulses and the pulse repetition period, respectively. As only the RF linewidth is a figure-of-merit, the pulse-to-pulse rms timing jitter of PMLL is obtained preferentially using this method.

An optical time-domain method to directly determine timing fluctuations and jitter is based on cross correlations generated by means of a second-harmonic generation in a nonlinear crystal [37]. The increase of the cross-correlation width as compared to that of the autocorrelation is caused by the timing jitter. The advantage of this method is its frequency scalability, whereas electrical spectrum analyzers and telecommunication-bands photodetectors are limited to 67 GHz (higher frequencies are only accessible using mixers) and 100 GHz, respectively. However, in case of wide pulses exhibiting much smaller timing jitter, the measurement error of the optical method strongly increases.

4.2.1 Hybrid Mode-Locking

A common technique to reduce the timing jitter and establish a frequency locking range of a two-section MLL is HML. Since the first demonstration using a 10 GHz QD MLL [38], a steady device improvement has led to decisive jitter reductions, e.g. 190 fs for a 40 GHz laser [39]. However, the HML capabilities of monolithic QD-based lasers, especially in terms of the locking range, still lag behind the one of QW devices [23].

HML is based on electrical modulation of the absorber reverse bias and enables reliable synchronization with other devices in a system and a precise matching of desirable frequencies. The DC bias of the absorber section together with a superimposed low-noise sinusoidal RF signal defines precisely the timing of the net gain window and suppresses background noise. Sidebands of each longitudinal mode are generated with spacing equal to the MLL frequency resulting in strong reduction of jitter. The external modulation of the reverse absorber voltage, which is provided by a low-noise and large RF-power synthesizer or voltage-controlled oscillator, is described by $\tilde{V}\,[1 + A\cos(\Omega t)]$, where $\tilde{V}$, A and Ω are the DC component, the amplitude and frequency of the absorber bias modulation, respectively. HML allows tuning of the device repetition rate by changing Ω.

For various absorber and gain section bias parameters, the external frequency is varied around the cavity repetition rate of the MLL in order to ascertain the locking range. The criterion to consider operation as locked to the external signal is a sideband-suppression ratio larger than 30 dB, which is determined from electrical spectra. Figure 4.2a and b illustrate representative electrical spectra for a synthesizer frequency of 40.93 and 40.96 GHz, namely a locked and an unlocked solution, respectively. If the MLL is synchronized to the external signal, a single and narrow carrier line is present. Otherwise, multiple beat products of the external and the PML frequency are observed. The transition from locked to unlocked regime is rather sharp.

The frequency of the external RF signal is usually chosen to be close to the PML frequency. The timing jitter and the locking range strongly depend on the bias point of the MLL and the modulation voltage of the external signal. Within the locking range at any bias point, the MLL follows the external RF signal, but exhibits

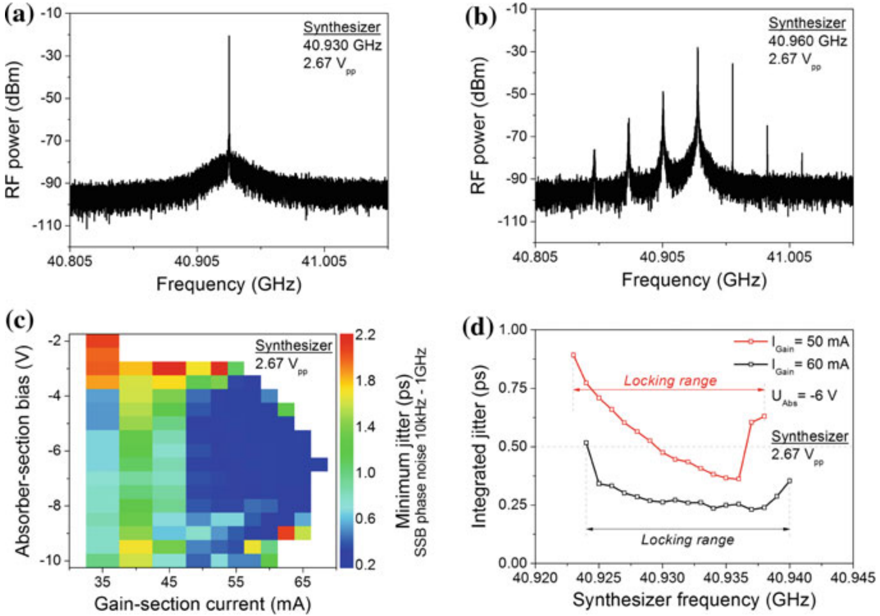


Fig. 4.2 Electrical spectra at the 60 mA/−6 V bias point for an external RF of **a** 40.93 GHz (locked case) and **b** 40.96 GHz (unlocked case). **c** Two-dimensional map of the minimum integrated jitter as a function of gain-section current and absorber-section bias. **d** Integrated jitter values as a function of the synthesizer RF within the respective locking range for two different operating points (*red* 50 mA, −6 V and *black* 60 mA, −6 V)

different jitter values. Lowest integrated jitter values ranging from 190 fs to 2.2 ps for various pairs of gain current and absorber bias are depicted as color-coded map in Fig. 4.2c. The modulation voltage of the external RF signal is 2.67 $V_{p\text{-}p}$ (peak-to-peak voltage). Jitter values below 500 fs are observed in a wide range above 50 mA and thus in regions where broader pulses are generated. HML enables a decisive improvement of the timing jitter by reducing the electrical RF linewidth, while the optical properties of the pulse train are unaffected, i.e. the width and shape of the pulses as well as the optical spectra remain the same.

A maximum locking range of 32 MHz for a single operating point is measured at low bias values. The absorber section of a MLL chip is an unmatched load and the external modulation voltage is partially reflected from it. A more sufficient seeding of the external RF signal is achieved by impedance matching, as described for the MLL module in the first section, and the maximum locking range extends to 53 MHz even at a lower modulation voltage of 2.0 $V_{p\text{-}p}$ (10 dBm). As expected, an optimized impedance-matched electrical circuit translates to an efficient RF power transfer to the laser chip and thus to improved HML characteristics. At the same time, a lower RF power consumption is achieved. Only such devices have the potential to reduce noticeably the operating cost and replace expensive RF synthesizers with RF amplifiers by compact voltage-controlled oscillators. However,

under HML, QW-based devices have still outperformed QD MLLs in terms of timing jitter [40] and locking range [41]. This disadvantage is mainly attributed to the smaller confinement factor and quantum-confined Stark shift of QD absorbers, reducing the modulation efficiency [42, 43].

Figure 4.2d shows the dependence of the integrated jitter on the synthesizer frequency measured at two different operating points, 50 mA/−6 V and 60 mA/−6 V, for a constant RF voltage of 2.67 $V_{p\text{-}p}$. The jitter values range from 361 to 892 fs and from 231 to 517 fs, respectively. For both operating points similar and sharply bounded locking ranges of 15–16 MHz are observed. However, a non-negligible jitter deviation for the first operating point (50 mA/−6 V) is found, whereas for the second operating point the jitter remains relatively constant and increases significantly only at the locking-range boarders. The jitter gradient within the locking range obviously strongly depends on the bias point of the MLL. Generally, the locking range decreases with increased absorber voltage and increases again for large reverse biases (>8 V). The pulse width follows the same trend: shorter pulses with larger peak power exhibit smaller locking ranges [44] as well as larger timing jitter (Fig. 4.2c). This phenomenon is in agreement with simulations of the locking-range dependence on the ratio of the saturation energies in the absorber and gain sections [45], which is a key mode-locking parameter [46]. An increase of this ratio leads usually to more stable mode locking and improved pulse quality [47]. A similar decrease of the locking range is observed with the increase of the bandwidth of the spectral filtering, which also yields narrower pulses [45].

Both values, the locking range and timing jitter depend on the modulation voltage of the external RF signal. The locking range increases linearly with the modulation amplitude for all operating parameters. The boundaries of the locking range for modulation voltages from 0 V (PML case) to 2.89 $V_{p\text{-}p}$ image a locking cone (black curves in Fig. 4.3a) with a maximum width of 17.6 MHz, which is almost symmetric to the PML frequency (grey line) at low voltages, but exhibit an asymmetry to lower frequencies for increasing voltage. The operating point was

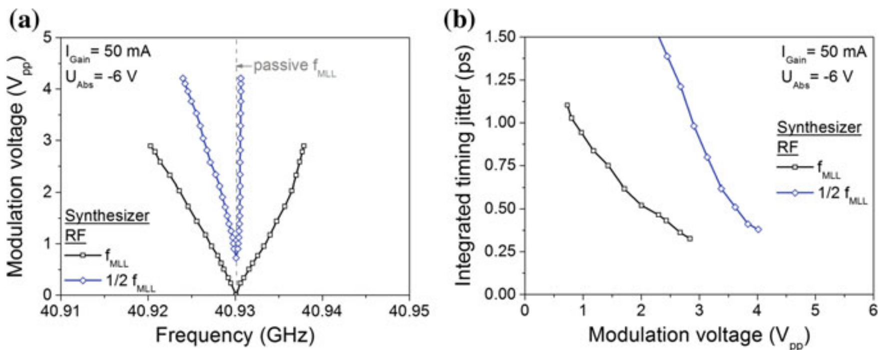


Fig. 4.3 **a** Modulation voltage dependence of the locking range cones and **b** minimum jitter values of the ~40 GHz HMLL under synthesizer RF around the native (*black*) and half of the MLL frequency (*blue*). The operating point is 50 mA/−6 V

50 mA/−6 V. Depending on the bias point of the MLL and the properties of its surroundings, the strength of the asymmetry varies and is related to the nonlinear pulse frequency dependence on the pulse amplitude, a nonlinear resonance phenomenon [48].

For hybrid MLL at much larger frequencies than ~40 GHz it is challenging to provide an adequate synthesizer signal. However, hybrid mode locking can also be initiated by applying an external RF at approximately half of the MLL frequency (here around 20.465 GHz). This approach has led to a strongly red-shifted asymmetry of the locking cone [49], typical of a sub-harmonic resonance [50], with a maximum locking range of 6.6 MHz at 4.21 $V_{p\text{-}p}$ (blue curves in Fig. 4.3a). Modulation voltage dependence of timing jitter for native and sub-harmonic modulation RF signals is illustrated in Fig. 4.3b showing minimum values of 326 fs and 381 fs for native and sub-harmonic modulation RF signals, respectively. The implementation of sub-harmonic RF signals enables HML at twice the synthesizer frequency, still exhibiting sufficient jitter reduction at the expense of a larger modulation voltage and a narrower locking range [51].

4.2.2 *Optical Injection*

Light injection of an external laser source, exhibiting narrow optical linewidth, into a MLL is known to be an effective optical method for jitter reduction. Besides the jitter improvement, optical injection is utilized to decisively narrow the optical linewidth of the MLL. The optical linewidth of a single longitudinal MLL mode is in the order of 10 MHz and thus too larger for emitters used in coherent systems [52]. Single-tone CW injection into an actively mode-locked extended-ring-cavity laser has led to a linewidth narrowing down to 20 kHz [53]. Phase locking of all longitudinal modes to the external laser in QW devices has been demonstrated [54]. For QD-PMLLs oscillating in the ground state, phase- and amplitude-noise reduction is achieved by pulse injection at its excited-state wavelength [55] as well as by CW ground-state injection showing no instabilities at zero detuning frequency [56]. In [57], single-tone injection at a cavity mode is shown to shift the emission spectrum of the MLL. The optical spectrum, detuned by up to 15 nm, does not overlap with the native one, thus enabling the MLL to operate on at least two different wavelength channels. All these observations suggest that the noise and tuning characteristics of HMLLs could be further improved by optical injection seeding. However, the jitter and optical linewidth reduction and the wide frequency tuning capability are usually accompanied by residual distortion of the pulse emission due to fundamental changes of the emitted optical spectrum under injection. Moreover, the stability and particularly the reproducibility of the injection seeding techniques have to be judged carefully.

In case of CW single-tone injection, the light from a tunable laser source (master laser), featuring a narrow optical linewidth of below 200 kHz and allow fine tuning of the wavelength and power, is injected through the as-cleaved gain-section facet

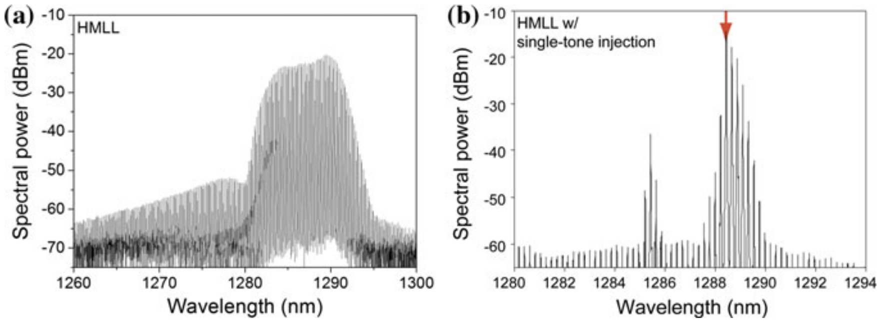


Fig. 4.4 Optical spectrum of **a** the HMLL and **b** injection-locked HMLL. The seed wavelength is indicated by the *arrow*. The injection power and wavelength are 770 μW and 1288.44 nm, respectively

of the MLL (slave laser) via a fiber circulator. A polarization controller is necessary to match the polarization of the injection seed and the slave. The MLL, operating in the hybrid regime using a 14 dBm modulation swing, exhibits, for the chosen bias conditions, a 620 fs timing jitter, a frequency tuning range of 10 MHz, a pulse duration of 3.1 ps and a 10 nm optical spectrum width at −10 dB level (Fig. 4.4a). When the slave laser is locked to the master, its optical spectrum narrows with the majority of modes red shifted from the injection wavelength, here 1288.44 nm (Fig. 4.4b). Optical spectrum narrowing causes pulse broadening, here an increase from 3.1 to 4.8 ps. Although still well distinguished pulses are observed, the pulse train is tainted with a significant CW background (40% of the pulse amplitude) due to the large injection power, which becomes a critical parameter.

For a 770 μW injection power and a wavelength of 1288.44 nm, the optical linewidth of the injection-locked HMLL is similar to the one for the master laser (~200 kHz), demonstrating complete optical linewidth reduction for the MLL. The integrated jitter is reduced by factor of ~3, down to 240 fs. By varying the wavelength of the external-cavity laser on the sub-nm scale, the frequency of the MLL can be changed. The frequency tuning range, benefiting from the narrower spectral width of the injection-locked MLL, is strongly increased compared to the free-running HMLL and reaches a maximum of 167 MHz. When the absorber section of the MLL is modulated using a lower RF swing (9 dBm), while the optical injection settings are kept constant, the locking range decreases down to 40 MHz and the integrated jitter increases to 365 fs. This result confirms the interaction of the optical injection seeding and the electrical modulation and their correlated influence on the laser emission. Although improved properties of the HMLL under optical single-tone injection and, once the slave laser is locked, long-term stability are provided, the system is accompanied by locking hysteresis and is difficult to control.

Injection locking via two coherent lines has been demonstrated for both, longitudinal single-mode lasers [58] and MLLs [59], yielding improved control of the slave laser parameters. For dual-tone injection, the light from the tunable laser

source is additionally modulated by means of a Mach-Zehnder modulator (MZM). The modulator, which is biased at its transmission minimum, is driven using a sinusoidal signal with a 2 × V_π swing at half the repetition frequency of the MLL. Consequently, two coherent sidebands with a distance corresponding to the MLL repetition frequency are generated, while the optical carrier is suppressed. In contrast to signal-tone injection, the repetition rate tuning of the MLL is controlled by varying the spacing between the sidebands, i.e. by changing the modulation frequency applied to the modulator.

Similar to the CW injection seeding, dual-tone injection leads to optical spectrum narrowing and a red shift as well as pulse width broadening. Optical linewidth measurements reveal a ~135 kHz wide beat product, thus showing an all-modal optical linewidth reduction. The injection power has a major influence on the slave characteristics. Two representative points for strong (570 μW) and weak (240 μW) dual-tone injection are presented in Fig. 4.5. In case of weak injection, a broad optical spectrum and only a slight CW background in the autocorrelation trace are observed. The pulse width is increased to moderate 3.6 ps. On the other hand, injection at larger optical power clearly narrows the spectrum and thus yields a low jitter of 121 fs and an ultra-wide frequency tuning range of 342 MHz, but at the expense of 5.2 ps broad pulses and increased CW background.

Figure 4.6 shows the locking range (green dots) and timing jitter (black dots) of the dual-tone injection-locked PMLL as a function of the injection power. At large powers, the sideband injection seeding dominates and the MLL parameters do not rely on HML, which is on the contrary to single-tone injection. However, at lower powers (below 400 μW) the dual-tone injection is insufficient to tune the frequency neither to essentially improve the jitter.

4.2.3 Optical Self-Feedback

Both techniques, hybrid HML and optical injection, present suitable solutions for effective jitter reduction and frequency tuning of the emitted pulse train. Indeed, very low timing jitter values and wide locking ranges were reported, but for HML a low-phase-noise high-frequency electrical oscillator is required, accompanied by large cost. The optical injection approach suffers from the same disadvantage caused by the additional external light source. Optical self-feedback (OFB), where a part of the master laser light is injected back into the device, has been demonstrated several decades ago for distributed-feedback lasers [61] and has been revisited as an effective way to reduce the phase noise of pulse trains [57, 62]. A reduction of the RF linewidth in the electrical spectrum from 46 to 1.1 kHz is reported for a 5.25 GHz QD MLL [63]. For a 10 GHz single-section InP-based quantum-dash laser, a reduction of the 3-dB linewidth from 30 kHz to 200 Hz was shown [64]. However, OFB is known to degrade the properties of MLL pulses [65]. In this section the impact of three crucial OFB parameters on two-section PMLLs is discussed. Five fundamentally different regimes, including a decisive one leading to

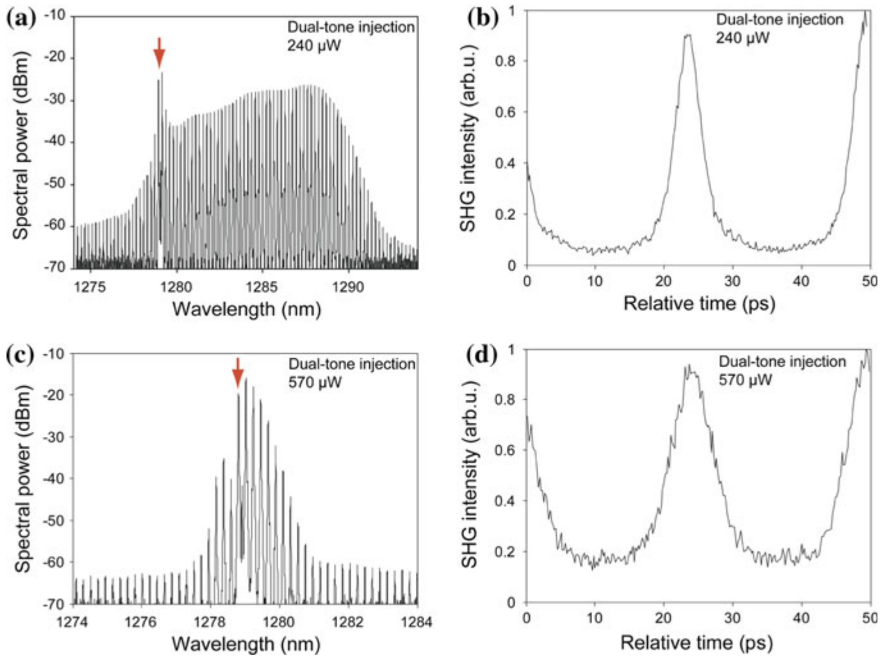


Fig. 4.5 **a** Optical spectrum and **b** autocorrelation trace of the MLL with dual-tone sideband injection locking using 240 µm optical power. **c** MLL optical spectrum and corresponding **d** autocorrelation trace for a 570 µW injection. The injection wavelengths are represented by the *arrows* [60]

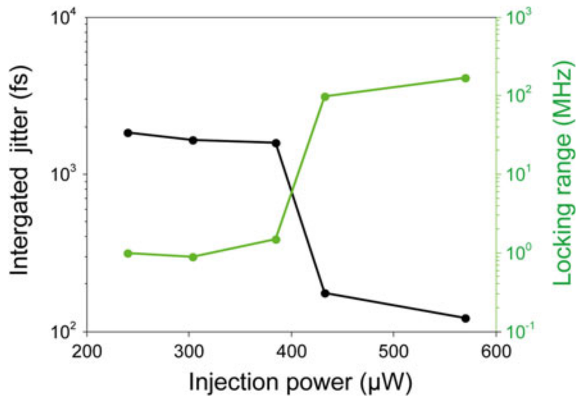


Fig. 4.6 Locking range (*green dots*) and integrated jitter (*black dots*) of a MLL as a function of the dual-tone injection power

enormous improvements of the pulse properties, are reported. These results are not restricted to the particular wavelength or frequency of the laser used here, but are of a general nature.

A MLL, operating in the passive regime and at the bias point where the most stable mode locking occurs (integrated timing jitter of 3.8 ps), exhibits fundamental mode locking at a frequency of 38.616 GHz and an average optical power in fiber of 2.8 mW. The OFB is realized by injecting a part of the laser light back to the cavity through the facet at the gain section. The experimental SMF-28-based OFB setup comprises an optical circulator and a 90:10 splitter, whereas following components build the feedback loop: a variable attenuator, an optical delay line and a polarization controller as shown in [66]. This fiber-based OFB is fully characterized by three essential parameters. The length of the optical fiber defines the round-trip time and the number of pulses inside the feedback loop. A fine loop-length adjustment by means of the optical delay line, which is in the order of the inverse MLL repetition rate, describes the relative position of the reinjected pulses with respect to the emitted ones. The value of the feedback strength, which is controlled by the variable attenuator, defines the overall attenuation inside the feedback loop relative to the laser output power.

By varying these OFB parameters, changes in frequency and peak power in the RF spectra, the integrated timing jitter and in the autocorrelation of the pulses are observed, defining different feedback regimes. Figure 4.7a depicts the peak power of RF spectra as a function of the delay and feedback attenuation for a constant fiber length of 16.6 m. The effects of OFB are periodic with ~26 ps pulse distance of the MLL. Five different OFB regimes are determined, being similar to the regimes reported for distributed feedback lasers [61]. Figure 4.7b shows representative SSB-phase-noise measurements for each regime, corresponding RF spectra can be found in [67]. Regime 1, located at feedback attenuations smaller than 20 dB, i.e. strong feedback conditions, is characterized by low peak power of the carrier line in the RF spectrum, whereas its width remains constant. These properties translate to a phase-noise-density distribution up to carrier-offset frequencies of 5 MHz, being similar to the one without OFB. However, the carrier line is surrounded by strong sidebands in the RF spectrum across a frequency range of 1.5 GHz, which appear as peaks in the phase-noise trace at frequencies of multiples of the inverse OFB round-trip time (12.4 MHz). The autocorrelation measurements of the pulses exhibit a strong CW background, thus confirming the distortion of mode locking in regime 1 caused by OFB. In regime 2, the MLL exhibits chaotic behavior with random changes of the RF spectrum. In regime 3, the sidebands in the RF spectrum are weaker than in regime 1 but the width of the carrier line is enhanced, yielding a larger phase noise at low carrier-offset frequencies. The CW-background in the autocorrelation traces is not as large as in regime 1, but larger than without OFB. In regime 3, the mode locking is deteriorated by the OFB as well. The presence of sidebands stands for a strong amplitude jitter, provoked by photon density changes as a consequence of the OFB. Moreover, simulations have revealed, that the amplitude modulation is accompanied by an additional phase modulation, being particularly responsible for the chaotic behavior in regime 2. The resonant OFB observed in regime 4 shows a carrier-peak-power increase of 10 dB. The corner frequency of the white-noise plateau in the phase-noise trace is shifted from 80 to 10 kHz and the thermal noise level is reached at a frequency offset of 1 MHz. The

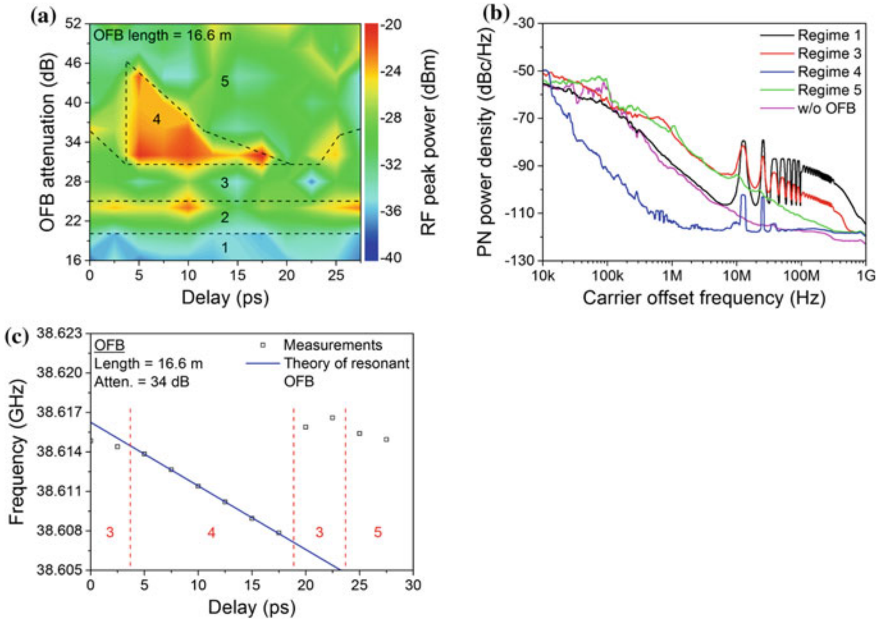


Fig. 4.7 **a** Color coded map of the peak power as a function of delay and feedback attenuation. **b** The SSB-phase-noise traces in different OFB regimes. **c** The repetition rate of the MLL as a function of the delay at a feedback attenuation of 34 dB. For all figures the feedback length is 16.6 m

sidebands are 30 dB weaker than in regime 1 and 3. Thus, the SSB phase noise is strongly suppressed compared to the free running case. At the same time, the pulses are not effected by OFB. In regime 5 only slight influences of the OFB can be noted. The pulses are unaffected, but the carrier linewidth in the RF spectrum fluctuates between the one without OFB and larger widths causing instabilities of the pulse emission.

The timing jitter of the MLL is only improved within the resonant regime. In this case, an integer number of pulses propagate through the feedback loop [62]. The loop round-trip time then equals an integer multiple of the MLL pulse distance. If the feedback length is changed, the MLL frequency shifts in order to meet this criterion (Fig. 4.7c). Within the resonant regime, the frequency is pulled by the change in the delay and exhibits a definite delay dependence. If the OFB is operating in regimes 3 and 5, i.e. being out of resonance, such dependence is not present, confirming a fundamental difference to regime 4. The frequency pulling effect explains also the triangular shape of the resonant regime in Fig. 4.7a. For feedback attenuations larger than 45 dB the influence of the OFB is too weak, whereas stronger OFB leads to the onset of the pulling at a relative delay of 3.5 ps.

At this point, the MLL frequency differs only slightly from the one in the free-running case (38.616 GHz). For larger delays the MLL frequency decreases. Under stronger feedback the frequency can be pulled further and the frequency range, which amounts to 8 MHz at 34 dB attenuation, is extended reaching its maximum at a feedback attenuation of 30 dB just before crossing to regime 3.

Based on the frequency pulling phenomenon of a MLL induced by resonant feedback from an external reflecting target, a simple displacement and vibration sensing technique is demonstrated [68]. The measurement of the laser frequency allows accurate determination of the relative position of the reflecting object. The sensor theoretically exhibits sub-wavelength resolution, but is limited by the RF linewidth of the laser and the resolution of the frequency measurements.

For all practical applications the MLL must operate in the resonant regime. In order to maximize this region, the influence of the third parameter, i.e. overall fiber loop length, is investigated. At 31.7 m, corresponding to ~6000 pulses in the loop cavity, the lowest timing jitter and the maximum extent of the resonant regime are determined (Fig. 4.8a and b). The regime 3 vanishes and the resonant one extents to a feedback attenuation of 27 dB and boarders directly with regime 2. At attenuations larger than 31 dB, the delay dependence of regime 4 is similar to the one known from the shorter loop length. At the border to regime 5 the jitter can exceed 7 ps (b/w pattern region in Fig. 4.8b). Between 27 and 31 dB the resonant feedback regime stretches across the complete delay, which is favorable for practical applications since fluctuations of the loop length do not cause switching into other regimes. Note that the delay values are always relative and do not imply the same position for the two different fiber lengths. The integrated jitter is reduced from 3.8 ps for the free-running case to values lower than 1 ps.

The optimum OFB parameters are found to be a feedback length incorporating ~6000 pulses, a 29 dB feedback attenuation and a delay of 19 ps. At this operating point, the carrier peak power in the RF spectrum is −21.3 dBm, the carrier signal-to-noise ratio (SNR) is 66.3 dB and the sidebands, located at multiples of 6.5 MHz from the carrier, are suppressed by at least 45.6 dB (Fig. 4.8c). Compared to the free-running case, the 3 dB linewidth and the integrated timing jitter are decisively suppressed from 187 kHz and 3.8 ps down to 1.9 kHz and 219 fs, respectively, corresponding to a 94% and 99% reduction of the values, respectively. For comparison, operating in the hybrid mode-locking locking regime under an external 2.7 $V_{p\text{-}p}$ RF modulation, this particular MLL exhibits an integrated timing jitter of 250 fs. Due to the fact, that the operating point is surrounded by a large resonant regime, disruptive external influences do not cause switching to other regimes. OFB is therefore indeed an alternative technique for effective phase-noise reduction of PMLLs and features the huge advantages of low cost and no additional power consumption.

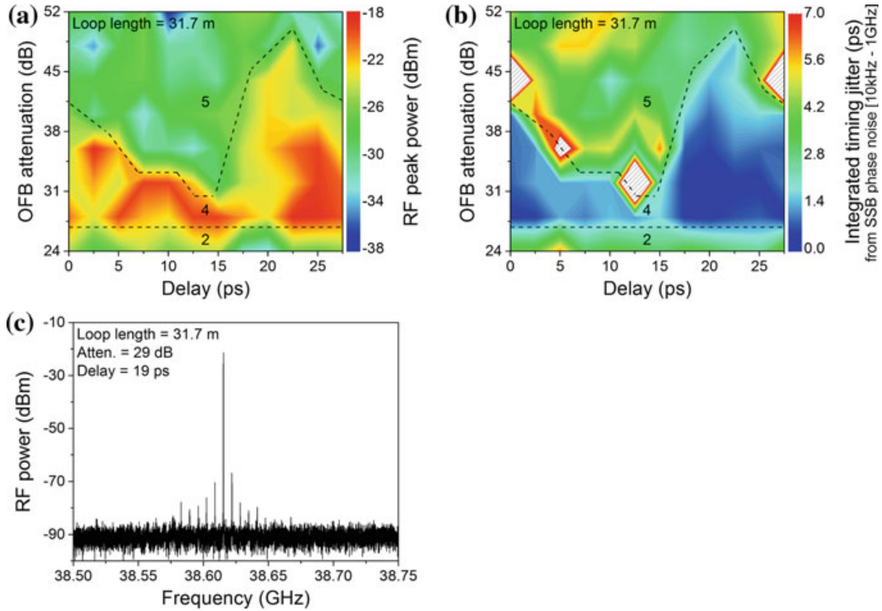


Fig. 4.8 A color-coded map of the RF peak power (**a**) and the integrated jitter (**b**) as a function of delay and feedback attenuation at a feedback length of 31.7 m. (**c**) The RF-spectrum of the MLL subject to optimum OFB

4.3 Applications

4.3.1 Millimeter-Wave-Signal Generation

Photonic signal sources in the range of microwave up to terahertz frequencies are of great interest for a variety of applications, e.g. medical and security imaging, ultra-broadband wireless and satellite communications as well as radar systems [69]. Compared to electronic circuit approaches, the photonic ones benefit from lower cost and lower power dissipation as well as larger frequency tunability. The O/E conversion is a key factor in terahertz-signal generation and requires large-bandwidth and high saturation-current photodetectors. Waveguide-coupled pin photodiodes used in high-speed optical communication exhibit −3 dB bandwidths exceeding 100 GHz. Travelling-wave metal–semiconductor–metal detectors present an alternative interesting approach [70]. A specially designed detector type for terahertz photonics is the uni-travelling carrier photodiode based on separate absorber and depletion layers [71], for which a bandwidth of 310 GHz has already been demonstrated [72]. Moreover, the generated electrical power is enhanced by 10–20 dB as compared to conventional pin photodiodes [73]. Integrated with various types of antennas, a compact terahertz sources can be realized.

Various optical techniques based on semiconductor lasers for generation of electrical microwave to terahertz signals have been explored, e.g. optical heterodyne generation, dual-tone lasing, optical injection or mode locking [69]. In case of optical heterodyning, the light of two free-running single-mode lasers emitting at different wavelengths is combined and converted to an electrical signal by a detector. The photocurrent contains a beat signal at the frequency defined by to the wavelength difference. This technique enables wide frequency tunability of the electrical signal, in principle from sub-MHz to several THz, given by the wavelength tunability of the laser sources and the bandwidth of the photodetector. Dual-mode lasers, emitting simultaneously on two different longitudinal resonator modes, provide directly a signal at a frequency equal to the wavelength difference and simplify the approach. 485 GHz electrical signal generation using a two-color Fabry-Pérot laser has been reported [74]. Compared to two free-running lasers, the tunability is however limited and the steps are coarse. In both approaches, the two tones are not precisely phase-locked and consequently the phase of the electrical signal varies in time.

Semiconductor lasers under optical injection exhibit a number of interesting nonlinear dynamic states. For moderate injection strengths and positive detuning from the Hopf-bifurcation boundary, period-one oscillations of the slave laser cause periodic intensity oscillation of the CW optical signal [75]. The modified and undamped relaxation oscillation of the slave laser, that is injection-locked to the frequency of the master laser, produces sidebands at frequency distances exceeding the intrinsic relaxation resonance frequency. The laser enables the generation of continuously tunable RF signals across a wide range [76]. However, the phase noise of RF signals generated using the single-tone injection approach is relatively large. By combining optical sideband injection locking and an optical phase-locked loop, a strongly suppressed SSB phase-noise power spectral density of −92 dBc/Hz at 10 kHz offset is achieved for a 36 GHz millimeter-wave signal, while the locking bandwidth amounts to 30 GHz [77].

MLLs represent another method for photonic micro- and millimeter-wave generation. As the locked longitudinal modes are equally spaced, a RF signal corresponding to the frequency distance is generated upon O/E conversion. Additionally, beat signals at higher harmonics are generated as well. The MLL-based approach benefits from low-noise performance, compactness and low power consumption. However, the main drawback is the low frequency tunability. By optical filtering of two longitudinal modes of an actively MLL, a millimeter-wave signal has been generated [78]. Selecting a higher harmonic of a low-frequency MLL, where modes located between the selected ones are suppressed, yields a high-frequency RF signal. At 37.1 GHz, a phase-noise power spectral density of −75 dBc/Hz at 5 kHz offset is found [78]. However, the requirement of a high-power electrical reference source counteracts the low-power consumption advantage of a single laser source. No electrical sinusoidal signal is necessary for PML. By means of a high-speed

photodetector, the optical pulse train of a PMLL is converted to an electrical signal at the same frequency, which is the conventional way to generate RF signals. A more convenient method is based on the direct extraction of the electrical signal from the saturable absorber of a monolithic PMLL, thus decisively simplifying the photonic RF-signal source. The absorber serves as an intra-waveguide photodetector enabling electrical signals to be generated with high conversion efficiencies, simultaneously along with the optical pulse train. Electrical pulses were directly extracted from the absorber of a three-section passively mode-locked external-cavity laser at a repetition rate of 2.5 GHz [79]. Compared to the external detection of the optical pulse train using a photodiode, the extracted pulses exhibit much larger rise and even longer fall times. The latter is attributed to the limited QW absorber dynamics. In order to overcome these limitations, QD MLLs are utilized, as QD-based absorbers feature sub-picosecond recovery times [14]. For a 10 GHz QD MLL with a gain-to-absorber section ratio of 5:1, a large differential DC-to-photocurrent conversion efficiency of 33% as well as a 86% extraction efficiency were demonstrated [80]. However, the extracted signals still suffer from very large phase-noise levels. Moreover, at larger repetition rates, the RF-signal generation is limited by electrical parasitics and requires shorter absorber lengths. These challenges along with the simultaneous generation of optical pulse trains are discussed in the following part on the basis of a ~40 GHz QD MLL with a 9:1 gain-to-absorber length ratio. A millimeter-wave signal just below 40 GHz is located in the IEEE Ka-band (27–40 GHz) [81] as well as in the ITU EHF-band (30–300 GHz) [82], thus being of great interest for radar systems and wireless communications.

For the electrical signal, the extraction losses of the experimental setup add up to 4.4 dB. For the present, the absorber impedance is not matched to 50 Ω, thus additionally causing coupling losses to the measurement system. Regarding the optical signal, the pulse train is detected using a 50 GHz photodetector, exhibiting a responsivity of 0.57 A/W. The overall losses of the optical path amount to 6.1 dB. A comparison of the RF spectra reveals a peak frequency of 38.61701 GHz and 38.61706 GHz for the directly extracted electrical signal and for the one generated from the optical pulse train, respectively (Fig. 4.9a). The 50 kHz deviation lies well within the carrier 3 dB linewidth of 187.3 kHz. Moreover, as no photodiode is required, the direct extraction yields a noise level lower by 6.3 dB. On the other hand, the peak power of the extracted signal amounts to −32.8 dBm, thus being smaller by 9.8 dB as compared to the optical pulse train. Phase-noise power spectral density traces of both signals show a very similar progression, except for the thermal-noise segment (above 10 MHz carrier offset) (Fig. 4.9b). Nevertheless, the properties of the directly extracted electrical signal from the absorber section essentially conform to the ones obtained for the optical pulse train.

The photocurrent generated by the 110 μm long saturable absorber is 6.37 mA. According to the definition of the differential DC-to-photocurrent conversion

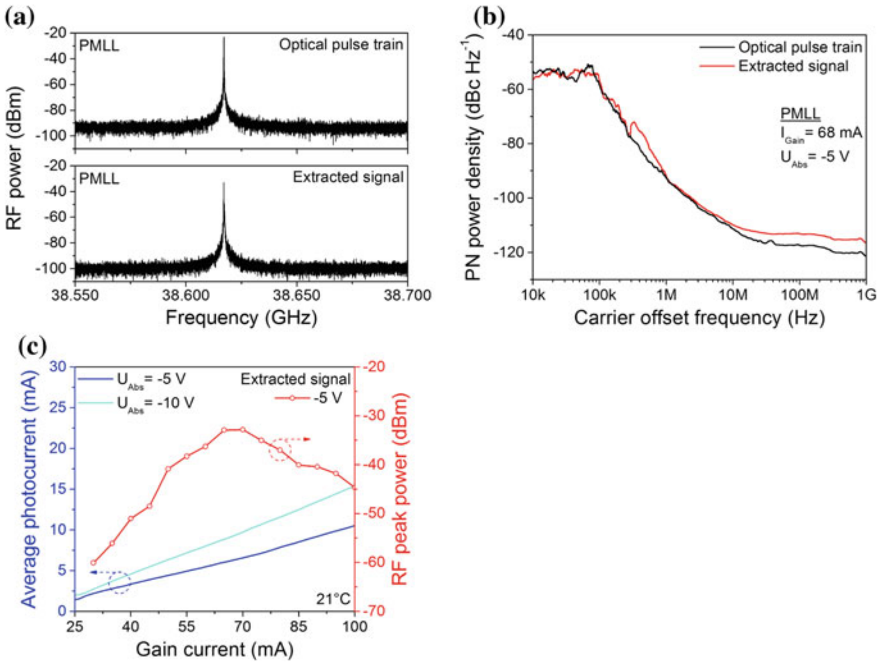


Fig. 4.9 Electrical spectra (**a**) and corresponding SSB phase-noise traces (**b**) of the directly extracted electrical signal from the saturable absorber and the one generated from the optical pulse train using a photodiode, respectively. **c** Absorber-photocurrent and RF-peak-power dependence on gain current and reversed absorber bias. The MLL is operating in the PML regime

efficiency in [80], i.e. the ratio of the average absorber photocurrent to the gain-section DC above threshold, the corresponding value is 16%. An extraction efficiency of 40% is found, thus, as well as the differential DC-to-photocurrent conversion efficiency, being half as large as compared to the values presented in [80]. However, the device used here incorporates a seven times shorter absorber section, which is the most critical design parameter of a MLL-based RF-signal source. MLLs, that are designed exclusively for RF-signal generation, differ from those optimized for emission of short optical pulses. A possibility to increase the photocurrent is to vary the bias point of the MLL. Figure 4.9c depicts the average absorber-photocurrent dependence on the gain current for two different absorber biases (−5 and −10 V). An increasing optical power in the cavity, attributed to increasing gain current, translates to a linearly ascending photocurrent, which is further enhanced at larger reverse biases. However, this progression is not observed for the RF peak power of the extracted signal. Increasing the gain current leads to higher RF peak powers as long the optical pulse energy increases. At a certain point (70 mA in case of −5 V absorber voltage), the pulse forming ability of the absorber

saturates. From this point on, larger gain currents yield wider pulses accompanied by rising CW background. Consequently, the RF peak power is reduced and the carrier linewidth broadens resulting in less stable mode locking, although an increased photocurrent is extracted. By means of a variation of the absorber length from 100 to 250 μm, an up to 8 dB enhancement of the RF peak power of extracted electrical signals is achieved. However, the narrow carrier line in the electrical spectrum rises then from a wide-frequency pedestal, resulting in larger phase-noise power-density values at low carrier offset frequencies. The peak-power benefit vanishes.

Nevertheless, also for MLLs with the best suitable gain-to-absorber length ratio, the still large timing jitter of PML translates to the electrical signal. Thus, OFB, which is described in the previous section, is also introduced here. Figure 4.10a and b shows the RF spectrum and the phase-noise trace for the extracted electrical signal, respectively, while the MLL is operating under optimum OFB conditions. A narrow carrier line with an increased peak power of −27.5 dBm and OFB typical sidebands at 6.5 MHz distance are observed. The SNR and sideband-suppression-ratio (SBSR) are 68.2 dB and 43.5 dB, respectively. Compared to the optical pulse train, a slightly wider white-noise region with a corner frequency of 5 kHz (3 kHz for the optical signal) is present. The 3-dB linewidth and the integrated timing jitter are 3.25 kHz and 359 fs (1.92 kHz and 219 fs for the optical signal), respectively. Although the

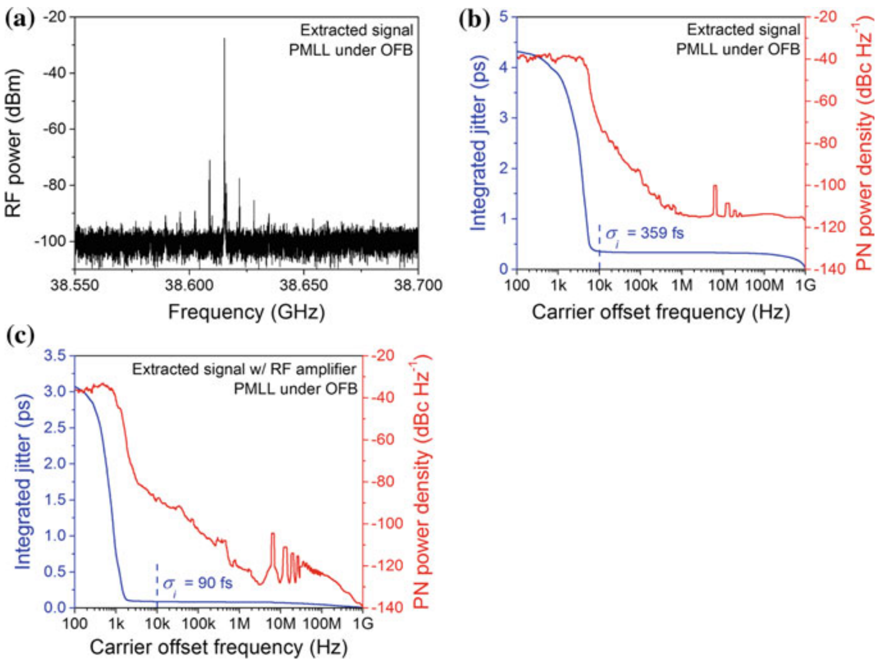


Fig. 4.10 Electrical spectrum (**a**), SSB phase-noise trace and integrated timing jitter (**b**) of the extracted electrical signal for the PMLL under optimal OFB. **c** The SSB phase-noise trace and integrated timing jitter for the extracted signal that is amplified by 24 dB

values are somewhat larger than those for the optical pulse train, a high-quality millimeter-wave signal is obtained [83].

The potential of the QD MLL under OFB operating as a RF-signal source is pointed out in Fig. 4.10c. The extracted electrical signal is amplified by means of a 24 dB broadband power amplifier. A corresponding integrated timing jitter of 90 fs is determined, a reduction by factor of four. Hence, the minimal phase-noise power spectral density of the extracted signal is not given by the MLL itself, but presently limited by extraction capabilities.

4.3.2 Optical Communication

Fiber-based communication systems enable the internet based on broadband data services. Large capacity, low cost and low energy consumption wavelength division multiplexing (WDM) networks with high spectral efficiencies are of particular importance. One of the key approaches for such networks are advanced modulation formats [84]. By means of advanced modulation formats larger spectral efficiencies or the ability to operate the system at a lower symbol rate for a desired data rate are feasible.

Direct modulation and electro-absorption modulated lasers are compact and simple to implement, but suffer from residual chirp and relatively low extinction ratios. On the other hand, MZMs, which are bulky and require broadband RF driver amplifiers, feature larger extinction ratios and the ability to modulate the intensity and phase of the optical field. In a MZM, two voltage-controlled phase modulators induce a relative phase difference to the two optical fields. Their constructive or destructive interference translates to an intensity modulation. For a dual-drive MZM, data and inverted data streams are used to generate chirp-free intensity modulation. Additional phase modulation can be imprinted on the optical carrier, if the two drive voltages differ [85].

A WDM transmitter usually incorporates transverse and longitudinal single-mode CW lasers operating at different wavelengths. Owing to their broad gain spectrum of several nanometers and phase-matched longitudinal modes with free-spectral ranges of tens of gigahertz, quantum-dash or QD MLLs generate a large number of carriers. Instead of multiple laser sources emitting at different wavelengths, WDM transmission can be realized using a single MLL. After filtering of individual carrier and multilevel modulation, aggregated data rates of 0.971 Tbit/s were achieved [52]. Capacities of 26 Tbit/s are feasible using orthogonal frequency-division multiplexing (OFDM) [6].

In this section, on-off keying (OOK) and differential (quadrature) phase-shift keying (D(Q)PSK) return-to-zero (RZ) transmissions based on QD MLL as light sources are demonstrated. A RZ-data signal can be generated by imprinting an electrical RZ waveform on the CW optical carrier using a single MZM and complex electronics [86, 87] or by carving an optical non-return-to-zero (NRZ) signal by means of an additional sinusoidally modulated MZM [88]. In contrast, a QD MLL

is providing an optical pulse train, which, upon NRZ modulation, is translated to RZ signaling without additional components or driver electronics. However, mode-locked pulse emission requires several phase-matched longitudinal modes, which is contrary to the WDM approach.

Different from the NRZ format, the optical intensity in case of the RZ format returns to zero within each bit slot. By means of RZ signaling and the OTDM technique, high bit-rate transmitters can be realized. Based on a 10 GHz MLL with compressed 300 fs pulses, a serial data rate of 10.2 Tbit/s is achieved using 16-QAM and 128-fold OTDM per polarization [89]. RZ signals show generally larger robustness to propagation distortions in optical fibers [84] as well as to intersymbol interference, which leads to lower requirements in terms of optical signal-to-noise ratio (OSNR) as compared to NRZ [90, 91].

For the simplest intensity modulation (OOK), the MZM is biased at the quadrature point of the transfer function and modulated from its minimum to maximum transmission using a swing of V_{π}. Operating at the minimum transmission point and modulated by a $2 \times V_{\pi}$ swing, differential phase modulation (DPSK) signals are generated. The binary signal is then encoded on the phase change between consecutive bits. DPSK reduces the required OSNR for any bit-error ratio (BER) by nominally 3 dB as compared to OOK [92]. DQPSK is based on four phase shifts ($0, +\pi/2, -\pi/2, \pi$) [93]. For a given symbol rate, the bit rate is doubled. A DQPSK signal is usually generated by means of two integrated MZMs operating as phase modulator and producing two modulated optical fields which are then combined with a $\pi/2$ shift [94]. The optical spectrum of DQPSK exhibits the same shape like a DPSK spectrum, but it is compressed in frequency by factor of two [94]. Moreover, only 2–3.1 dB larger OSNR is theoretically required for a laser with optical linewidth of 10–30 MHz [95].

4.3.2.1 On-Off Keying

For the data transmission experiments, QD MLL modules like those presented above are used as light sources. The modules are operating at the SONET/SDH OC-768/STM-256 frequency of 39.813 GHz and in the hybrid regime under a relatively low modulation voltage of 2.0 $V_{p\text{-}p}$. The large jitter in the PML regime is thus reduced down to 338 fs. The pulse train is boosted by means of a QW semiconductor optical amplifier (SOA) and then spectrally filtered using a 2.8 nm wide optical filter. The SOA compensates for the losses induced by the optical filter and the subsequent modulator. The resulting optical spectrum shows 13 longitudinal modes centered around 1310 nm. From corresponding autocorrelation measurements, a pulse FWHM of 2.1 ps is determined. The narrow optical spectrum around 1310 nm is beneficial for data transmission across long fiber distances without regeneration of the signal. The integrated jitter and the pulse width of 338 fs and 2.1 ps, respectively, fulfill the requirements for OTDM up to at least

80 GBd. An amplified, delay-matched and 127-bits-long pseudo-random binary sequence (PRBS7) is superimposed to the MLL pulse train using a single-drive MZM, generating a 39.813 Gbit/s single-polarization RZ-OOK signal. By means of an optical time-division multiplexer (OMUX), in which the data signal is divided into two paths, appropriately delayed and combined, the bit rate is doubled to 79.626 Gbit/s with a true PRBS7. The receiver unit comprises a variable optical attenuator and two praseodymium-doped fiber amplifiers (PDFA) ensuring shot-noise limited operation of the 100 GHz photodetector. Eye diagrams are detected with a 70 GHz sampling oscilloscope. As the maximum bit rate of the error analyzer is below the signal bit rate, an electrical time-division-multiplexing receiver with an integrated 1:2 demultiplexer (DEMUX) is installed to retrieve BERs. The native 79.626 Gbit/s RZ-OOK data signal is then construed as two 39.813 Gbit/s NRZ-OOK output signals (Out 0P and Out 1P) of the DEMUX.

The 79.626 Gbit/s RZ-OOK eye diagram for the maximum received power in back-to-back configuration, depicted in Fig. 4.11a, exhibits a signal-to-noise ratio (SNR) of 11.8 and a root-mean-square (RMS) timing jitter of 452 fs. Some smearing of the optically multiplexed data streams is observed, mainly caused by the bandwidth limitation of the sampling oscilloscope. Slight differences in successive eye crossings can be identified due to some interleaving non-precisions of the two pulse trains in the OMUX. The 39.813 Gbit/s NRZ-DEMUX output eye diagrams at the error-free threshold (BER of 10^{-9}) are shown as inset in Fig. 4.11b. BER curves of both signals demonstrate error-free performance without error floor to below 10^{-10} (Fig. 4.11b). At a BER of 10^{-9} and around −22 dBm of received power, a 0.6 dB penalty is found between the two streams, a consequence of thermal instabilities of the OMUX unit [51].

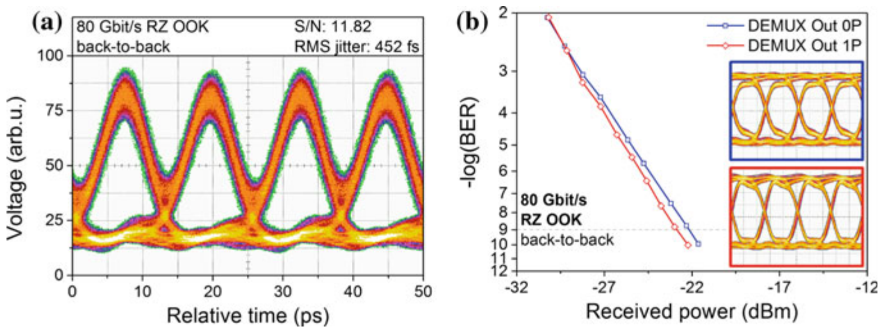


Fig. 4.11 **a** Back-to-back eye diagram of the 80 Gbit/s RZ-OOK data signal for maximum received optical power. **b** BER versus received power of DEMUX outputs 0P and 1P and corresponding 40 Gbit/s eye diagrams at BER of 10^{-9} as *inset*

4.3.2.2 Differential (Quadrature) Phase-Shift Keying

Single-polarization D(Q)PSK data transmission is tested using the same operating conditions of the MLL module as for OOK. The 39.813 GBd DPSK signal is generated by means of a dual-drive MZM biased at its minimum transmission point and modulated with a swing of $2 \times V_\pi$. DQPSK is usually generated using an in-phase/quadrature (I/Q) modulator, which independently modulates both components of the optical field, providing precise phase changes. Here, DQPSK is realized by successively sequenced dual-drive MZMs and a phase modulator (PM), where the latter generates 0 or $\pi/2$ phase shifts. The PM demands an exact drive voltage to obtain precise $\pi/2$ phase modulation, which is hard to achieve. This transmitter setup may cause additional power penalties and requires larger OSNRs due to imperfect phase shifts.

The modulation of the phase of the electrical field cannot be directly decoded by simple square-law photodiodes. Hence, the receiver is extended by a delay interferometer (DI) and a 40 GHz balanced photodetector. In the DI, a part of the data signal is tapped, delayed by the free spectral range (FSR) and then merged with the other signal path. Two DI outputs, the destructive and the constructive port, are available. At the first one, the optical fields of two successive bits interfere destructively in absence of a phase change and constructively in presence of a phase change. At the constructive DI port, the logically inverted data signal is generated. The deconstructive and constructive outputs are detected using a balanced detector, consisting of two matched photodiodes, which yields the photocurrent difference of the two DI ports. The 2.7 dB larger receiver sensitivity of DPSK as compared to OOK is preserved using a balanced detector. In order to detect DQPSK signals, the data stream is usually split into two equal paths, each equipped with a DI (with a $+\pi/4$ and $-\pi/4$ phase differences in the arms of the DIs) and a balanced detector, enabling simultaneous measurement of the in-phase and quadrature tributaries. Here, the D(Q)PSK signals are encoded using a single receiver unit, by setting a 39.813 GHz FSR and a voltage-controlled, carrier-wavelength related phase difference between the two DI arms. The DPSK demodulation is obtained at a phase difference of 0 and π. In case of DQPSK, the additional tributary is detected at $\pi/2$ and $3/2\pi$. Consequently, DQPSK signaling is completely characterized by two tributaries having a $\pi/2$ phase difference.

BER curves of the 39.813 GBd RZ-DPSK and DQPSK data transmission experiments, obtained back-to-back (b-t-b) and across 45 km of SMF-28, are depicted in Fig. 4.12a. For DPSK, the error-free threshold is reached at a received optical power of −31.2 dBm. The DQPSK signals are processed by two tributaries, termed Tr 1 and Tr 2, having a $\pi/2$ phase difference. Error-free operation of Tr 1 and Tr 2 require a received power of −25.8 dBm and −24.5 dBm, respectively. The implementation of DQPSK using successively sequenced MZM and PM has led to a 6 dB penalty in receiver sensitivity as compared to DPSK. The imperfection of the transmitter and receiver units become apparent, resulting in somewhat

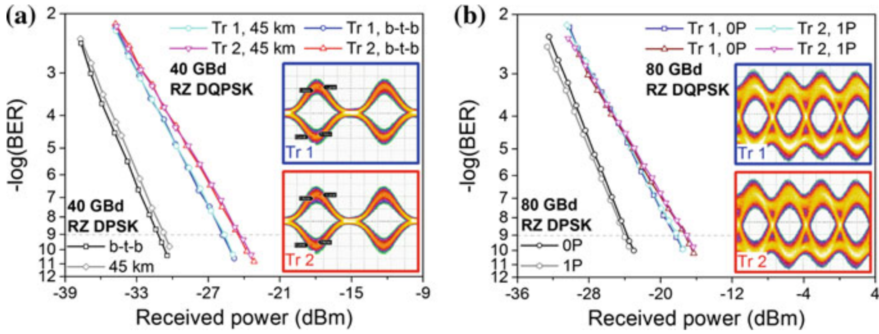


Fig. 4.12 **a** BER versus received power of 40 GBd RZ-DPSK and DQPSK signals in the back-to-back (b-t-b) configuration and across 45 km of SMF-28. The *inset* shows RZ-DQPSK eye diagrams of Tr 1 and Tr 2 for back-to-back at the maximum received optical power. **b** 80 GBd RZ DPSK and DQPSK in back-to-back configuration. DPSK: BER curves of DEMUX outputs 0P/1P. DQPSK: BER versus received power of Tr 1 and Tr 2 and the DEMUX outputs 0P and 1P. Eye diagrams of Tr 1 and Tr 2 are shown as inset, detected at the maximum received power in front of the DEMUX

larger penalties than theoretically expected. The 39.813 GBd RZ-DQPSK eye diagrams of Tr 1 and Tr 2, shown as inset of Fig. 4.12a, exhibit SNR of 8.5 and 8.1, respectively. The DPSK and DQPSK transmissions across 45 km of SMF-28, performed without optical regeneration, does not induce any noticeably signal distortion and thus no power penalty. Owing to the relatively narrow emission spectrum of the QD MLL centered around 1310 nm, the transmission distance, which is mainly limited by the fiber attenuation, may be largely extended by insertion of PDFAs or SOAs.

For 79.626 GBd RZ DPSK and DQPSK, an OMUX is installed after the modulators in order to double the bit rate, the same way as described for OOK signaling. The FSR of the DI is now set to 79.626 GHz, equal to the symbol duration. Back-to-back BER curves for DPSK DEMUX outputs 0P and 1P, as well as for the DQPSK tributaries and corresponding DEMUX outputs are presented in Fig. 4.12b, all complying with BERs below 2.5×10^{-10} without error floor. The average power penalty upon OTDM is 7.3 dB. The 79.626 GBd RZ-DQPSK eye diagrams of Tr 1 and Tr 2 at maximum received power, depicted as inset of Fig. 4.12b, exhibit a SNR of 6.7 and 6.5, respectively. The eye opening margin does not support error-free data transmission across 45 km of SMF-28. The comparison of the optically multiplexed RZ-DPSK and RZ-OOK signals reveal a 1.5 dB receiver sensitivity advantage of the phase modulation. Note that the RZ-OOK transmission system benefits from the 100 GHz bandwidth of the photodetector, being 2.5 times larger than the bandwidth of the balanced detector used in case of RZ DPSK [96, 97].

4.4 Conclusion

1.31 μm QD MLLs operating in several configurations allowing repetition frequency tuning and jitter reduction have been presented. HML using sub-harmonic RF signals offers the possibility to generate low-jitter optical signals even at high frequencies, at which it is challenging to provide an adequate electrical signal. CW and dual-tone injection seeding have led to optical linewidth narrowing of all MLL modes, strong jitter reduction and wide frequency tunability, but at the expense of somewhat wider pulses. The influence of OFB on PMLLs has revealed five different regimes, including a resonant one being of practical interest and offering a cost-effective way to stabilize the pulse emission. Moreover, this technique does not consume additional power. Operating under OFB, a single MLL can provide low-jitter optical and electrical pulse trains at the same time, thus radically simplifying photonic sources for millimeter-wave signal generation. Up to 160 Gbit/s data transmission based on advanced modulation format is demonstrated using a HMLL module combined with modulators.

Acknowledgements Many colleagues contributed to the success of this work. The authors would like to thank expressly for each individual who has been in recent years with us and especially Moritz Kleinert and Marc Spiegelberg for their untiring efforts. The work was funded by DFG in the framework of the SFB 787.

References

1. H.F. Liu, M. Fukazawa, Y. Kawai, T. Kamiya, Gain-switched picosecond pulse (<10 ps) generation from 1.3 μm InGaAsP laser-diodes. IEEE J. Quantum Electron. **25**, 1417–1425 (1989)
2. P.P. Vasilev, Ultrashort pulse generation in diode-lasers. Opt. Quant. Electron. **24**, 801–824 (1992)
3. F. Van Dijk, B. Charbonnier, S. Constant, A. Enard, S. Fedderwitz, S. Formont, et al., Quantum dash mode-locked lasers for millimeter wave signal generation and transmission, in *Annual Meeting of the IEEE Photonics Society*, Denver, CO, 2010, pp. 187–188
4. A. Stohr, S. Babiel, P.J. Cannard, B. Charbonnier, F. van Dijk, S. Fedderwitz et al., Millimeter-wave photonic components for broadband wireless systems. IEEE Trans. Microw. Theory Tech. **58**, 3071–3082 (2010)
5. W.H. Knox, Ultrafast technology in telecommunications. IEEE J. Sel. Top. Quantum Electron. **6**, 1273–1278 (2000)
6. D. Hillerkuss, R. Schmogrow, T. Schellinger, M. Jordan, M. Winter, G. Huber et al., 26 Tbit/s line-rate super-channel transmission utilizing all-optical fast Fourier transform processing. Nat. Photonics **5**, 364–371 (2011)
7. C. Dorrer, High-speed measurements for optical telecommunication systems. IEEE J. Sel. Top. Quantum Electron. **12**, 843–858 (2006)
8. X. Huang, A. Stintz, H. Li, L.F. Lester, J. Cheng, K.J. Malloy, Passive mode-locking in 1.3 μm two-section InAs quantum dot lasers. Appl. Phys. Lett. **78**, 2825 (2001)
9. A. Gubenko, D. Livshits, I. Krestnikov, S. Mikhrin, A. Kozhukhov, A. Kovsh et al., High-power monolithic passively modelocked quantum-dot laser. Electron. Lett. **41**, 1124 (2005)

10. M. Laemmlin, G. Fiol, C. Meuer, M. Kuntz, F. Hopfer, A.R. Kovsh et al., Distortion-free optical amplification of 20–80 GHz modelocked laser pulses at 1.3 [micro sign]m using quantum dots. Electron. Lett. **42**, 697 (2006)
11. D. Bimberg, Quantum dot based nanophotonics and nanoelectronics. Electron. Lett. **44**, 168–170 (2008)
12. R.L. Sellin, C. Ribbat, M. Grundmann, N.N. Ledentsov, D. Bimberg, Close-to-ideal device characteristics of high-power InGaAs/GaAs quantum dot lasers. Appl. Phys. Lett. **78**, 1207 (2001)
13. O.B. Shchekin, D.G. Deppe, 1.3 mu m InAs quantum dot laser with T-o = 161 K from 0 to 80 degrees C. Appl. Phys. Lett. **80**, 3277–3279 (2002)
14. D.B. Malins, A. Gomez-Iglesias, S.J. White, W. Sibbett, A. Miller, E.U. Rafailov, Ultrafast electroabsorption dynamics in an InAs quantum dot saturable absorber at 1.3 μm. Appl. Phys. Lett. **89**, 171111 (2006)
15. J. Gomis-Bresco, S. Dommers-Volkel, O. Schops, Y. Kaptan, O. Dyatlova, D. Bimberg, et al., Time-resolved amplified spontaneous emission in quantum dots. Appl. Phys. Lett. **97** (2010)
16. D.G. Deppe, H. Huang, O.B. Shchekin, Modulation characteristics of quantum-dot lasers: the influence of p-type doping and the electronic density of states on obtaining high speed. IEEE J. Quantum Electron. **38**, 1587–1593 (2002)
17. A.R. Kovsh, N.A. Maleev, A.E. Zhukov, S.S. Mikhrin, A.P. Vasil'ev, E.A. Semenova, et al., InAs/InGaAs/GaAs quantum dot lasers of 1.3 μm range with enhanced optical gain. J. Cryst. Growth **251**, 729–736 (2003)
18. G. Fiol, C. Meuer, H. Schmeckebier, D. Arsenijević, S. Liebich, M. Laemmlin, et al., Quantum-dot semiconductor mode-locked lasers and amplifiers at 40 GHz. IEEE J. Quantum Electron. **45**, 1429–1435 (2009)
19. J.K. Mee, R. Raghunathan, J.B. Wright, L.F. Lester, Device geometry considerations for ridge waveguide quantum dot mode-locked lasers. J. Phys. D Appl. Phys. **47**, 233001 (2014)
20. E. Rouvalis, D. Arsenijević, M. Spiegelberg, T. Sadeev, R. Ziegler, A.G. Steffan, et al., 40 GHz quantum quantum-dot mode-locked laser packaged module operating at 1310 nm, in *Asia Communications and Photonics Conference (ACP)*, Shanghai, China, 2014, pp. 1–3
21. H.A. Haus, Theory of mode locking with a slow saturable absorber. IEEE J. Quantum Electron. **11**, 736–746 (1975)
22. D.J. Derickson, R.J. Helkey, A. Mar, J.R. Karin, J.G. Wasserbauer, J.E. Bowers, Short pulse generation using multisegment mode-locked semiconductor-lasers. IEEE J. Quantum Electron. **28**, 2186–2202 (1992)
23. M.G. Thompson, A.R. Rae, X. Mo, R.V. Penty, I.H. White, InGaAs quantum-dot mode-locked laser diodes. IEEE J. Sel. Top. Quantum Electron. **15**, 661–672 (2009)
24. E.A. Viktorov, P. Mandel, M. Kuntz, G. Fiol, D. Bimberg, A.G. Vladimirov et al., Stability of the mode-locked regime in quantum dot lasers. Appl. Phys. Lett. **91**, 231116 (2007)
25. A.G. Vladimirov, U. Bandelow, G. Fiol, D. Arsenijević, M. Kleinert, D. Bimberg, et al., Dynamical regimes in a monolithic passively mode-locked quantum dot laser, J. Opt. Soc. Am. B-Opt. Phys. **27**, 2102–2109 (2010)
26. M.G. Thompson, A. Rae, R.L. Sellin, C. Marinelli, R.V. Penty, I.H. White et al., Subpicosecond high-power mode locking using flared waveguide monolithic quantum-dot lasers. Appl. Phys. Lett. **88**, 133119 (2006)
27. X.D. Huang, A. Stintz, H. Li, A. Rice, G.T. Liu, L.F. Lester et al., Bistable operation of a two-section 1.3-mu m InAs quantum dot laser—absorption saturation and the quantum confined Stark effect. IEEE J. Quantum Electron. **37**, 414–417 (2001)
28. M.G. Thompson, K.T. Tan, C. Marinelli, K.A. Williams, R.V. Penty, I.H. White et al., Transform-limited optical pulses from 18 GHz monolithic modelocked quantum dot lasers operating at ~1.3 [micro sign]m. Electron. Lett. **40**, 346 (2004)
29. M. Kuntz, G. Fiol, M. Lämmlin, D. Bimberg, M.G. Thompson, K.T. Tan et al., Direct modulation and mode locking of 1.3 μm quantum dot lasers. New J. Phys. **6**, 181 (2004)

30. H. Schmeckebier, G. Fiol, C. Meuer, D. Arsenijević, D. Bimberg, Complete pulse characterization of quantum-dot mode-locked lasers suitable for optical communication up to 160 Gbit/s. Opt. Express **18**, 3415–3425 (2010)
31. D. von der Linde, Characterization of the noise in continuously operating mode-locked lasers. Appl. Phys. B **39**, 201–217 (1986)
32. ITU, The control of jitter and wander within the optical transport network (OTN)—recommendation G.8251, in *Series G: Transmission Systems and Media, Digital Systems and Networks*, ITU-T (2010)
33. D. Eliyahu, R.A. Salvatore, A. Yariv, Noise characterization of a pulse train generated by actively mode-locked lasers. J. Opt. Soc. Am. B-Opt. Phys. **13**, 1619–1626 (1996)
34. H.A. Haus, A. Mecozzi, Noise of mode-locked lasers. IEEE J. Quantum Electron. **29**, 983–996 (1993)
35. D. Eliyahu, R.A. Salvatore, A. Yariv, Effect of noise on the power spectrum of passively mode-locked lasers. J. Opt. Soc. Am. B-Opt. Phys. **14**, 167–174 (1997)
36. F. Kefelian, S. O'Donoghue, M.T. Todaro, J.G. McInerney, G. Huyet, RF linewidth in monolithic passively mode-locked semiconductor laser. IEEE Photonics Technol. Lett. **20**, 1405–1407 (2008)
37. L.A. Jiang, S.T. Wong, M.E. Grein, E.P. Ippen, H.A. Haus, Measuring timing jitter with optical cross correlations. IEEE J. Quantum Electron. **38**, 1047–1052 (2002)
38. M.G. Thompson, C. Marinelli, K.T. Tan, K. A. Williams, R.V. Penty, I.H. White, et al., 10 GHz hybrid modelocking of monolithic InGaAs quantum dot lasers. Electron. Lett. **39**, 1121–1122 (2003)
39. M. Kuntz, G. Fiol, M. Laemmlin, C. Meuer, D. Bimberg, High-speed mode-locked quantum-dot lasers and optical amplifiers. Proc. IEEE **95**, 1767–1778 (2007)
40. B. Huettl, R. Kaiser, Monolithically integrated optical pulse sources for ultra-high speed applications, in *URSI General Assemblies*, New Delhi, India, 2005
41. R. Kaiser, B. Hüttl, W. Rehbein, H. Stolpe, H. Heidrich, S. Fidorra, et al., Repetition rate and wavelength tuning of monolithic 40 GHz mode-locked lasers based on InP, in *Conference on Indium Phosphide and Related Materials*, Santa Barbara, CA, 2003, pp. 255–258
42. M.G. Thompson, D. Larson, A.R. Rae, K. Yvind, R.V. Penty, I.H. White, et al., Monolithic hybrid and passive mode-locked 40 GHz quantum dot laser diodes, in *European Conference on Optical Communication (ECOC)*, Cannes, France, 2006, pp. 1–2
43. K. Yvind, P.M. Smowton, D. Larsson, J. Mørk, J.M. Hvam, M. Thompson, et al., Low-noise monolithic mode-locked semiconductor lasers through low-dimensional structures, in *SPIE Photonics West*, San Jose, CA, 2008, pp. 69090A-1–69090A-9
44. G. Fiol, D. Arsenijević, D. Bimberg, A.G. Vladimirov, M. Wolfrum, E.A. Viktorov, et al., Hybrid mode-locking in a 40 GHz monolithic quantum dot laser. Appl. Phys. Lett. **96**, 011104–011104-3 (2010)
45. A.G. Vladimirov, M. Wolfrum, G. Fiol, D. Arsenijević, D. Bimberg, E. Viktorov, et al., Locking characteristics of a 40-GHz hybrid mode-locked monolithic quantum dot laser, in *SPIE Photonics Europe*, Brussels, Belgium, 2010, pp. 77200Y-77200Y-8
46. H.A. Haus, Modelocking of semiconductor-laser diodes. Jpn. J. Appl. Phys. **20**, 1007–1020 (1981)
47. B. Hüttl, R. Kaiser, C. Kindel, S. Fidorra, W. Rehbein, H. Stolpe et al., Experimental investigations on the suppression of Q switching in monolithic 40 GHz mode-locked semiconductor lasers. Appl. Phys. Lett. **88**, 221104 (2006)
48. A.H. Nayfeh, D.T. Mook, *Nonlinear Oscillations* (Wiley, New York, NY, 1979)
49. R. Arkhipov, A. Pimenov, M. Radziunas, D. Rachinskii, A.G. Vladimirov, D. Arsenijević, et al., Hybrid mode locking in semiconductor lasers: simulations, analysis, and experiments. IEEE J. Sel. Top. Quantum Electron. **19**, 1100208–1100208-8 (2013)
50. V.I. Arnold, *Geometrical Methods in the Theory of Ordinary Differential Equations*, 2nd edn. (Springer-Verlag, New York, NY, 1988)
51. D. Arsenijević, M. Kleinert, M. Spiegelberg, M. Stubenrauch, D. Bimberg, 1.31 μm quantum-dot hybrid mode-locked lasers for optical time-division multiplexing, in

International Conference on Transparent Optical Networks (ICTON), Budapest, Hungary, 2015, pp. 1–4
52. W. Freude, J. Pfeifle, R. Watts, I. Shkarban, S. Wolf, V. Vujicic, et al., Phase-noise compensated carriers from an optical frequency comb allowing terabit transmission, in *International Conference on Transparent Optical Networks (ICTON)*, Budapest, Hungary, 2015, pp. 1–4
53. A. Takada, W. Imajuku, Linewidth narrowing and optical phase control of mode-locked semiconductor ring laser employing optical injection locking. IEEE Photonics Technol. Lett. **9**, 1328–1330 (1997)
54. M. Teshima, K. Sato, M. Koga, Experimental investigation of injection locking of fundamental and subharmonic frequency-modulated active mode-locked laser diodes. IEEE J. Quantum Electron. **34**, 1588–1596 (1998)
55. J. Kim, P.J. Delfyett, Interband optical pulse injection locking of quantum dot mode-locked semiconductor laser. Opt. Express **16**, 11153–11161 (2008)
56. T. Habruseva, G. Huyet, and S.P. Hegarty, Dynamics of quantum-dot mode-locked lasers with optical injection. IEEE J. Sel. Top. Quantum Electron. **17**, 1272–1279 (2011)
57. G. Fiol, M. Kleinert, D. Arsenijević, D. Bimberg, 1.3 μm range 40 GHz quantum-dot mode-locked laser under external continuous wave light injection or optical feedback. Semicond. Sci. Technol. **26**, 014006–014006-5 15 (2011)
58. L. Goldberg, H.F. Taylor, J.F. Weller, Fm sideband injection locking of diode-lasers. Electron. Lett. **18**, 1019–1020 (1982)
59. T. Habruseva, S. O'Donoghue, N. Rebrova, D.A. Reid, L.P. Barry, D. Rachinskii et al., Quantum-dot mode-locked lasers with dual-mode optical injection. IEEE Photonics Technol. Lett. **22**, 359–361 (2010)
60. T. Habruseva, D. Arsenijević, M. Kleinert, D. Bimberg, G. Huyet, S.P. Hegarty, Optimum phase noise reduction and repetition rate tuning in quantum-dot mode-locked lasers. Appl. Phys. Lett. **104**, 021112–021112-4 (2014)
61. R. Tkach, A. Chraplyvy, Regimes of feedback effects in 1.5-μm distributed feedback lasers. J. Lightw. Technol. **4**, 1655–1661 (1986)
62. C. Otto, K. Lüdge, A.G. Vladimirov, M. Wolfrum, E. Schöll, Delay-induced dynamics and jitter reduction of passively mode-locked semiconductor lasers subject to optical feedback. New J. Phys. **14**, 113033 (2012)
63. C.Y. Lin, F. Grillot, N.A. Naderi, Y. Li, L.F. Lester, rf linewidth reduction in a quantum dot passively mode-locked laser subject to external optical feedback. Appl. Phys. Lett. **96**, 051118 (2010)
64. A. Akrout, A. Shen, A. Enard, G.H. Duan, F. Lelarge, A. Ramdane, Low phase noise all-optical oscillator using quantum dash modelocked laser. Electron. Lett. **46**, 73 (2010)
65. E.A. Avrutin, S. Xibin, B.M. Russell, Optical feedback tolerance of mode-locked laser diodes and some feedback reduction methods: a numerical investigation. Opt. Quant. Electron. **40**, 1175–1180 (2008)
66. D. Arsenijević, M. Kleinert, D. Bimberg, Phase noise and jitter reduction by optical feedback on passively mode-locked quantum-dot lasers. Appl. Phys. Lett. **103**, 231101–231101-4 (2013)
67. D. Arsenijević, M. Kleinert, D. Bimberg, Breakthroughs in photonics 2013: passive mode-locking of quantum-dot lasers. IEEE Photonics J. **6**, 0700306–0700306-6 (2014)
68. C. Simos, H. Simos, T. Nikas, D. Syvridis, Compact optical displacement sensing by detection of microwave signals generated from a monolithic passively mode-locked laser under feedback, vol. 9506, (2015), p. 95060F
69. X.Q. Qi, J.M. Liu, Photonic microwave applications of the dynamics of semiconductor lasers. IEEE J. Sel. Top. Quantum Electron. **17**, 1198–1211 (2011)
70. E.H. Bottcher, E. Droge, D. Bimberg, 200 GHz distributed InGaAs metal-semiconductor-metal photodetectors for the long-wavelength regime, in *International Symposium on Compound Semiconductors*, St. Petersburg, Russia, 1997, pp. 55–60

71. H. Ito, S. Kodama, Y. Muramoto, T. Furuta, T. Nagatsuma, T. Ishibashi, High-speed and high-output InP-InGaAs unitraveling-carrier photodiodes. IEEE J. Sel. Top. Quantum Electron. **10**, 709–727 (2004)
72. H. Ito, T. Furuta, S. Kodama, T. Ishibashi, InP/lnGaAs uni-travelling-carrier photodiode with 310 GHz bandwidth. Electron. Lett. **36**, 1809–1810 (2000)
73. M.J. Fice, E. Rouvalis, L. Ponnampalam, C.C. Renaud, A.J. Seeds, Telecommunications technology-based terahertz sources. Electron. Lett. **46**, S28–S31 (2010)
74. S. Osborne, S. O'Brien, E.P. O'Reilly, P.G. Huggard, B.N. Ellison, Generation of CW 0.5 THz radiation by photomixing the output of a two-colour 1.49 mu m Fabry-Perot diode laser. Electron. Lett. **44**, 296–298 (2008)
75. S.C. Chan, Analysis of an optically injected semiconductor laser for microwave generation. IEEE J. Quantum Electron. **46**, 421–428 (2010)
76. A. Hurtado, J. Mee, M. Nami, I.D. Henning, M.J. Adams, L.F. Lester, Tunable microwave signal generator with an optically-injected 1310 nm QD-DFB laser. Opt. Express **21**, 10772–10778 (2013)
77. L.A. Johansson, A.J. Seeds, Millimeter-wave modulated optical signal generation with high spectral purity and wide-locking bandwidth using a fiber-integrated optical injection phase-lock loop. IEEE Photonics Technol. Lett. **12**, 690–692 (2000)
78. D. Novak, Z. Ahmed, R.B. Waterhouse, R.S. Tucker, Signal generation using pulsed semiconductor-lasers for application in millimeter-wave wireless links. IEEE Trans. Microw. Theory Tech. **43**, 2257–2262 (1995)
79. D.J. Derickson, R.J. Helkey, A. Mar, J.G. Wasserbauer, Y.G. Wey, J.E. Bowers, Microwave and millimeter wave signal generation using mode-locked semiconductor lasers with intra-waveguide saturable absorbers, in *IEEE MTT-S International Microwave Symposium Digest*, Albuquerque, NM, 1992, pp. 753–756
80. C.Y. Lin, Y.C. Xin, J.H. Kim, C.G. Christodoulou, L.F. Lester, Compact optical generation of microwave signals using a monolithic quantum dot passively mode-locked laser. IEEE Photonics J. **1**, 236–244 (2009)
81. IEEE, IEEE Standard for Letter Designations for Radar-Frequency Bands—IEEE Std 521 (2002)
82. ITU, Nomenclature of the Frequency and Wavelength Bands used in Telecommunications—Recommendation ITU-R V.431–7, ITU-T (2000)
83. D. Arsenijević, M. Kleinert, D. Bimberg, Optoelectronic Oscillator, Germany Patent PCT/DE 2014/200257, WO 2014/202074, US 2016/0149377, 2014
84. P.J. Winzer, R.J. Essiambre, Advanced optical modulation formats. Proc. IEEE **94**, 952–985 (2006)
85. H. Kim, A.H. Gnauck, Chirp characteristics of dual-drive Mach-Zehnder modulator with a finite DC extinction ratio. IEEE Photonics Technol. Lett. **14**, 298–300 (2002)
86. N.M. Froberg, G. Raybon, U. Koren, B.I. Miller, M.G. Young, M. Chien et al., Generation of 12.5-Gbit/s soliton data stream with an integrated laser-modulator transmitter. Electron. Lett. **30**, 1880–1881 (1994)
87. X. Liu, Y. Kao, Generation of RZ-DPSK using a single mach-zehnder modulator and novel driver electronics, in *European Conference on Optical Communication (ECOC)*, Stockholm, Sweden, 2004, pp. 1–2
88. J. Leibrich, C. Wree, W. Rosenkranz, CF-RZ-DPSK for suppression of XPM on dispersion-managed long-haul optical WDM transmission on standard single-mode fiber. IEEE Photonics Technol. Lett. **14**, 155–157 (2002)
89. T. Richter, E. Palushani, C. Schmidt-Langhorst, M. Nölle, R. Ludwig, J.K. Fischer, et al., Single wavelength channel 10.2 Tb/s TDM-Data capacity using 16-QAM and coherent detection, in *Optical Fiber Communication Conference and Exposition (OFC), National Fiber Optic Engineers Conference (NFOEC)*, Los Angeles, CA, 2011, pp. 1–3
90. L. Boivin, G.J. Pendock, Receiver sensitivity for optically amplified RZ signals with arbitrary duty circle, in *Optical Amplifiers and Their Applications (OAA)*, Nara, Japan, 1999

91. W. Idler, A. Klekamp, R. Dischler, J. Lazaro, A. Konczykowska, System performance and tolerances of 43 Gb/s ASK and DPSK modulation formats, in *European Conference on Optical Communication (ECOC)*, Rimini, Italy, 2003, pp. 1006–1007
92. A.H. Gnauck, P.J. Winzer, Optical phase-shift-keyed transmission. J. Lightwave Technol. **23**, 115–130 (2005)
93. J.G. Proakis, *Digital Communications* (McGraw-Hill, Boston, 2001)
94. R.A. Griffin, A.C. Carter, Optical differential quadrature phase-shift key (oDQPSK) for high capacity optical transmission, in *Optical Fiber Communication Conference (OFC)*, Anaheim, CA, 2002, pp. 367–368
95. N.S. Avlonitis, E.M. Yeatman, Performance evaluation of optical DQPSK using saddle point approximation. J. Lightwave Technol. **24**, 1176–1185 (2006)
96. D. Arsenijević, D. Bimberg, Quantum-dot lasers for 35 Gbit/s pulse-amplitude modulation and 160 Gbit/s differential quadrature phase-shift keying, in *SPIE Photonics Europe*, Brussels, Belgium, 2016, pp. 98920S–98920S-10
97. D. Arsenijević, H. Schmeckebier, M. Kleinert, E. Rouvalis, R. Ziegler, A.G. Steffan, et al., Quantum-dot mode-locked lasers for microwave-signal generation and 160 Gbps optical communication, in *IEEE Photonics Conference (IPC)*, Reston, VA, 2015

Chapter 5
Nanophotonic Approach to Energy-Efficient Ultra-Fast All-Optical Gates

Grégory Moille, Sylvain Combrié and Alfredo De Rossi

Abstract All-optical processing is based on fast nonlinear effects, such that light can be used to control light. The development of a novel class of low-loss semiconductor optical resonators, capable of field confinement close to the diffraction limit, has decreased the power level required to trigger nonlinear effects by several orders of magnitude. We review a decade of research on all-optical gates aiming both at fast and energy-efficient operation, with the prospect of integration on a silicon photonics platform.

5.1 Introduction: A Case for All-Optical Signal Processing

Data signals are ubiquitous in the everyday life and the resulting ever increasing demand in bandwidth for communication and processing is raising concerns about the alarming share of electric power they consume. This situation is producing a paradigm change in information technologies where photonics is promised to play a crucial role. The application domain of photonic technologies, originally centred on long-distance communication, is now extending to short-range data links and optical interconnects[1] as well as wireless and other microwave-related technologies [12, 44], particularly radar systems [22].

In future optical interconnect and other photonic assisted data processing systems, a large number of signals will be transferred into the optical domain and routed to the appropriate destination. Here, the question of minimizing the number of signal

[1]This topic is broadly covered elsewhere in this book.

G. Moille · S. Combrié (✉) · A. De Rossi
Thales Research and Technology, 1, av. A. Fresnel, Campus Polytechnique, 91767 Palaiseau, France
e-mail: sylvain.combrie@thalesgroup.com

A. De Rossi
e-mail: alfredo.derossi@thalesgroup.com

© Springer International Publishing AG 2017

G. Eisenstein and D. Bimberg (eds.), *Green Photonics and Electronics*,
NanoScience and Technology, DOI 10.1007/978-3-319-67002-7_5

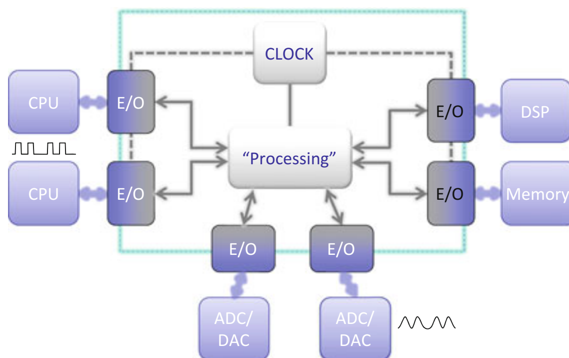


Fig. 5.1 Possible synergy between electronics and fast optical processing where all-optical routing is used to avoid additional conversion in the electric domain

conversions from and to the electric domain becomes relevant. As an example, performing any sort of "operation", for instance detecting the header of a message or extracting a subset of data from a high speed channel, requires the conversion into the electric domain. Thus, these signals need to be detected, processed, and reconverted into the optical domain. Thus, wherever possible, it would be convenient to have some of these operations to be performed in the optical domain. This situation is represented in Fig. 5.1, where digital and analog data are converted in the optical domain for transmission and for a few critical operations such as pattern search. Optical signals are also needed for synchronisation and precise analog to digital conversion [22, 76].

Optical correlation techniques are in principle extremely effective for pattern search, and could be used to reduce the amount of data to be processed by electronics drastically [81]. This is shown in Fig. 5.2. Another example of all-optical processing being implemented with fibre optics and bistable optical elements is the all-optical header recognition and signal routing [19].

In this book chapter, a critical optical function named all-optical gating (AOG) is considered. This consists in sampling an optical signal at an extremely well defined time, which is controlled by another optical signal. A related application domain of all-optical gates is Optical Time Domain Division (OTDM). OTDM is a convenient alternative to more complex signal modulation formats when high-speed and simplicity, but not spectral efficiency, matters, e.g. in dedicated communication links, optical interconnects or other kinds of short-range but high-speed channels. Also, high-speed signal measurement techniques are based on optical gating [79].

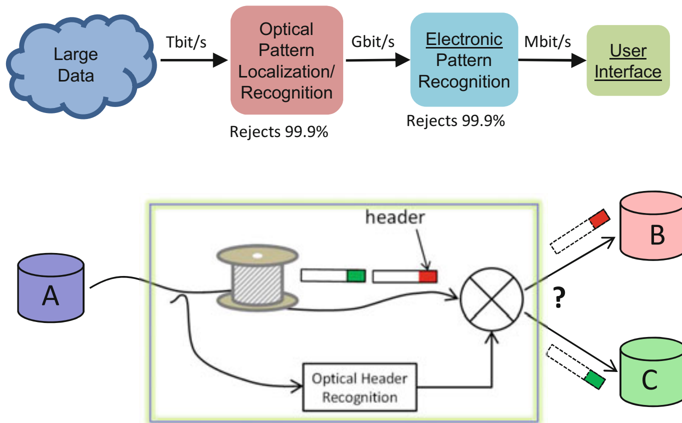


Fig. 5.2 All-optical processing could be effective for fast information retrieval, as a front-end to electronics. Top, optical techniques could be used to search for a specific information within a large amount of data; bottom, scheme for all-optical routing. Adapted from [81]

5.2 Integrated All-Optical Gate

5.2.1 Technologies for Integrated On-Chip All-Optical Processing

All-optical gates are based on fast nonlinear effects. One of the most popular implementation of this concept is the nonlinear optical loop mirror (NOLM), widely used for ultra-fast all-optical manipulation of high-speed optical signals [51]. It exploits the nonlinear index change (Kerr effect) in a Sagnac fibre interferometer [18]. The NOLM is an extremely fast (sub-ps gating time), simple and robust device where the intrinsically weak nonlinear response of silica is compensated by the large interaction length (> 100 m) and very low losses, which makes operation with Watt-level peak power pump possible.

The seek for an integrated equivalent of the nonlinear optical fibre started in the late 80's and considered semiconductors [68] because of their nonlinear response, about 3 orders of magnitude larger than in Silica [62]. Semiconductor integrated optics has taken off almost two decades later, however, with the fabrication of low-loss and tightly confined waveguides [21, 65].

In parallel, the large and relatively fast optical nonlinearities in active semiconductor devices, in particular semiconductor optical amplifiers SOA, have raised considerable interest [26, 43] for high-speed all-optical signal processing. Recently, SOAs have been operated as fast optical gates with close to 100 Gb/s data rate, based on Cross-Phase-Modulation in a Quantum Dot active region [47]. Even faster operation have been achieved in similar devices by exploiting Four-Wave-Mixing [46]. The underlying dynamics of free carriers here is characterized by a slow time constant

(≈ 100 ps) requiring special care (e.g. differential mode architectures with two SOAs) to avoid undesired patterning effects.

An original and conceptually elegant approach consists in inducing a change of absorption and refractive index via an intersubband transition. This prevents any slow dynamics related to the generation/recombination process to interfere with the evolution of the distribution of free carriers in the conduction band, which is a much faster process. The concept has been implemented using a double quantum well structure with transition energy in the Telecom spectral range owing to Antimonide alloy structure grown on Indium Phosphide [17]. Truly picosecond nonlinear switching with picojoule pump energies have been achieved and an integrated switch capable of demultiplexing a data stream at 160 Gb/s has been demonstrated [3]. A very strong nonlinear response of Silicon nanocrystals has been reported recently [45].

An overview of the possible all-optical gating technology is shown in Fig. 5.3, with particular emphasis on resonant devices, namely ring resonators or photonic crystals. For comparison, non-resonant devices such as nonlinear waveguides and interferometers are also represented.

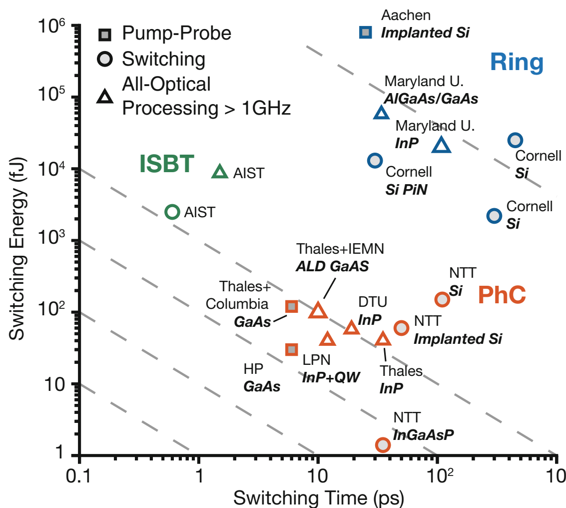


Fig. 5.3 All-optical processing based on semiconductor nonlinear devices, compared in terms of the speed (the switching time or recovery time) and the energy required for triggering the switching action. Symbols refer to different situations: pump-probe measurement of the response, direct observation of switching (or gating) and operation at high rate. Dashed line represent a constant *time* × *energy* product, which is approximately conserved for the same material and same cavity design

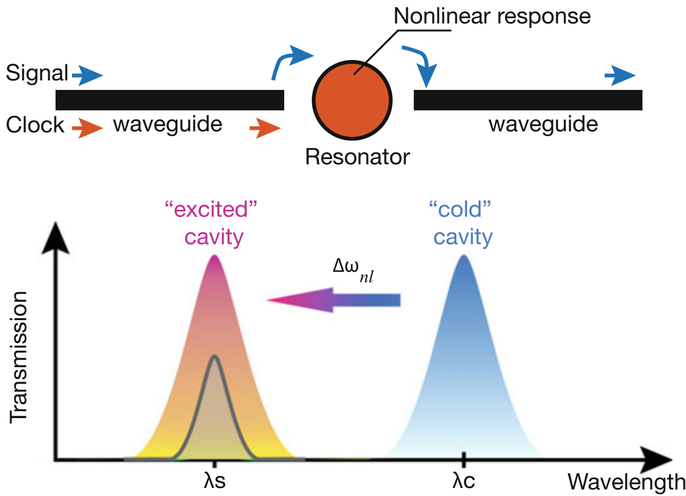


Fig. 5.4 Principle of the resonant all-optical switch. Signal is modulated by the dynamical control of the detuning

5.2.2 Energy-Efficient All-Optical Gates

The first attempt to build what at that time was described as the optical equivalent of a transistor dates back to the 70's. It was shown that an optical resonator (a Fabry-Pérot interferometer) filled with a nonlinear material (Na vapour excited near the *D* absorption lines) responds with a characteristic bistable input-output dependence [24]. III–V semiconductors have quickly been considered as more suitable materials for miniaturization and for improving speed and energy efficiency [23]. That leads to the concept of nonlinear semiconductor resonator as all-optical gates meant to operate at reasonably low optical power levels.

Considerations about the benefits of scaling down the volume of the resonator have motivated research on monolithic semiconductor resonators made of GaAs [33], where it was shown that recovery time and the pulse energy required for optical control decrease with the size of the resonator. Fast recovery time (30 ps) and strong all-optical modulation have been achieved with a pump pulse energy of about 600 fJ, in the smallest device.

Optical integrated circuits built around nonlinear resonators have been proposed later [77] to perform some all-optical processing experiments, particularly wavelength conversion, whereas a modulated signal at wavelength λ_1 is used to modulate a CW carrier at λ_2. As for semiconductor waveguides, the recent improvement of processing capabilities, enabling high-quality etching and, even more importantly, sub-micron patterning, has enabled a radically new class of optical microresonators, combining both small size and large quality factor.

In order to appreciate how this point is important in the context of all-optical gates, it is useful to consider two parameters characterizing their operation: the energy required for switching (E_{sw}) and the recovery time (τ_r). In the specific case of nonlinear resonators, with quality factor Q and mode volume V, and in the continuous wave limit, it is shown that the minimum power for observing bistable switching is $P_{min} \propto n_2 V Q^{-2}$ [67], when assuming instantaneous (e.g. Kerr) index change $\Delta n = n_2 I$. Based on scaling rules for n_2 with the energy gap of the semiconductor, it has been shown that the nonlinear response within a given spectral range can be maximized with a suitable choice of the semiconductor compound alloy [63, 68].

Assuming the choice of the most suitable material has been made, which also depends on other parameters such as manufacturability, cost, and other optical properties, the same relation suggests a connection between the device bandwidth, namely the recovery time, $\Delta\omega = 1/\tau_r \propto Q^{-1}$, and the operating power $P_{min} \approx E_{sw}/\tau_r$, *i.e.*:

$$E_{sw}\tau_r = const \times n_2 V \tag{5.1}$$

Therefore, increasing the Q factor will reduce the switching energy, but at the cost of reducing the speed too. Instead, the reduction of the modal volume could, in principle, improve energy efficiency without affecting speed. There is an interesting similarity with microelectronics, where the decrease of energy consumption *per transistor* has decreased drastically by mere geometrical scaling considerations.

The modal volume express the capacity of a resonator to store light within a certain volume, and it is related to the spatial distribution of the field $\vec{E}(\vec{r})$ of a particular normal mode. More precisely, V is the volume that an uniform field $E = \max(\vec{E}(\vec{r}))$ would need to store the same amount of energy as in the resonator, namely: $W = \frac{1}{2}\int \varepsilon_0\varepsilon_r|\vec{E}(\vec{r})|^2 d\vec{r} = V \times \frac{1}{2}\int \varepsilon_0\varepsilon_r \max(|\vec{E}|^2)$, as illustrated in Fig. 5.5.

Extremely large Q-factors are possible in dielectric resonators [75], while, as the equations above suggests, V is only loosely related to their physical footprint. The important point here is that if the index contrast is large enough, the optical mode is in general well confined inside the resonator and can approach the diffraction limit $(\frac{\lambda}{2n})^3$.

Plasmon resonators can beat this limit [52], but at the cost of large optical losses, because in this limit the magnetic energy of the resonant mode is vanishing and is replaced by the kinetic energy of electrons in the metal, which is dissipated within the short timescale of inelastic collisions (10–100 fs) [36], setting a limit to the Q factor which is too severe for exploiting the resonant enhancement.

Dielectric resonators made of silicon (or other high index materials) such as micro-rings or micro disks (Fig. 5.5), can achieve a much larger Q factor ($> 10^6$) while the smallest ring resonator [58] with radius = 1.5 μm features a mode with volume $V \approx 1\,\mu m^3$ which is way smaller than macroscopic resonators or other dielectric resonators with low refractive index contrast, but still nearly 2 orders of magnitude larger than the diffraction limit, namely $(\frac{\lambda}{2n})^2 = 0.015\,\mu mm^3$ in the telecom C band. The measured intrinsic Q factor is 2×10^4, about a factor two below the bending loss limit for this radius.

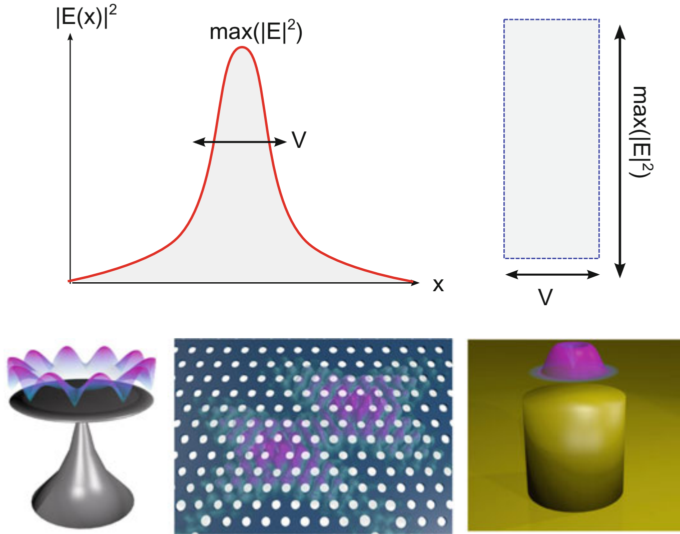


Fig. 5.5 Optical resonators. *Top* Concept of the modal volume: V is such that it contains the same energy as the resonator when filled with uniform field $E = \max E(\mathbf{x})$. *Bottom* representation of some categories of resonators, micro-disk, PhC and plasmonic resonator

Instead, Photonic Crystal (PhC) resonator can approach this limit much closer. The design developed by Zhang and Qiu [88] results into a mode volume of 0.03 μm^3, only twice as large the diffraction limit. Thus, semiconductor PhC resonators are extremely suited for maximizing speed and energy efficiency.

5.2.3 III–V Photonic Crystals Resonators

Photonic Crystals are optical structures whose properties are modulated periodically at the scale of the wavelength. They can viewed as a generalisation of the Bragg reflector, where the scattered waves add coherently to produce a reflection with strength growing with the number of periods of the reflector. Consequently, a perfectly ordered and infinite structure, periodically modulated along any direction, forbids the propagation of light within a given spectral range. The existence of a forbidden gap for photons implies fundamental properties in the control of the light-matter interaction [82]. In analogy to semiconductors and crystals, a “defect” in a otherwise perfectly periodic structure results into resonant localized states [34].

The fabrication of such structures has been attempted with *ad hoc* techniques (e.g. holography) but is still extremely challenging because of the tight tolerances required. In contrast, the planar semiconductor patterning and etching processes, derived from microelectronics, can meet such tolerances. Crucial milestones have

been the demonstration of a complete photonic band gap along the plane of such structures [39] and the first laser [56]. These two-dimensional structures needed to be optically isolated from the substrate, in order to avoid off-plane radiation losses [35]. This was achieved through the removal by chemical etching of the “sacrificial” layer beneath the patterned area, which leaves a self-standing membrane, as shown in Fig. 5.6.

The careful design of the “defect” is crucial in order to induce a strong optical confinement and a large Q-factor [2]. The continuous improvement of the design and of the processing has lead to the achievement of very large Q (up to 10^7) factors [61].

A different optimisation leads to the minimisation of the mode volume, which is more relevant for fast optical gates. A very suitable design [88] consists in displacing two holes by a fraction (0.16) of the lattice period, which is shown in Fig. 5.6. This produces a resonance close to the high frequency side of the forbidden photonic gap of the lattice (and it is therefore referred as a “donor” mode). The corresponding distribution of the optical field is extremely localized, mostly between the two displaced holes. The calculated mode volume is $0.24(\lambda/n)^3$, which is less than twice the diffraction limit.

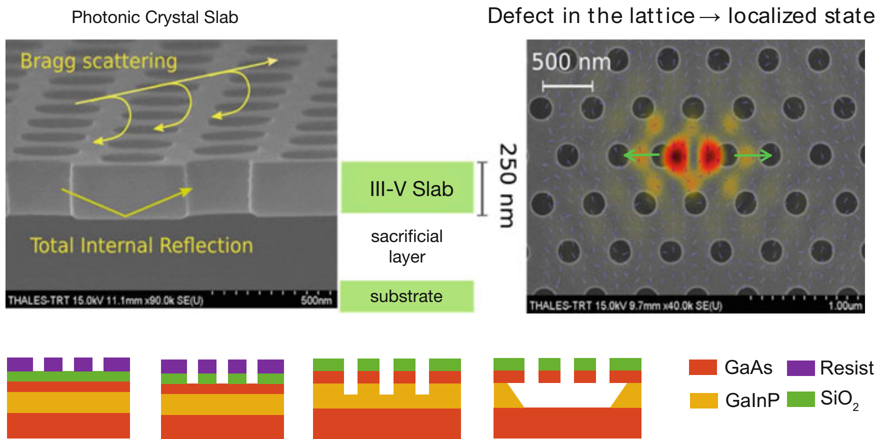


Fig. 5.6 Photonic crystal made of a semiconductor membrane patterned with a two-dimensional lattice of holes. *Left* in-plane confinement results from the Bragg scattering (the photonic band gap), out-of-plane confinement is ensured by total internal reflection. *Right* defect-mode resonator resulting from the outward displacement of two holes; intensity of the electric field is superimposed. The cavity design is based on [88]. Process flow, from *left* to *right* resist pattering by e-beam lithography, etching of the SiO_2 hard mask, etching of the III–V by an Inductively Coupled Plasma (ICP) and removal of the sacrificial layer. The fabrication process is detailed in [14]

5.3 Nonlinear Dynamics in PhC Resonators

Long ago, S. John suggested *the utilisation of localisation as a trigger mechanism for nonlinear or bistable response, as in the Kerr electro-optic effect* in photonic crystals [34]. The nonlinear dynamics of PhCs has indeed been investigated broadly [32, 66], highlighting phenomena such as non-adiabatic dynamics [83], wavelength conversion [54], self-pulsing [13, 41], to cite a few. The focus here will be on how such cavities can be designed to achieve energy-efficient and fast all-optical gating in a device which could be compatible with photonic integrated circuits. These considerations have strong implications in the choice of the materials and designs, which will be discussed hereafter.

5.3.1 Microwatt Nonlinear Response

Photonic crystals cavities are almost single mode resonators with mode size comparable to the wavelength. The density of the optical power can be extremely high compared with the level of the power at the input. In the coupled mode approximation [42], the energy in the cavity W is simply related to the input power P through the cavity photon lifetime $\tau_c = Q\omega_0^{-1}$, namely $W = P\tau_c$. Let us consider a cavity with a mode volume $V = 1.3(\lambda/n)^3$ and $Q \approx 10^6$ fed with $P = 1\,\mu\text{W}$ and $\lambda = 1.5\,\mu\text{m}$. While the stored energy at steady state is only 5 fJ, the energy density is as large as in a beam with $I = 1\,\text{GW/cm}^2$, which is enough to trigger the nonlinear absorption in semiconductors such as GaAs.

This situation is apparent in Fig. 5.7, where the resonance of a high-Q cavity (5×10^5) is broadened due to nonlinear absorption as the coupled input power is in the micro-Watt range [15]. The heat generated by the nonlinear absorption in the cavity can be estimated through its thermal resistance from the observation of the rise of the local temperature. In GaAs, the thermal resistance of these structures is typically about 10^5 K/W. This is a large value, as at steady state, or a times scales longer than the thermal time constant (in the μs range), a few microwatts of absorbed optical power induce an index change much larger than any other nonlinear effect (Kerr or due to free carriers). Thermal nonlinearity is therefore the dominant effect at long time scales. Bistable transitions are easily observed when the excitation is red detuned relative to the cavity resonance [80].

In the context of ultra-fast all-optical processing, however, the thermo-optic effect is too slow to be exploited, but still need to be considered carefully in order to achieve a stable operation point.

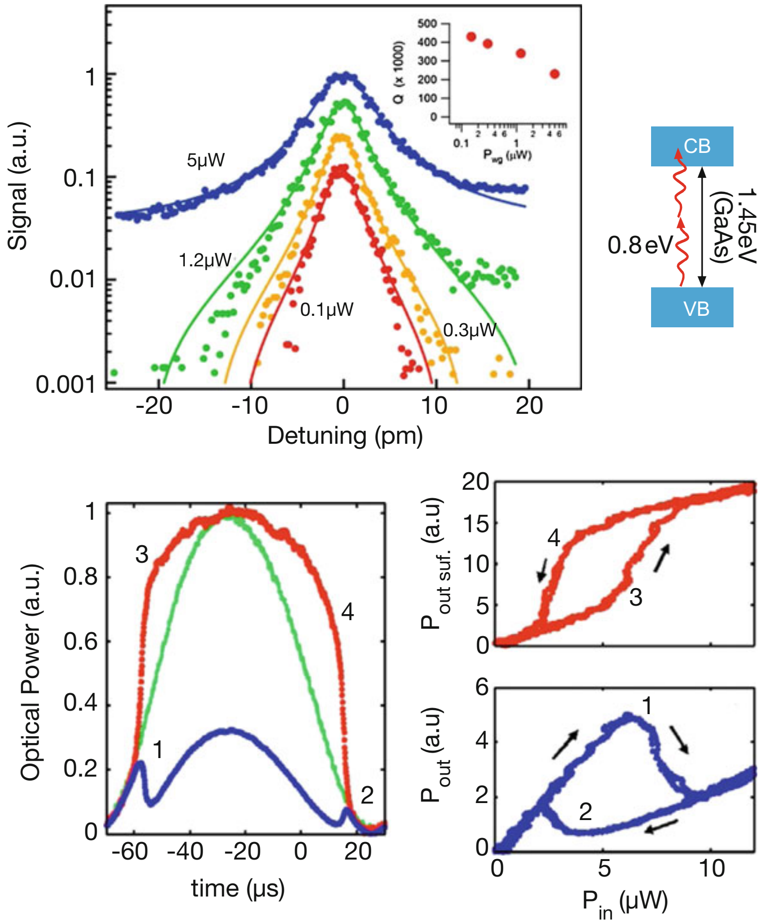


Fig. 5.7 Broadening of the resonance of a high-Q GaAs PhC cavity due to nonlinear absorption, from [15]; bistable response of a GaAs cavity due to thermally induced index change, from [80]

5.3.2 Fast Optical Nonlinearities in Semiconductors

In semiconductors, the contribution to the nonlinear index change due to the anharmonicity of the electron potential, i.e. the Kerr effect, is in general weaker than that resulting from the excitation of free carriers. This latter contribution consists of a variety of different mechanisms. The first of these is captured by the Drude model, describing a plasma of charged particles interacting with an electromagnetic field. As the density of free carriers increases, this results into a decrease of the refractive index at optical frequencies above the plasma resonance, which, in this case, is located in the THz spectral range or below. The second mechanism, “band filling”, is related to the saturation of the absorption as states in the conduction band are filled

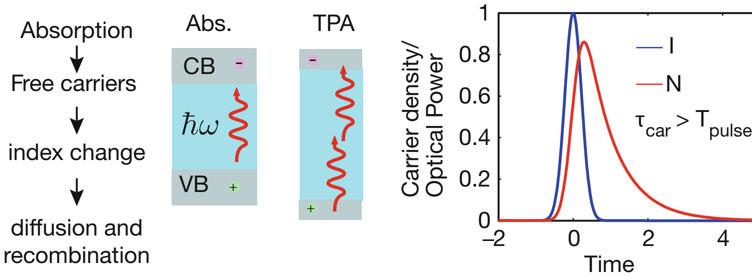


Fig. 5.8 The nonlinear response (normalized units) to optical excitation I in semiconductors is dominated by the free-carriers dependent refractive index, which is not instantaneous, as it follows the carrier density N

up. This also induces a decrease of the refractive index, described by the Kramers-Kronig equations. The third effect is explained as a many-body effect inducing a "shrinkage" of the electronic band gap, hence a change of the dispersion [9].

On the other hand, carrier dynamics is not instantaneous, as it is regulated by diffusion and recombination, with a rate τ_{carr} typically in the picosecond to nanosecond time scale (Fig. 5.8). The simplest model is based on the rate equation $\dot{N} = -N/\tau_{carr} + \frac{\alpha}{\hbar\omega}I$, with α the linear absorption. The change of refractive index $\Delta n = \frac{dn}{dN}$ is dispersive, and it is particularly large near the absorption band edge of the semiconductor [9].

In the limit where recombination is much faster than the excitation, $\tau_{carr} << T$, an effective Kerr nonlinear response can be defined as $n_2 = \frac{dn}{dN}\frac{\tau_{carr}}{\hbar\omega}\alpha$. For $GaInAsP$ with electronic gap $E_g = 0.84$ eV (1470 nm) the absorption at $\lambda = 1.55\,\mu$m is $\alpha = 300\,\mathrm{m}^{-1}$ and $\frac{dn}{dN} = -8\times10^{-20}\,\mathrm{cm}^{-3}$, thus $n_2 = -3\times10^{-15}\,\mathrm{m}^2\mathrm{W}^{-1}$, assuming $\tau_{carr} = 20$ ps. This value is more than 2 orders of magnitude larger than the electronic Kerr contribution. Based on a combination of optimized nonlinear response of the material and confinement in a ultra-small cavity, all-optical switching with femto-Joule control pulses has been demonstrated in PhC cavities [55].

Because of the large density of the optical field in the cavity, non-linear absorption is in general not negligible. In $GaInAsP$ (0.84 eV), the nonlinear absorption is comparable to direct absorption as the power intensity is $0.2\,\mathrm{GW\,cm}^{-2}$, which is easily reached in PhC cavities. Thus, semiconductor with relatively large gap, such as GaAs (1.42 eV), InP (1.35 eV) and Silicon, and, therefore, strictly transparent at the telecom wavelengths, can be used for nonlinear gates in this spectral range. The weaker index dependence on the carrier density implies however slightly larger power levels.

Nonetheless, there are two advantages in using binary material such as GaAs and InP: the much better thermal conductivity than ternaries, and the transparency of the material in the linear regime. Thus, the signal (probe) can be made weak enough not to induce nonlinear effects itself, and, furthermore, the same material can be used for in-plane propagation of signal and pump. Thus, it is not necessary to excite the cavity off-plane or to develop a complex 3D optical circuitry.

5.3.3 Nonlocal Nonlinear Response of PhC Cavities

The geometry of the optical resonators considered here has two important implications on the dynamics of free carriers, hence of the nonlinear response. First, the average distance of the photo-generated carriers from the surface L_s is very short (typically 100 to 200 nm) compared to the diffusion length. Thus, surface recombination S tends to dominate any other recombination process. In the case of *GaAs*, S might approach the saturation velocity in absence of any passivation treatment [53]. Considering a patterned surface such as in PhC crystal structures, an effective carrier lifetime is readily derived within the Shockley [64] approximation: $\tau_{eff}^{-1} = \tau^{-1} + S/L_s$, which can be shorter than 1 ps. Such a fast dynamics was first observed in GaAs Photonic Crystal structures [11].

The second implication of the nanoscale geometry appears when the photogeneration is localized within a range smaller than the diffusion length, or the distance to the surface. Then, a new time constant appears in the evolution of the carrier density, which is related to the migration of carriers away from the area where the optical cavity mode is localized. This becomes apparent when the recombination process is less important, such as in material with weaker surface recombination, such as InP or Silicon. In order to appreciate this point, Fig. 5.9 shows the decay of the carrier density in the area where the optical field is localized, which we refer to as *effective density* N_{eff}. The very early stage is characterized by a time constant which is almost two orders of magnitude faster than the effective carrier lifetime (about 500 ps in InP PhCs). The origin of this behaviour is apparent when observing the spatial distribution of photogenerated carriers, initially localized within about half a micron, and expanding very quickly in the first picoseconds. This effect was first observed in silicon PhCs [72] and later in other semiconductors, namely InP [29, 87] and InGaAsP [55].

When the surface recombination is large, then it can compete with diffusion. Figure 5.10 compares the decay of the effective density N_{eff} with the averaged density, i.e. the total carrier population, as the surface recombination is increased. While the total carrier population decays exponentially with a time constant clearly related to the recombination, the effective density is also influenced by the diffusion, which is increasingly apparent as S decreases.

It has been realised that, in nanoscale semiconductor resonators, the exponential decay described by a time constant is not well adapted to describe the nonlocal response of carriers. Very recently, a model based on the Green's function has been introduced to provide a more accurate description of the diffusion process [48].

5.4 PhC All-Optical Gate

The concept of a resonant all-optical gate, as illustrated in Fig. 5.4, has been first implemented in a photonic crystal structures on a Silicon on Oxide platform [71],

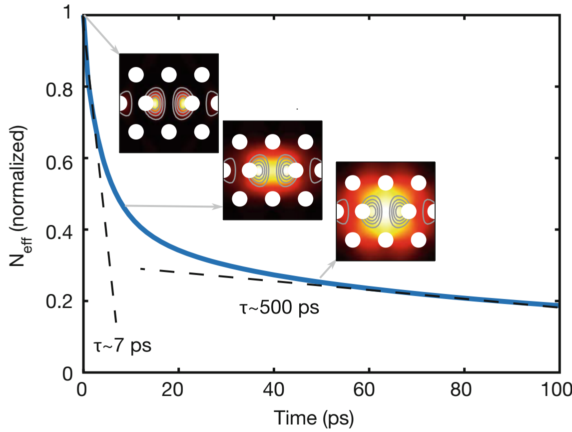


Fig. 5.9 InP PhC cavity. Calculated effective density of carriers N_{eff}, generated by TPA in a PhC cavity. Snapshots of the spatial distribution are shown at the excitation, after fast diffusion (≈ 7 ps) and after 50 ps. The lattice period 450 nm gives the spatial scale. Gray contour lines denote the initial carrier distribution, added for comparison. Dashed straight lines denote fitted fast and a slow time-constants

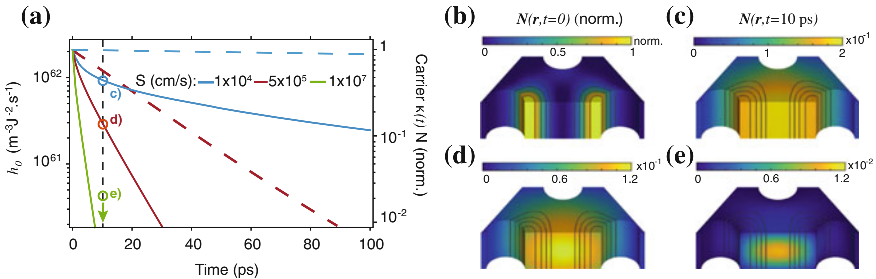


Fig. 5.10 H_0 cavity: calculated impulse response h_0 (*solid*) and the spatially averaged and normalized carrier density $\kappa(t) = overlineN/max(\overline{N})$ (*dashed*), depending on the surface recombination velocity S. Initial ($t = t_{0^+}$) spatial distribution N of the carriers (**b**) and at $t = t_{0^+} + 10$ ps with (**c**) $S = 10^4$ cm s^{-1}, **d** $S = 5 \times 10^5$ cm s^{-1} and **e** $S = 1 \times 10^7$ cm s^{-1}. Adapted from [48]

shortly following the demonstration of a nonlinear ring resonator on the same material [4]. Arguably, this opened the perspective of making these nonlinear devices accessible through mass-production, and attractive because they can coexist with the wide palette of photonic functions already available on this platform.

Still, the nonlinear absorption coefficient, the carrier mobility and lifetime in Silicon are not the most favourable for fast all-optical processing, compared to other materials, particularly III–V semiconductors. Faster and more energy-efficient devices are possible by selecting materials with faster carrier lifetime and stronger nonlinear response. The choice of the appropriate cavity design, minimizing the modal volume is also crucial, as discussed in Sect. 5.2.2.

Figure 5.11a represents a PhC all-optical gate obtained by evanescently coupling a H_0 resonator to two PhC waveguides. The H_0 cavity design, described in detail in Fig. 5.6, provides the smallest modal volume for a dielectric resonator. Moreover, the choice of *GaAs* allowed to exploit its large surface recombination and one order of magnitude larger nonlinear absorption compared to Silicon, translating into much faster and more energy-efficient operation [30].

The quality factor of the resonator is estimated from the transmission spectrum (Fig. 5.11b). The coupling loss into the waveguide, here ∼−5 dB/facet, is reasonably small owing to a mode-adapter which is detailed in [74]. The linear transmission is estimated to be 90% with the cavity on-resonance, which corresponds to the the ratio between the loaded $Q_L = \lambda_0/\Delta\lambda = 1200$ and the intrinsic $Q_i = 2.5 \times 10^4$. The cavity photon lifetime $\tau_{ph} = Q_L\lambda_0/2\pi c \approx 1$ps is short enough not to limit the switching speed.

Homodyne wavelength-degenerate pump-probe measurements reveal all-optical modulation, with an increase or a decrease of the signal, depending on the detuning (Fig. 5.11c, d). In both cases, a decay constant close to 6 ps is observed, suggesting a much faster carrier dynamics than in other semiconductor, e.g. Silicon [71], no matters if it is implanted with ions to decrease the carriers lifetime [70]. Recent measurements on GaAs PhC cavities designed to operate in the near infrared spectral range have also concluded a fast carrier lifetime [10]. It is likely that τ_{carr} might however vary substantially in GaAs nanoscale structures, likely in the $1 - 10$ ps range, depending on the detail of the etching process used to pattern the PhC structure and,

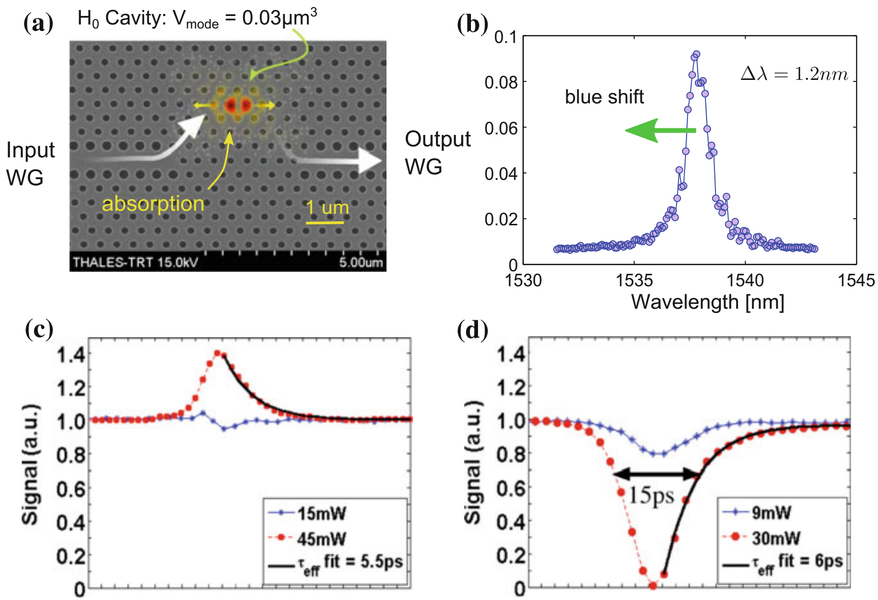


Fig. 5.11 GaAs-based PhC resonant all-optical device, following [30]

more generally, on the properties of the surface. This point will be discussed further in Sect. 5.4.5.

Another remarkable aspect of this GaAs structure is the low power level inducing the all-optical modulation, which is below 50 mW peak, hence about 200 fJ for a 4 ps long pulse. This is the result of the extremely tight optical confinement and the large nonlinear absorption. How this compares with other technologies and materials is represented in Fig. 5.3.

5.4.1 Photon Molecule

The practical use of nonlinear gates requires that the optical control can be distinguished from the signal. In principle, the symmetries of a doubly degenerate resonance could be exploited to avoid unwanted crosstalk. Nevertheless, the spectral separation of the signal and the control is much easier to implement, as this requires a doubly resonant cavity and filters. The so called L_3 and L_5 cavities (3 and 5 missing holes in line) contain several modes and have been used for frequency non-degenerate all-optical switching [71].

A crucial point here is that the strength of a non-degenerate nonlinear interaction depends on the spatial overlap of the electric fields. This is strictly true for the Kerr interaction (and Four Wave Mixing). It could be speculated that this conditions could be relaxed, since the free-carriers induced index change, which is exploited here as nonlinear effect, is non-local, as discussed in Sect. 5.3.3. In practice, and particularly if the effective carrier lifetime is short, achieving a good field overlap, e.g. spatial matching of the two resonances, remains critical for maximizing the nonlinear response.

The above constraints lead to the concept of "photon molecule", where coupling two single-mode cavities results into a doublet of resonances [27, 31, 73], is used here to obtain two modes maximizing their overlap. The photon molecule is implemented here using two H_0 cavities separated by 4 lattice periods along the M direction [16, 50], as shown in Fig. 5.12a. The symmetric and anti-symmetric combination of the H_0 cavity mode are associated to the two resonances (Fig. 5.12b, c). The amplitude $|E|$ of these two super-modes overlaps almost perfectly (Fig. 5.12d), which is what matters here. As a consequence, the excitation of either modes results into a close to identical spectral shift of both of them (Fig. 5.12b), thereby enabling efficient non-degenerate switching, i.e. with spectrally separated signal and control optical channels.

5.4.2 The Role of the Carrier Lifetime

In a semiconductor all-optical device based on the excitation of free-carriers, the density of the effective carrier population N_{eff} in the cavity, to which the change

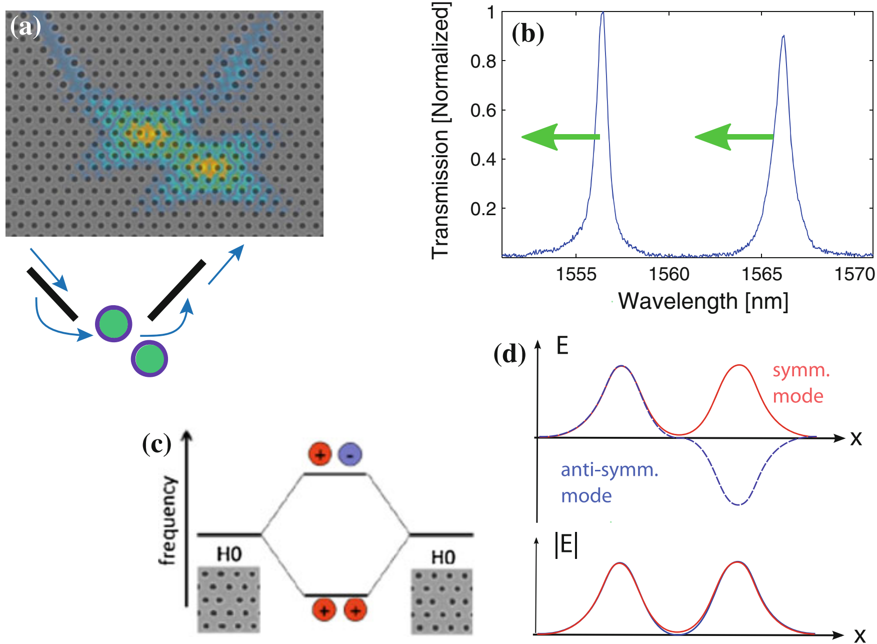


Fig. 5.12 **a** SEM image of a Photon Molecule and, superimposed, the electric field energy density at resonance; **b** experimental transmission spectrum, arrows indicate the effect of the excitation of free-carriers; **c-d** cocept of a Photon Molecule with associated odd and even modes

in the refractive index Δn is proportional, builds up depending on the simple rate equation $\partial_t N_{eff} = -N_{eff}/\tau_{carr} + G(t)$. Therefore, if the carrier lifetime τ_{carr} is shorter than the duration of the excitation $G(t)$, carriers cannot accumulate, and the response is weak. Conversely, if the carrier lifetime is long, the recovery time of the all-optical device increases accordingly. Considering this, the optimized carrier lifetime should be ideally close to the duration of the excitation. This is illustrated in Fig. 5.13, where free carriers are excited into a nonlinear resonator by a pump pulse with duration $t_{pulse} = 4.5$ ps, which is a sensible choice for processing data at a fast rate (e.g. time demultiplexing a 100 Gb/s data stream) and, therefore, suggests that carrier lifetime should be of the same order.

In the following, we will survey recent experiments where a variety of materials have been used to build a PhC all-optical gate. The recovery dynamics of a nonlinear photonic molecule is measured through a time resolved spectral transmission experiment (Fig. 5.14). The doublet of resonances allows to excite the cavity with a pump tuned to one resonance, while a spectrally broad pulse is used as a “white light” source to probe the instantaneous transmission spectrum of the other resonance. Both pump and probe are obtained from the same mode-locked laser (MLL) by spectral filtering.

The dynamical response of these nonlinear cavities is modelled by a combination of time-dependent coupled mode theory [42] and the dynamical equations

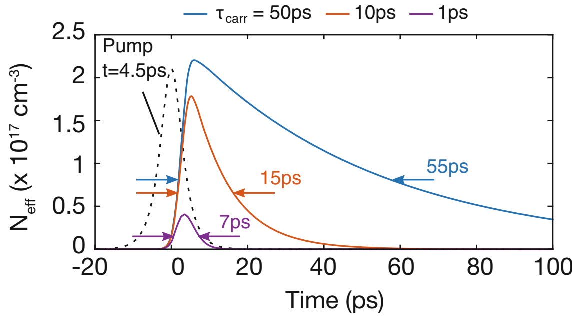


Fig. 5.13 Calculated temporal decay of the effective carrier concentration, depending on different carrier lifetime: τ_{carr} = 1, 10, and 50 ps. From [50]

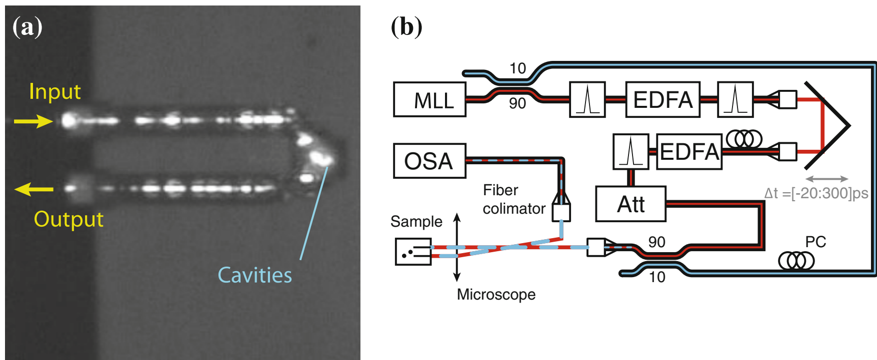


Fig. 5.14 **a** Infra-red image (*top view*) of the PhC all-optical gate showing the access waveguides and the cavities from the top. On resonance, light propagates through the photon molecule and returns to the end-facet. **b** Spectrally and temporally resolved pump-probe experiment. Pump path (*red*), probe (*blue*). EDFA: Erbium Doped Fibre Amplifier, PC: Polarization Controller, Att: Attenuator, OSA: Optical Spectrum Analyser. Experimental setup from [49]

describing the evolution of the population of free carriers inside the resonator. Here, the diffusion of the photo-excited carriers is taken into account rigorously, within the ambipolar approximation, using the formalism of the Green's function. The model is able to reproduce the experimentally measured response accurately and with a minimal set of fitting parameters, namely the surface recombination velocity [48, 50].

5.4.3 InP

Nonlinear photonic crystal resonators, all based on the H_0 design and made of InP, have been introduced recently [16, 29, 78, 86] and all the measurements of the

dynamical response have evidenced a fast and a slow time constant. A similar behaviour, but at a much lower pump excitation power, has also been reported in InGaAsP resonators [55]. The observed fast time constant, describing the decay of the transmitted probe, has been reported to be as low as 10 ps. Interestingly, this does not correspond to the relaxation of the carrier population, which is much longer. Rather, it results from the much faster diffusion process. Moreover, the transmission also depends on the detuning of the probe relative to the cavity, which can be optimized to obtain a short gating function. This is apparent in Fig. 5.15, showing the measured decay of the excited carriers, which is directly related to the instantaneous detuning of the resonator. Our model is in very good agreement with the experiment when assuming surface recombination velocity $S = 10^3$ m/s which the only free parameter here (all the others are either calculated, or measured independently). Put in other words, the surface recombination can be deduced from pump-probe measurements [48].

It has been shown that the dynamical response can be further improved by modifying the lineshape and the symmetry of the resonator [85]. Still, the underlying dynamics is much slower than the observed response. This implies that free carriers will continue to accumulate if the repetition rate of the pulsed excitation is faster than the time required for a complete recovery of the carrier population.

5.4.4 P-Doped InP

In an undoped, unbiased semiconductor, the decay of the effective density of the photo-excited carriers is governed by the ambipolar diffusion constant. In III–V materials such as GaAs, the process is dominated by the heavy holes. A much faster diffusion is achieved owing to p-doping, such that only the dynamics of the minority

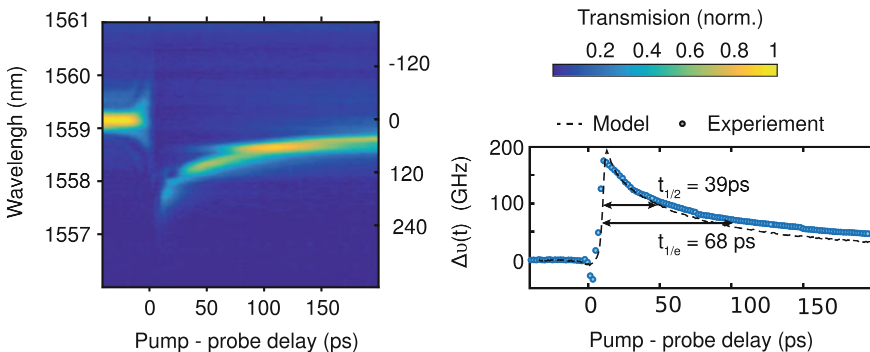


Fig. 5.15 Pump-probe measurement of the InP gate described in [16]. *Left* spectrally and temporally resolved transmission map; *right* extracted instantaneous spectral shift (*circles*) and model (*dashed line*) with $S = 10^3$ m/s

carriers, the highly mobile electrons, matters. This is a well known technique to accelerate the carrier dynamics and was, for instance, already used in uni-traveling carrier photodiodes [69].

This idea has been exploited to accelerate the response of an InP photonic molecule, which is doped with acceptors to the level of about 10^{17} cm^{-3}, such that free-carrier losses do not affect the Q factor ($\alpha \simeq 1$ cm^{-1}). The response is shown in Fig. 5.16, which clearly reveals a faster recovery, compared to the identical yet undoped structure in Fig. 5.15. More precisely, from the instantaneous spectral shift of the cavity it is deduced that the density of the excited carriers has halved within only $\tau_{1/2} = 25$ ps, almost twice as fast as the undoped case.

As the optical excitation becomes stronger and the level of injected carriers becomes larger than the doping level, the ambipolar diffusion regime is eventually restored and the fast time constant increases again, as apparent in Fig. 5.17. A model

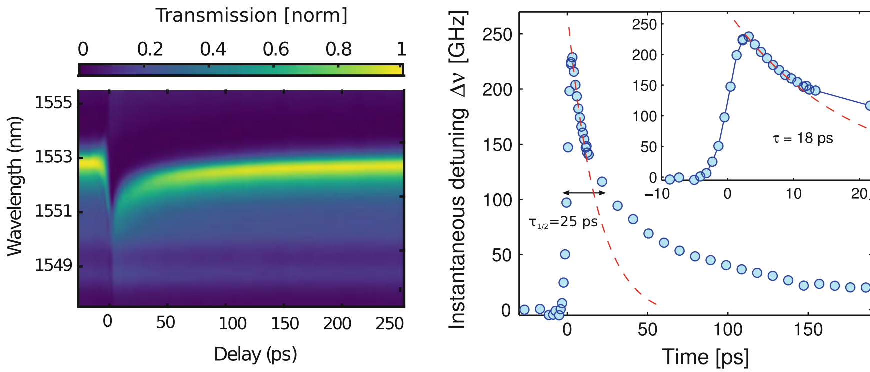


Fig. 5.16 **a** Time-resolved experimental transmission spectra centred around the low-energy resonance **b** Extracted time-resolved instantaneous detuning. The double arrow indicates the time for halving the detuning $\tau_{1/2}$. The inset is an enlarged view with exponential fit (dashed line) and corresponding time constant τ. From [49]

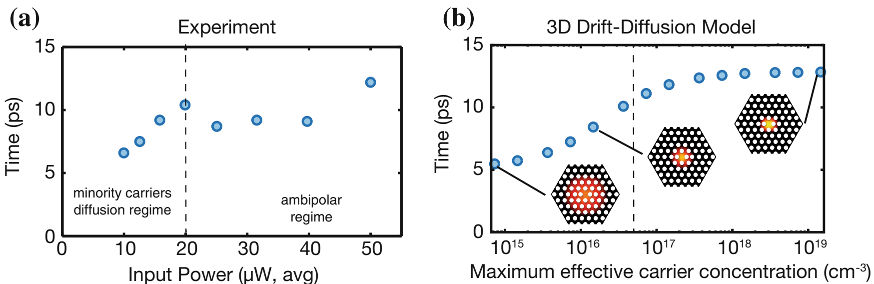


Fig. 5.17 **a** Measured fast time constant as a function of the input pump power coupled to the cavities. **b** Calculated time constant versus the maximum of the effective carrier density. The dotted line represents the level of p-doping. The distribution of the electrons at $t = 10$ ps is represented in the insets for different excitation levels. From [49]

accounting for the drift-diffusion process in the geometry of the PhC cavity reproduces this behaviour fairly well. P-doping is shown to be effective in enhancing the fast diffusion dynamics in PhC microresonators, without increasing the switching energy (still in the 100 fJ range) [49].

5.4.5 Passivated GaAs

As discussed in Sect. 5.4, the optical nonlinear response related to the excitation of free carriers is particularly fast in Gallium Arsenide, because the strong surface recombination induces a fast recovery by quickly removing carriers. However, it turns out that this effect is so strong in photonic crystal devices that it prevents the build-up of the free carrier population, thereby dwarfing nonlinear response. Another issue which has been observed in GaAs PhC is the degradation of the optical properties of the resonator, due to photo-induced oxidation [40].

The passivation of semiconductor surfaces exposed to large optical power densities, such as at the end facets of high-power laser diodes is well known and it is dealt with conventional deposition techniques such as PECVD (plasma-assisted chemical vapour deposition). These known passivation techniques have been used to improve the quality of the PhC surfaces. The trouble here is the complex geometry of PhC membranes, requiring a conformal coating of the surface, including that facing the semiconductor substrate, which is hardly accessible. Nevertheless, a new technique, known as Atomic Layer Deposition (ALD) has been developed, mainly in the context of high-speed electronics based on III–V semiconductors [6, 84] and has been applied to PhC very recently [38]. However, the impact on the surface recombination of ALD, depending on the specific treatment of the surface and the material deposited have been evidenced later [50]. Figure 5.18 shows a 30 nm thick layer of Al_2O_3 covering all the surfaces of a GaAs PhC uniformly.

Spectrally and temporally resolved pump-probe experiments have been carried out to compare the relaxation dynamics before and after passivation with Al_2O_3. The main result is summarized in Fig. 5.19a, b, demonstrating that the specific surface passivation treatment used here results into an increase of the instantaneous detuning from about 30 GHz to 140 GHz. Furthermore, the non-instantaneous recovery of the detuning is now apparent, which is consistent with an increase of the carrier lifetime. Importantly, the two measurements (passivated and unpassivated) are both reproduced accurately by the model [48]. The change of a single parameter, the surface recombination velocity, accounts for the different response. We conclude that a fairly large value, $S = 10^5$ ms^{-1}, about the saturation velocity in GaAs, is characteristic of the uncoated sample, while ALD reduces drastically the recombination to 3×10^3 ms^{-1}, i.e. by nearly two orders of magnitude. The modelled decay of the carrier density in the cavity (Fig. 5.19c) is insightful and reveals the drastic change of dynamics induced by the ALD coating.

Spectrally non-degenerate homodyne pump-probe measurements are used to estimate the response of the switch. When the pump and the probe are set with optimal

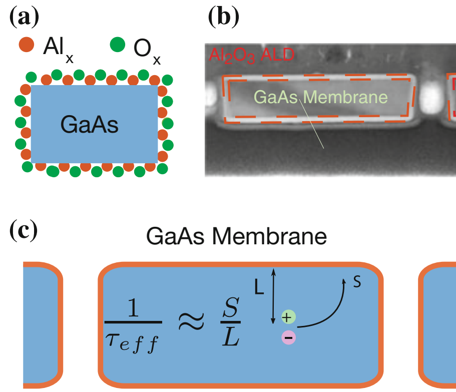


Fig. 5.18 **a** Atomic Layer Deposition process **b** Cross-sectional transmission electron micrograph (TEM) of the PhC slab revealing a conformal Al_2O_3 coating [50] and **c** schematics of the surface recombination process

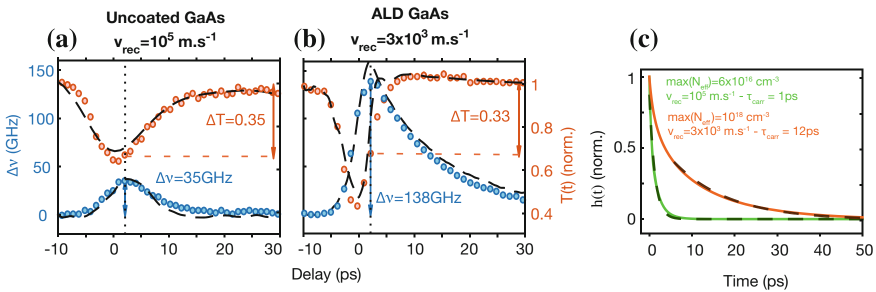


Fig. 5.19 Instantaneous cavity frequency shift $\Delta\nu(t) = \nu_c(t) - \nu_c(-\infty)$ (*left scale*) and time-resolved transmission $T(t)$ (*right scale*), for respectively the uncoated (**a**) and coated (**b**) samples. Note that in the case of the uncoated GaAs (a) the response is almost instantaneous, only limited by the pulse duration (4 ps) and the cavity lifetime. In both cases, the coupled pump energy is estimated to 300 fJ. The calculated response is superimposed (*solid dashed line*) **c** Calculated normalized effective carrier density $h(t)$ for both reference (*green*) and ALD (*orange*) samples in order to extracted the carrier lifetimes τ. From [50]

detuning relative to the resonances (Fig. 5.20), a switching contrast of 7 dB is achieved with a maximum of 200 fJ per pulse. If the pulse energy is decreased to about 100 fJ, the contrast is sill 3 dB, enough to perform some all-optical processing operations corresponding to a power consumption of 1 mW for 10 giga operations/s, which can be considered as energy-efficient. It is remarkable that the transmission has recovered completely in about 20 ps, the $1/e$ decay being about 12 ps, which is extremely fast for a semiconductor device based on free carriers.

A stronger evidence of a fast and complete recovery is the capability to respond to a sequence of closely spaced pulses. In fact, this implies that any possible "internal" dynamics has relaxed. Figure 5.21 reports the excitation of the AOG with two pump pulses with spacing decreasing from 40 ps to 10 ps. The probe reveals that the

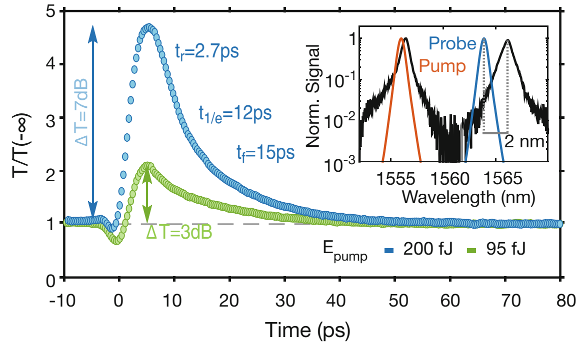


Fig. 5.20 Time-resolved homodyne non-degenerate pump-probe measurements at probe detuning $\Delta\lambda = -2.2$ nm. Relative probe transmission $T/T(-\infty)$ versus pump-probe delay and extracted rise-time t_r, fall time t_f and width at $1/e$. From [50]

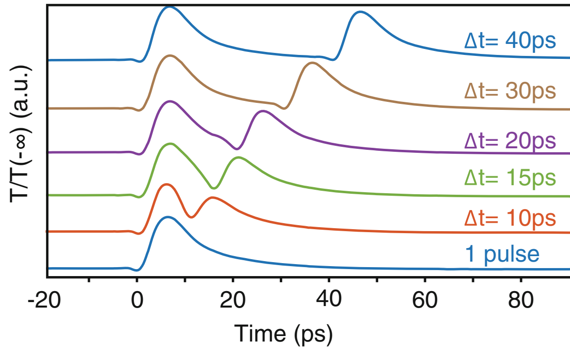


Fig. 5.21 Same as Fig. 5.20 but with a sequence of two pump pulses with separation Δt varying from 10 ps to 40 ps. From [50]

response to the following pulse is identical, hence not influenced by any transient dynamics due to the first pulse, until these are very close.

5.4.6 Integration with Silicon Photonics

In the context of all-optical switching based on the excitation of free carriers, compound III–V semiconductors can offer a larger electron mobility, hence a potentially faster response, than Silicon. Moreover, their direct electronic band gap can be adjusted by a suitable choice of the compound, which here is exploited to maximize the nonlinear response. Still, Silicon Photonics provides an industrial-grade platform for integrated optical circuits with a broad range of functionalities already available and the capacity of addressing complicated architectures.

The hybrid integration of III–V devices on a silicon photonic chip address this point as it allows to combine the advantages of the Silicon platform with the unique properties of III–V. The most mature and accessible approach to hybrid integration relies on the adhesive wafer bonding [28].

The integration of nanopatterned structures such as PhCs is however challenging because of the extremely accurate alignment required to control the coupling to the

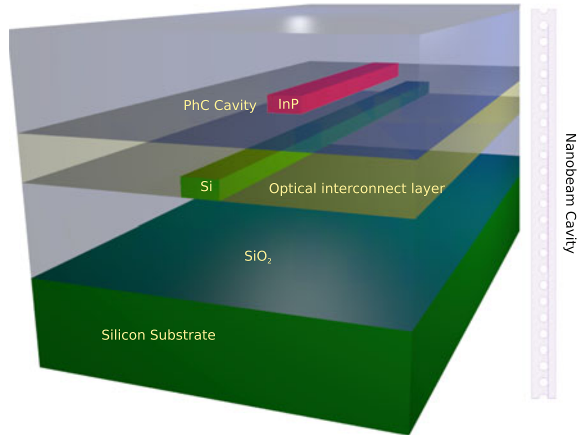


Fig. 5.22 Nanobeam cavity made of InP heterogeneously integrated on a Silicon Photonic Chip, [7]

underlying silicon waveguide (Fig. 5.22). Nonetheless, a “nanobeam” PhC resonator made of InP has been fabricated on top of a silicon photonic circuit [25]. The III–V structure is entirely encapsulated in Silica, which provides an additional advantage in terms of better thermal sinking, particularly important for laser sources [8].

By adding a InGaAsP thin layer at both surfaces of the InP membrane, surface recombination is greatly increased, which translates into a much faster recovery of the nonlinear response. By exploiting this property, a fast optical gate with a recovery time of about 30 ps has been demonstrated. Furthermore, all-optical wavelength conversion and signal recovery have been demonstrated at a rate as high a 10 Gb/s [7]. This landmark experiment demonstrates the feasibility of complex photonic circuits including sources, amplifiers and all-optical gates.

5.5 Application Example: All-Optical Signal Sampling

The “digital world” surrounding us is actually physically supported by analog signals, as well as any other signal associated to sounds, images, or any kind of physical measurement. The transition from the analog to the digital domain for processing, requires analog-to-digital converters (ADC). An A/D-converter produces a digital representation, a sample, of an analog signal. The two main characteristics of ADCs are the sampling rate and the accuracy of the sample expressed as the number of representative bits (ENOB). The need for fast ADCs with a large dynamic range (equivalently large ENOB) originates from some application domains such as medical imaging, broadband communications and radar [22]. Despite the steady progress of electronics, the current ADCs are struggling to meet the required performances.

While the thermal noise has been identified as the dominant limiting factor for high-resolution ADCs operating at moderate sampling rates (e.g. audio), the performances of high-speed ADCs are limited by both the comparator ambiguity and by

the timing error (jitter) in opening the gate. Concerning the latter issue, it has been shown that timing jitter in state-of-the-art electronic clocks is hardly better than 100 fs. This affects the accuracy of the representation, explaining why the ENOB has not increased much in fast ADCs [5]. For instance, Fujitsu has demonstrated a 65 GSa/s ADC providing 8-Bit resolution [1].

In contrast to electronic clocks, mode locked lasers (MLLs) can generate a train of optical pulses spaced by extremely regular time intervals, and therefore used as a high-quality optical clock. The lowest time jitter reported is below 1 fs [37], i.e. orders of magnitude better than electronic clocks. Not surprisingly, this has prompted a radically new approach to ADCs involving the use of an optical clock signal [5]. We note that if the signal to be sampled is also in the optical domain, an AOG is required.

In fact, the combination of a MLL and a electro-optic modulator is likely to be the simplest approach for converting an analog microwave signal into a train of samples. This architecture also allows a substantial degree of parallelisation owing to photonic integration [5]. If, however, the microwave signal is provided in the optical domain, which is the common situation of microwave photonics [60], then one might speculate about the possibility of performing sampling all-optically, by mixing the optical clock and the signal.

Another important performance factor in ADCs is the power consumption, in particular regarding high-speed applications in embarked devices, where the power dissipation and the operating time of the battery are critical. In fact, it has been shown that the power consumption of an electronic ADCs scales approximately quadratically with the sampling frequency (see Fig. 5.23), and therefore this becomes a concern when large bandwidth is needed. Moreover, power consumption also depends critically on the number of bits. For instance, ADCs with sampling frequency equal to 20 GHz consume 1.2 W with a 6 bit resolution and 10 W with 8 bits [57, 59]. Thus, the power consumption of an AOG operating at high-speed and with a large dynamics (switching contrast) is extremely important. A less critical of power consumption on the speed is therefore a crucial and achievable goal in optical signal processing.

5.5.1 All-Optical Sampling

An example of all-optical sampling involving a PhC AOG is shown here. A RF sinusoidal signal with frequency $f_{MZ} = 2.777$ GHz is transposed in the optical domain using a Mach-Zehnder modulator (MZM) and then it is sampled by PhC AOG described in Sect. 5.4.1, made of Indium Phosphide. The device is activated by train of pulses generated by a Mode Locked Laser (MLL) serving as optical clock with frequency equal to $f_{clock} = 2.525$ GHz. The wavelengths of the two optical inputs are adjusted such as the clock is set on the resonance at higher frequency whereas the optical carrier is blue detuned by about 1 nm from the low frequency resonance. The sampled signal is collected at the output, separated from the clock by spectral

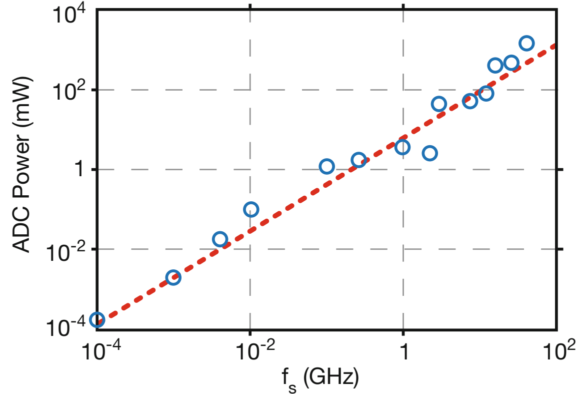


Fig. 5.23 Power consumption of an Analog to Digital Converter depending on the sampling rate (image adapted from [20])

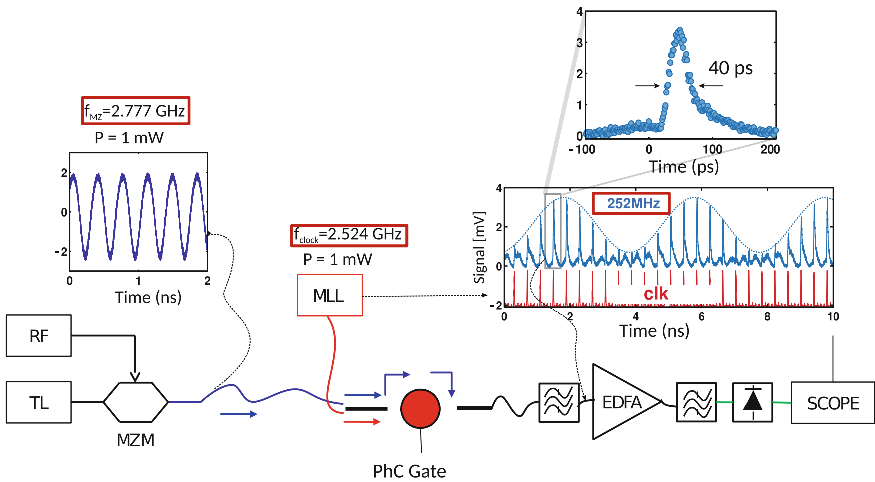


Fig. 5.24 Experiment of all-optical sampling involving a PhC AOG. *Inputs* tunable laser (TL) modulated by a Mach-Zehnder Modulator (MZM) shown in the inset, and Mode Locked Laser (MLL). Output is filtered and amplified by an Erbium Doped Fiber Amplifier (EDFA), detected and imaged at the oscilloscope (insets)

filtering and amplified by an optical low-noise amplifier (EDFA) to raise the power level before the InGaAs photodetector. The sampled signal is then displayed together with the clock and the input signal by an oscilloscope synchronised with the MLL (Fig. 5.24).

In this experiment, the sampling frequency is smaller than the signal frequency, thus the sampled signal represents an aliased version of the original signal, which is a sinusoid with frequency $f = f_{MZ} - nf_{clock} = 252$ MHz, as the original signal is located in the second Nyquist band ($n = 1$). Considering a more general signal centred at some frequency f_0 with spectral width smaller than $f_{clock}/2$, then the sampling generates a representation of the original signal in the base (low frequency)

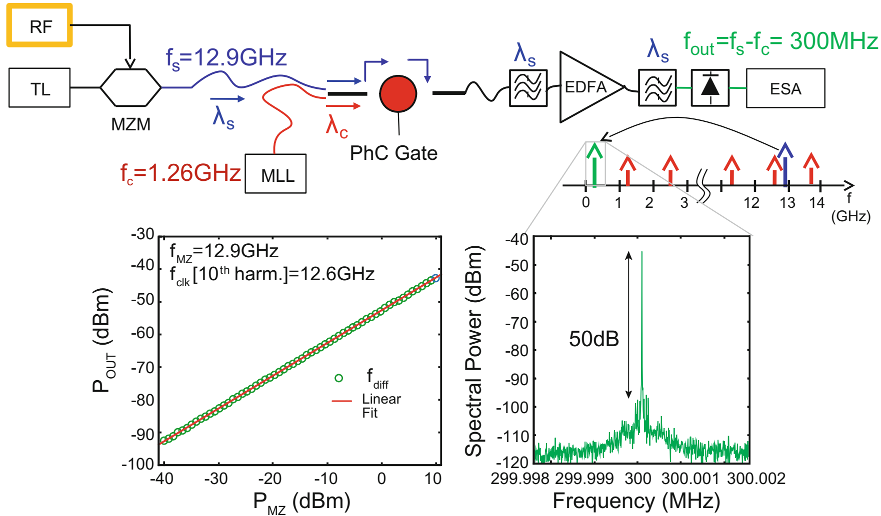


Fig. 5.25 All-optical sampling involving a PhC AOG, observed in the electric spectral domain. Experimental set up (*top*), as in Fig. 5.24, with an electrical spectrum analyser (ESA) replacing the oscilloscope. The harmonics of the clock are associated to the Nyquist's bands. Signals with frequency larger than the sampling frequency are translated into the first Nyquist band. Electric spectral power of the sampled signal as a function of the driving RF power applied to the MZM (*bottom left*) and electric spectrum of the sampled signal (*bottom right*)

band which is "exact" in the sense of the Shannon criterion. This has an extremely important implication in the domain of wide band radio receivers and radars, as there sampling also implements the frequency down-conversion.

This is clearly seen in another measurement. Here, a RF signal at a much higher frequency (12.9 GHz) is sampled with a clock at 1.26 GHz. The all-optical sampling generates a replica in the first Nyquist band at 300 MHz, corresponding to the mixing with the 10th harmonic of the clock of the signal. The RF spectrum is a narrow lineshape with noise being 50 dB below (Fig. 5.25). Another interesting point is the relationship between the input and sampled signal amplitudes. This is remarkably linear when the input RF power is spanned over 50 dB.

This linear dependence is not obvious, given the nonlinear character of the AOG. Nonetheless, an intuitive explanation for this is in the mechanism involved in the sampling operation i.e. the two-photon absorption, as explained in Sect. 5.3. In spite the fact that the average power of the two incoming optical signals is similar (about 1 mW), the peak power of the clock is much larger than the signal that only the former triggers the nonlinear response trough the two-photon absorption.

The most attractive aspect of this approach to photonic ADC is that the implementation of a series to parallel scheme is conceptually easy. There, a high-speed signal is sampled by time-interleaved ADCs operating in parallel and feeding lower speed

electronic digitizers. Besides ADCs, AOGs can be used for wavelength conversion, pulse analysis [79], signal regeneration and optical time demultiplexing.

5.6 Conclusions

Historically, nonlinear optical effects have been evidenced using large optical power levels, which have been made possible by the invention of the laser. This is why energy-efficiency and nonlinear optics are regarded to be at odds. The main point of this chapter is to bring evidences that this is no longer true. New material processing techniques have enabled the creation of a variety of dielectric optical resonators capable of nearly diffraction-limited optical confinement. Consequently, electric fields large enough to trigger nonlinear effects are achieved in tiny volumes by exciting the resonator with optical power levels in the mW or even the μW range, namely many orders of magnitude weaker than in more common nonlinear optic experiments.

The kind of resonators considered here are made of III–V semiconductor compounds and the optical confinement is based on the Bragg scattering, and are named Photonic Crystals. The nonlinear mechanism exploited here is the excitation of free carriers due to the nonlinear absorption, which reach a large density in the resonator. This produces a change in the refractive index which can be exploited to detune the resonance dynamically and, hence, to create an all-optical gate.

Aiming both at a fast response and energy-efficiency implies a trade off, as, in general, faster effects are weaker, e.g. free-carrier versus thermo-optic index change. Nonetheless, extremely fast nonlinear devices which could be operated with mW of optical pump have been demonstrated, by exploiting the choice offered by compound III–V semiconductors, and the possibility to control the dynamics of free carriers through surface treatment, doping, and the geometry. Moreover, it has been shown that such devices can be integrated on a silicon photonic platform, and thereby offering radically novel functionalities which could be combined with the panoply already available there, to create advanced devices for optical signal processing.

Acknowledgements We acknowledge support from the French Research Agency (ANR) through the contracts Qswitch, Auctopuss and Ethan, and the European Commission, through the contracts Qphoton, Gospel, Copernicus. We thank Fabrice Raineri, Johann-Peter Reithmaier and Gadi Eisenstein for fruitful discussions.

References

1. Fujitsu europe press release, fujitsu launches second generation ultra-fast 65 GSa/s 8-bit ADC technology for 100G optical transport (2010), http://www.fujitsu.com/emea/news/pr/fseu-en_20100913-978.html
2. Y. Akahane, T. Asano, B.S. Song, S. Noda, High-Q photonic nanocavity in a two-dimensional photonic crystal. Nature **425**(6961), 944–947 (2003)

3. R. Akimoto, S. Gozu, T. Mozume, H. Ishikawa, Monolithically integrated all-optical gate switch using intersubband transition in InGaAs/AlAsSb coupled double quantum wells. Opt. Exp. **19**(14), 13386–13394 (2011)
4. V.R. Almeida, M. Lipson, Optical bistability on a silicon chip. Opt. Lett. **29**(20), 2387–2389 (2004)
5. M.E. Grein, A.H. Nejadmalayeri, C.W. Holzwarth, M.Y. Sander, M.S. Dahlem, Y. Michael, P. Michael, W. Geis Nicole, A. DiLello Jung, U. Yoon Ali Motamedi... E.P. Ippen, A. Khilo, S.J. Spector, F.X. Krtner, Photonic ADC: Overcoming the bottleneck of electronic jitter. Optics Express **20**(4), 4454 (2012)
6. T. Aoki, N. Fukuhara, T. Osada, H. Sazawa, M. Hata, T. Inoue, Nitride passivation reduces interfacial traps in atomic-layer-deposited Al_2O_3/GaAs (001) metal-oxide-semiconductor capacitors using atmospheric metal-organic chemical vapor deposition. Appl. Phys. Lett. **105**(3), 033513 (2014)
7. A. Bazin, K. Lenglé, M. Gay, P. Monnier, L. Bramerie, R. Braive, G. Beaudoin, I. Sagnes, R. Raj, F. Raineri, Ultrafast all-optical switching and error-free 10 Gbit/s wavelength conversion in hybrid InP-silicon on insulator nanocavities using surface quantum wells. Appl. Phys. Lett. **104**(1), 011102 (2014)
8. A. Bazin, P. Monnier, X. Lafosse, G. Beaudoin, R. Braive, I. Sagnes, R. Raj, F. Raineri, Thermal management in hybrid InP/silicon photonic crystal nanobeam laser. Opt. Express **22**(9), 10570–10578 (2014)
9. B.R. Bennett, R.A. Soref, J.A. del Alamo, Carrier-induced change in refractive index of InP, GaAs and InGaAsP. Quantum Electron. IEEE J. **26**(1), 113–122 (1990)
10. R. Bose, J.S. Pelc, S. Vo, C.M. Santori, R.G. Beausoleil, Carrier dynamics in GaAs photonic crystal cavities near the material band edge. Opt. Express **23**(10), 12732 (2015)
11. A.D. Bristow, J.-P.R. Wells, W.H. Fan, A.M. Fox, M.S. Skolnick, D.M. Whittaker, A. Tahraoui, T.F. Krauss, J.S. Roberts, Ultrafast nonlinear response of AlGaAs two-dimensional photonic crystal waveguides. Appl. Phys. Lett. **83**(5), 851 (2003)
12. J. Capmany, D. Novak, Microwave photonics combines two worlds. Nat. Photonics **1**(6), 319–330 (2007)
13. N. Cazier, X. Checoury, L.-D. Haret, P. Boucaud, High-frequency self-induced oscillations in a silicon nanocavity. Opt. Express **21**(11), 13626–13638 (2013)
14. S. Combrié, S. Bansropun, M. Lecomte, O. Parillaud, S. Cassette, H. Benisty, J. Nagle, Optimization of an inductively coupled plasma etching process of Ga In P/Ga As based material for photonic band gap applications a. J. Vac. Sci. Technol. B **23**(4), 1521–1526 (2005)
15. Sylvain Combrié, D. Rossi, Q.V. Tran, H. Benisty, *GaAs* photonic crystal cavity with ultrahigh Q: microwatt nonlinearity at 1.55 μm. Opt. Let. **33**, 1908–1910 (2008)
16. S. Combrié, G. Lehoucq, A. Junay, S. Malaguti, G. Bellanca, S. Trillo, L. Menager, J.P. Reithmaier, A. De Rossi, All-optical signal processing at 10 GHz using a photonic crystal molecule. Appl. Phys. Lett. **103**(19), 193510 (2013)
17. G.W. Cong, R. Akimoto, K. Akita, T. Hasama, H. Ishikawa, Low-saturation-energy-driven ultrafast all-optical switching operation in (CdS/ZnSe)/BeTe intersubband transition. Opt. Express **15**(19), 12123–12130 (2007)
18. N.J. Doran, D. Wood, Nonlinear-optical loop mirror. Opt. Letters **13**(1), 56–58 (1988)
19. H.J.S. Dorren, M.T. Hill, Y. Liu, N. Calabretta, A. Srivatsa, F.M. Huijskens, H. de Waardt, G.D. Khoe, Optical packet switching and buffering by using all-optical signal processing methods. J. Lightwave Technol. **21**(1), 2–12 (2003)
20. M. Elbadry, R. Harjani, *Quadrature Frequency Generation for Wideband Wireless Applications*. Springer (2015)
21. M.A. Foster, A.C. Turner, J.E. Sharping, B.S. Schmidt, M. Lipson, A.L. Gaeta, Broad-band optical parametric gain on a silicon photonic chip. Nature **441**(7096), 960–963 (2006)
22. P. Ghelfi, F. Laghezza, F. Scotti, G. Serafino, A. Capria, S. Pinna, D. Onori, C. Porzi, M. Scaffardi, A. Malacarne, V. Vercesi, E. Lazzeri, F. Berizzi, A. Bogoni, A fully photonics-based coherent radar system. Nature **507**(7492), 341–345 (2014)

23. H.M. Gibbs, S.L. McCall, T.N.C. Venkatesan, A.C. Gossard, A. Passner, W. Wiegmann, Optical bistability in semiconductors. Appl. Phys. Lett. **35**(6), 451 (1979)
24. H.M. Gibbs, S.L. McCall, T.N.C. Venkatesan, Differential gain and bistability using a sodium-filled fabry-perot interferometer. Phys. Rev. Lett. **36**(19), 1135–1138 (1976)
25. Y. Halioua, A. Bazin, P. Monnier, T.J. Karle, G. Roelkens, I. Sagnes, R. Raj, F. Raineri, Hybrid III-V semiconductor/silicon nanolaser. Opt. Express **19**(10), 9221–9231 (2011)
26. K.L. Hall, J. Mark, E.P. Ippen, G. Eisenstein, Femtosecond gain dynamics in InGaAsP optical amplifiers. Appl. Phys. Lett. **56**(18), 1740 (1990)
27. T.D. Happ, M. Kamp, A. Forchel, J.-L. Gentner, L. Goldstein, Two-dimensional photonic crystal coupled-defect laser diode. Appl. Phys. Lett. **82**(1), 4–6 (2003)
28. M.J.R. Heck, H.-W. Chen, A.W. Fang, B.R. Koch, D. Liang, H. Park, M.N. Sysak, J.E. Bowers, Hybrid silicon photonics for optical interconnects. IEEE J. Sel. Topics Quantum Electron. **17**(2), 333–346 (2011)
29. M. Heuck, S. Combrié, G. Lehoucq, S. Malaguti, G. Bellanca, S. Trillo, P.T. Kristensen, J. Mørk, J.P. Reithmaier, A. de Rossi, Heterodyne pump probe measurements of nonlinear dynamics in an indium phosphide photonic crystal cavity. Appl. Phys. Lett. **103**(18), 181120 (2013)
30. C. Husko, A. De Rossi, S. Combrié, Q. V Tran, F. Raineri, C.W. Wong, Ultrafast all-optical modulation in GaAs photonic crystal cavities. Appl. Phys. Lett. **94**(2), 021111 (2009)
31. S. Ishii, K. Nozaki, T. Baba, Photonic molecules in photonic crystals. Jap. J. Appl. Phys. **45**(8A), 6108–6111 (2006)
32. P. Bermel, S.G. Johnson, J.D. Joannopoulos, J. Bravo-Abad, A. Rodriguez, M. Soljačić, Enhanced nonlinear optics in photonic-crystal microcavities. *Opt. Express* **15**(24) (2007)
33. J.L. Jewell, S.L. McCall, A. Scherer, H.H. Houh, N.A. Whitaker, A.C. Gossard, J.H. English, Transverse modes, waveguide dispersion, and 30 ps recovery in submicron GaAs/AlAs microresonators. Appl. Phys. Lett. **55**(1), 22 (1989)
34. S. John, Strong localization of photons in certain disordered dielectric superlattices. Phys. Rev. Lett. **58**(23), 2486–2489 (1987)
35. S.G. Johnson, P.R. Villeneuve, S. Fan, J.D. Joannopoulos, Linear waveguides in photonic-crystal slabs. Phys. Rev. B **62**(12), 8212 (2000)
36. J.B. Khurgin, How to deal with the loss in plasmonics and metamaterials. Nat. Nanotechnol. **10**(1), 2–6 (2015)
37. T.K. Kim, Y. Song, K. Jung, C. Kim, H. Kim, C.H. Nam, J. Kim, Sub-100-as timing jitter optical pulse trains from mode-locked er-fiber lasers. Opt. Lett. **36**(22), 4443 (2011)
38. S. Kiravittaya, H.S. Lee, L. Balet, L.H. Li, M. Francardi, A. Gerardino, A. Fiore, A. Rastelli, O.G. Schmidt, Tuning optical modes in slab photonic crystal by atomic layer deposition and laser-assisted oxidation. J. Appl. Phys. **109**(5), 053115–053115 (2011)
39. T.F. Krauss, R.M. De La Rue, S. Brand, Two-dimensional photonic-bandgap structures operating at near-infrared wavelengths. Nature **383**(6602), 699–702 (1996)
40. H.S. Lee, S. Kiravittaya, S. Kumar, J.D. Plumhof, L. Balet, L.H. Li, M. Francardi, A. Gerardino, A. Fiore, A. Rastelli, O.G. Schmidt, Local tuning of photonic crystal nanocavity modes by laser-assisted oxidation. Appl. Phys. Lett. **95**(19), 191109 (2009)
41. S. Malaguti, G. Bellanca, A. De Rossi, S. Combrie, S. Trillo, Self-pulsing driven by two-photon absorption in semiconductor nanocavities. Phys. Rev. A **83**, 051802 (2011)
42. C. Manolatou, M.J. Khan, S. Fan, P.R. Villeneuve, H.A. Haus, J.D. Joannopoulos, Coupling of modes analysis of resonant channel add-drop filters. IEEE J. Quantum Electron. **35**(9), 1322–1331 (1999)
43. J. Mark, J. Mork, Subpicosecond gain dynamics in InGaAsP optical amplifiers: Experiment and theory. Appl. Phys. Lett. **61**(19), 2281 (1992)
44. D. Marpaung, C. Roeloffzen, R. Heideman, A. Leinse, S. Sales, J. Capmany, Integrated microwave photonics: Integrated microwave photonics. Laser Photon. Rev. **7**(4), 506–538 (2013)
45. A. Martínez, J. Blasco, P. Sanchis, V. José Galán, J. Garcá-Rupérez et al., Ultrafast all-optical switching in a silicon-nanocrystal-based silicon slot waveguide at telecom wavelengths. Nanoletters **10**, 15061511 (2010)

46. M. Matsuura, O. Raz, F. Gomez-Agis, N. Calabretta, H.J.S. Dorren, 320 Gbit/s wavelength conversion using four-wave mixing in quantum-dot semiconductor optical amplifiers. Opt. Lett. **36**(15), 2910 (2011)
47. C. Meuer, C. Schmidt-Langhorst, R. Bonk, H. Schmeckebier, D. Arsenijevi, G. Fiol, A. Galperin, J. Leuthold, C. Schubert, D. Bimberg, 80 Gb/s wavelength conversion using a quantum-dot semiconductor optical amplifier and optical filtering. Opt. Express **19**(6), 5134–5142 (2011)
48. G. Moille, S. Combrié, A. De Rossi, Modeling of the carrier dynamics in nonlinear semiconductor nanoscale resonators. Phys. Rev. A **94**, 023814 (2016)
49. G. Moille, S. Combrié, K. Fuchs, M. Yacob, J.P. Reithmaier, A. de Rossi, Acceleration of the nonlinear dynamics in p-doped indium phosphide nanoscale resonators. Opt. Lett **42**, 795–798 (2017)
50. G. Moille, S. Combrié, L. Morgenroth, G. Lehoucq, F. Neuilly, H. Bowen, D. Decoster, A. de Rossi, Integrated all-optical switch with 10ps time resolution enabled by ald. Laser Photon. Rev. **10**(3), 409–419 (2016)
51. H.C.H. Mulvad, M. Galili, L.K. Oxenlwe, H. Hao, A.T. Clausen, J.B. Jensen, C. Peucheret, P. Jeppesen, Demonstration of 5.1 Tbit/s data capacity on a single-wavelength channel. Opt. Express **18**(2), 1438–1443 (2010)
52. M.P. Nezhad, A. Simic, O. Bondarenko, B. Slutsky, A. Mizrahi, L. Feng, V. Lomakin, Y. Fainman, Room-temperature subwavelength metallo-dielectric lasers. Nat. Photon. **4**(6), 395–399 (2010)
53. D.D. Nolte, Surface recombination, free-carrier saturation, and dangling bonds in InP and GaAs. Solid-state Electron. **33**(2), 295–298 (1990)
54. M. Notomi, S. Mitsugi, Wavelength conversion via dynamic refractive index tuning of a cavity. Phys. Rev. A **73**, 051803(R) (2006)
55. K. Nozaki, T. Tanabe, A. Shinya, S. Matsuo, T. Sato, H. Taniyama, M. Notomi, Sub-femtojoule all-optical switching using a photonic-crystal nanocavity. Nat. Photon. **4**(7), 477–483 (2010)
56. O. Painter, R.K. Lee, A. Scherer, A. Yariv, J.D. O'brien, P.D. Dapkus, I. Kim, Two-dimensional photonic band-gap defect mode laser. Science **284**(5421), 1819–1821 (1999)
57. K. Poulton, R. Neff, B. Setterberg, B. Wuppermann, A 20-GSample/s 8b ADC with a 1-MByte Memory in 0.18-um CMOS. *IEEE International Solid State Circuits Conference Digest of Technical Papers* (2003)
58. D.F.Q. Xu, R.G. Beausoleil, Silicon microring resonators with 1.5-μm radius. Opt. Express **16**(6), 4309 (2008)
59. P. Schvan, J. Bach, C. Falt, P. Flemke, R. Gibbins, Y. Greshishchev, N. Ben-Hamida, D. Pollex, J. Sitch, S.-C. Wang, J. Wolczanski, A 24GS/s 6b ADC in 90 nm CMOS. In *2008 IEEE International Solid-State Circuits Conference—Digest of Technical Papers*, pp. 544–634. IEEE (2008)
60. A.J. Seeds, Microwave photonics. Microw. Theor. Tech. IEEE Trans. **50**(3), 877–887 (2002)
61. H. Sekoguchi, Y. Takahashi, T. Asano, S. Noda, Photonic crystal nanocavity with a q-factor of 9 million. Opt. Express **22**(1), 916–924 (2014)
62. M. Sheik-Bahae, D.J. Hagan, E.W. Van Stryland, Dispersion and band-gap scaling of the electronic kerr effect in solids associated with two-photon absorption. Phys. Rev. Lett. **65**, 96–99 (1990)
63. M. Sheik-Bahae, D.C. Hutchings, D.J. Hagan, E.W. Van Stryland, Dispersion of bound electron nonlinear refraction in solids. IEEE J. Quantum Electron. **27**(6), 1296–1309 (1991)
64. W. Shockley, *Electrons and holes in semiconductors: with applications to transistor electronics*. Van Nostrand (1950)
65. G.A. Siviloglou, S. Suntsov, R. El-Ganainy, R. Iwanow, G.I. Stegeman, D.N. Christodoulides, R. Morandotti, D. Modotto, A. Locatelli, C. De Angelis, Enhanced third-order nonlinear effects in optical AlGaAs nanowires. Opt. Express **14**(20), 9377–9384 (2006)
66. M. Soljačić, J.D. Joannopoulos, Enhancement of nonlinear effects using photonic crystals. Nat. Mater. **3**, 211–219 (2004)
67. M. Soljačić, M. Ibanescu, S.G. Johnson, Y. Fink, J.D. Joannopoulos, Optimal bistable switching in nonlinear photonic crystals. *Phys. Rev. E***66**(5) (2002)

68. G.I. Stegeman, E.M. Wright, N. Finlayson, R. Zanoni, C.T. Seaton, Third order nonlinear integrated optics. Lightwave Technol. J. **6**(6), 953–970 (1988)
69. N. Shimizu, T. Ishibashi, S. Kodama, T. Furuta, High-speed response of uni-traveling-carrier photodiodes
70. T. Tanabe, K. Nishiguchi, A. Shinya, E. Kuramochi, H. Inokawa, M. Notomi, K. Yamada, T. Tsuchizawa, T. Watanabe, H. Fukuda, H. Shinojima, S. Itabashi, Fast all-optical switching using ion-implanted silicon photonic crystal nanocavities. Appl. Phys. Lett. **90**(3), 031115–031115 (2007)
71. T. Tanabe, M. Notomi, S. Mitsugi, A. Shinya, E. Kuramochi, All-optical switches on a silicon chip realized using photonic crystal nanocavities. Appl. Phys. Lett. **87**(15), 151112 (2005)
72. T. Tanabe, H. Taniyama, M. Notomi, Carrier diffusion and recombination in photonic crystal nanocavity optical switches. J. Lightwave Technol. **26**(11), 1396–1403 (2008)
73. G. Tayeb, D. Maystre, Rigorous theoretical study of finite-size two-dimensional photonic crystals doped by microcavities. J. Opt. Soc. Am. A **14**(12), 3323–3332 (1997)
74. Q. Vy Tran, S. Combrié, P. Colman, A. De Rossi, Photonic crystal membrane waveguides with low insertion losses. Appl. Phys. Lett. **95**(6), 231104–231104 (2009)
75. K.J. Vahala, Optical microcavities. Nature **424**, 839–846 (2003)
76. G.C. Valley, Photonic analog-to-digital converters. Opt. Express **15**(5), 1955–1982 (2007)
77. V. Van, T.A. Ibrahim, P.P. Absil, F.G. Johnson, R. Grover, P.-T. Ho, Optical signal processing using nonlinear semiconductor microring resonators. Sel. Topics Quantum Electron. IEEE J. **8**(3), 705–713 (2002)
78. D. Vukovic, Y. Yu, M. Heuck, S. Ek, N. Kuznetsova, P. Colman, E. Palushani, J. Xu, K. Yvind, L. Oxenloewe et al., Wavelength conversion of a 9.35 Gb/s RZOOK signal in an Inp photonic crystal nanocavity. *Photon. Technol. Lett.***26**(257) (2014)
79. I.A. Walmsley, C. Dorrer, Characterization of ultrashort electromagnetic pulses. Adv. Opt. Photon. **1**(2), 308 (2009)
80. E. Weidner, S. Combrié, A. De Rossi, N.V. Quynh, S. Cassette, Nonlinear and bistable behavior of an ultrahigh-Q GaAs photonic crystal nanocavity. Appl. Phys. Lett. **90**(1), 101118 (2007)
81. A.E. Willner, S. Khaleghi, M.R. Chitgarha, O.F. Yilmaz, All-optical signal processing. J. Lightwave Technol. **32**(4), 660–680 (2014)
82. E. Yablonovitch, Inhibited spontaneous emission in solid-state physics and electronics. Phys. Rev. Lett. **58**(20), 2059–2062 (1987)
83. A.M. Yacomotti, F. Raineri, C. Cojocaru, P. Monnier, A. Levenson, R. Raj, Nonadiabatic dynamics of the electromagnetic field and charge carriers in high-Q photonic crystal resonators. Phys. Rev. Lett. **96**, 093901 (2006)
84. P.D. Ye, G.D. Wilk, M.M. Frank, Processing and characterization of III–V compound semiconductor MOSFETs using atomic layer deposited gate dielectrics. In Athanasios Dimoulas, Evgeni Gusev, PaulC. McIntyre, and Marc Heyns, editors, *Advanced Gate Stacks for High-Mobility Semiconductors*, vol. 27 of *Advanced Microelectronics*, pp. 341–361. Springer, Berlin Heidelberg (2007)
85. Y. Yi, Y. Chen, H. Hao, W. Xue, K. Yvind, J. Mork, Nonreciprocal transmission in a nonlinear photonic-crystal Fano structure with broken symmetry: Nonreciprocal transmission in a nonlinear Fano structure. Laser Photon. Rev. **9**(2), 241–247 (2015)
86. Y. Yi, M. Heuck, S. Ek, N. Kuznetsova, K. Yvind, J. Mørk, Experimental demonstration of a four-port photonic crystal cross-waveguide structure. Appl. Phys. Lett. **101**(25), 251113–251113 (2012)
87. Y. Yi, E. Palushani, M. Heuck, N. Kuznetsova, P.T. Kristensen, S. Ek, D. Vukovic, C. Peucheret, L.K. Oxenløwe, S. Combrié et al., Switching characteristics of an Inp photonic crystal nanocavity: Experiment and theory. Opt. Express **21**(25), 31047–31061 (2013)
88. Z. Zhang, M. Qiu, Small-volume waveguide-section high Q microcavities in 2D photonic crystal slabs. Opt. Express **12**(17), 3988–3995 (2004)

Chapter 6
Alternative Logic Families for Energy-Efficient and High Performance Chip Design

Itamar Levi and Alexander Fish

Abstract With advances in technology and the expansion of mobile applications, energy consumption, which is one of the fundamental limits in both high performance microprocessors and low to medium performance portable systems.

6.1 Introduction

With advances in technology and the expansion of mobile applications, energy consumption, which is one of the fundamental limits of both high performance microprocessors and low to medium performance portable systems, has become a primary focus of attention in VLSI digital design [1–6]. In high performance systems, energy and peak power are the limiting factors on further increases in clock speed and circuit density, due to the difficulties of conveying power to circuits and removing the heat they generate. Moreover, the integration of circuits with different workloads and activity profiles results in the formulation of hot spots and temperature gradients over the die. This can impact the long-term reliability and complicate the verification of the processor [7]. In portable battery operated devices such as cellular phones, bio-medical devices, sensor networks, etc., energy consumption is critical since it determines the lifetime of the battery (for non-rechargeables) or the time between recharges. It is also affects the packaging, cost and weight. Many architectures and techniques have been researched, analyzed and proposed for power reduction and energy minimization of combinational circuits [1, 8, 9]. In general, power reduction can be implemented at different levels of design abstraction: the algorithm, system, architecture, gate, circuit and/or the technology. At the algorithm level, for example, methods for simplifying the logic involved in computation and coding for smaller Hamming distances are being developed [10–13]. Examples of

I. Levi (✉) · A. Fish
Faculty of Engineering Bldg, Bar-Ilan University, 1102 Ramat-Gan, Israel
e-mail: itamar.levi@biu.ac.il

A. Fish
e-mail: Alexander.fish@biu.ac.il

© Springer International Publishing AG 2017

G. Eisenstein and D. Bimberg (eds.), *Green Photonics and Electronics*,
NanoScience and Technology, DOI 10.1007/978-3-319-67002-7_6

energy reduction at the system level include commonly used power-save modes, dynamic voltage scaling and leakage-power-reduction management [14–21]. The sleepy-stack/keeper approach to reduce leakage power [22], reduced swing techniques by special circuit families such as Current Mode Logic (CML) [22, 23] and the use of high cost functions for power hungry gates are examples of commonly used techniques at the gate level. A variety of design techniques, such as Input Vector Control (IVC) [24–26], Reverse Body Biasing (RBB) technique [27–30], Dynamic Threshold CMOS (DTMOS) [31, 32] and many others can be implemented at the circuit and transistor levels to reduce both dynamic and static power [3, 33–36]. Finally, at the technology level, supply voltage reduction can be utilized to effectively reduce both dynamic and leakage power [37–39].

During the process of Standard Design Flow (SDF) the logical representation of the Integrated Circuits (ICs) (in High Description Language, HDL) is synthesized to a library of standard digital cells. These logical cells are implemented according to a well optimized layout structure as well as the device dimensions and physical guidelines of the digital library provider. Digital libraries possess different flavors of gates. Most of the digital circuits are implemented using standard CMOS logic. Static CMOS Logic has been the most popular design approach for the past three decades. Many attempts have been made to propose a better alternative logic family to achieve lower power dissipation, smaller area and higher performance.

In the past, Pass-Transistor Logic (PTL) was proposed as a promising alternative to Static CMOS Logic [33, 40–45]. The leakage of PTL implementations of monotonic gates was shown to be much higher than that of CMOS implementations. Some PTL techniques, such as Double PTL (DPL), were proposed to solve the aforementioned problems [40–42]. However, most of these solutions resulted in increased transistor count and area, a large number of required buffers and degradation in signal integrity. Dynamic Logic [43–45] has been used as an efficient alternative for high performance operation. Computation using dynamic gates operates in two phases: pre-charge and evaluation. The advantages of dynamic logic include the high driving strength of the evaluation network and reduced transistor count. On the other hand, dynamic logic presents a number of significant drawbacks, such as charge leakages, charge sharing, signal integrity and restoration issues, high dynamic power consumption and susceptibility to glitches with no data recovery. These sensitivities are intensified with process scale-down, supply voltage reduction and process variations increase.

In a given standard Cell Library, the same gate can be designed in several flavors optimized for different objectives. While utilizing a digital logic family such as CMOS, the traditional paradigm assumes that high performance comes at the expense of energy efficiency; that is, one can design low energy cells that operate at low frequencies or high performance ones that consume high energy. Research and optimizations to disentangle this paradigm (i.e. offer the best of both worlds) are typically carried out at the algorithm or architectural levels of the design by operations in different modes, dynamic voltage and frequency scaling (DVFS), different power domains, algorithmic optimization, etc. The task is more difficult at lower abstraction levels such as the gate and transistor levels. When utilizing

standard CMOS gates, the optimization space at the gate level is quite small. For the sake of clarity, it should be noted that low energy flavors of cells (gates) are typically limited by the amount of energy reduction they can provide as compared to the high performance flavors for the same supply voltage and frequency.

In traditional sequential designs, performance is dominated by the circuit's most critical (slowest) path, whereas the energy consumption is basically the sum over all the design consumers (i.e., all the gates/cells in the design). The typical optimization carried out by the automated tools, targeted by a specification to achieve both low energy and high performance, is to find the design's most critical paths and to assign high performance gates to them (gates consuming relatively high energy) while the rest of the paths (that are not timing-critical) can remain in a fairly low performance mode that consumes low energy.

In this chapter, we present the recently proposed alternative dual mode logic (DML) gates family that enables a very high level of energy-delay optimization flexibility at the gate level. This flexibility can be utilized to dramatically improve both the energy efficiency and the performance of combinational circuits by manipulating their critical and noncritical paths. We present an approach that locates the design's critical paths and operates them in a boosted performance mode (denoted as *Dynamic*). The noncritical paths are operated in the low energy DML mode (denoted as *Static*), which does not affect the performance of the design, but results in a significant reduction in energy consumption. The key strength of this method is that the same gate can be selectively operated in each mode (in a gate-level granularity). This provides much greater room for optimization at higher algorithmic and system abstraction levels, in addition to the inherent gate level opportunities which are not possible with a CMOS based design. We discuss DML behavior in different operation scenarios, focusing on energy efficient low voltage operation.

6.2 Background

This section discusses energy-performance tradeoffs in digital designs, and focuses on operations in different regions. Then, we briefly discuss existing logical families for digital design and elaborate on their tradeoffs in design metrics. Finally, we analyze the typical energy-delay (*E-D*) optimization space paradigm and highlight how DML can extend it. The dual modularity of the DML gates family enables a clock-to-clock-cycle selective operation in one of the two DML operating modes, one of which (*dynamic*) enables high performance and the other (*static*) allows low energy consumption.

1. Energy-Delay (E-D) Tradeoffs.

Energy and performance have been the main metrics used to assess the efficiency of digital electronic systems over the years. Traditionally, energy efficiency comes at the expense of high performance and vice versa. R.G. Dreslinski et al. presented the

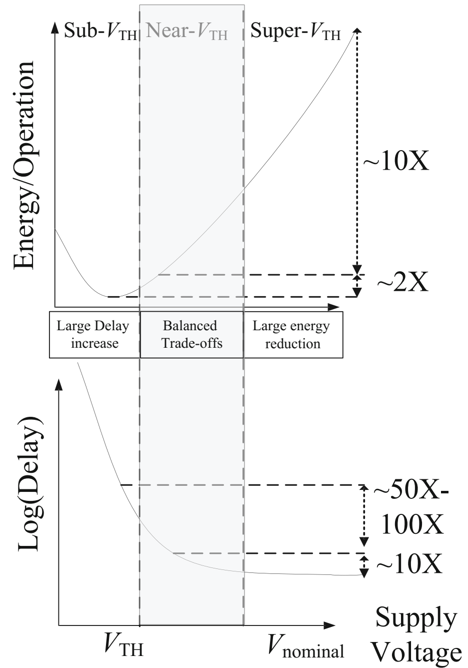


Fig. 6.1 Energy and performance tradeoff for different supply voltages. Reproduced from [46]

general tradeoff space of energy and performance as a function of the supply voltage (V_{DD}) [46]. Figures 6.1 and 6.2 show very popular illustrations of the tradeoffs between energy per operation and speed for different supply voltages.

The traditional designs that dominate today's market use supply voltages in the range of 0.9–1.8 V to operate their digital and analog circuits. In these designs all the "on" transistors operate in the so-called-Super-V_{TH} region (see Fig. 6.1), far above the switching threshold of a transistor. In this region the "I_{on}" current of a transistor used for switching digital gates from one state to another is very strong, which results in a ratio of many orders of magnitude between the "I_{on}" and the "I_{off}" (parasitic leakage) currents. This makes the "Super-V_{TH}" operation very fast and reliable. The minimum delay point (MDP), which indicates the best possible performance of a given circuit topology, is achieved in the "Super-V_{TH}" region, as shown in Fig. 6.2. However, these traditional designs consume very high energy and are not suitable for many modern applications where energy dissipation is the main concern.

Low voltage operation in the "sub-V_{TH}" or "near-V_{TH}" regions has been shown to be the ideal way to dramatically reduce energy dissipation. In the sub-V_{TH} designs all transistors are operated from the supply voltage which is below the transistor switching threshold voltage. This approach, which was proposed as far back as the 1960s, is radically different since operation in the sub-threshold regions exploits the parasitic leakage current, and uses it as its primary operation current. This means that according to the classical definition most of the transistors are in

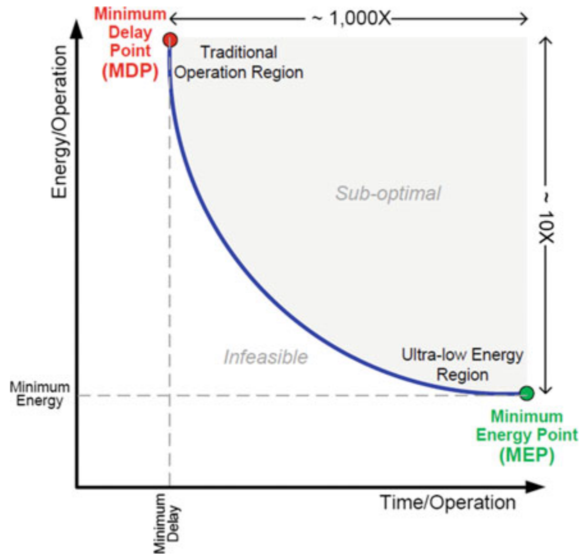


Fig. 6.2 Pareto-optimal energy-delay curve showing the Minimum Energy Point (MEP) and the minimum delay point (MDP) [102]

the "off" state even during their operation and switching. Sub-threshold operation can substantially reduce both leakage and switching (dynamic) energy dissipation, and thus result in minimum energy dissipation. Dynamic power is greatly reduced, primarily due to the quadratic dependency on supply voltage and the complete elimination of some of its components. Similarly, the static leakage is also much lower since it also depends on the supply voltage (exponential dependency at low voltages). The Minimum Energy Point (MEP) is usually located in the sub-threshold region, as shown in Figs. 6.1 and 6.2. Lowering the supply voltage below the MEP causes an increase in energy consumption (see Fig. 6.1). Below the MEP, the delay of a circuit is exponentially larger (exponential increase with supply voltage reduction). In turn this makes the leakage currents of sub-V_{TH} devices start to dominate the energy per operation.

While maximum energy efficiency is achieved in the sub-V_{TH} region, since the sub-threshold currents are much weaker than standard "super-threshold" currents, the time needed to change a digital gate state is significantly longer, which limits the operation frequency of the circuit considerably. At the MEP the delay is at least three orders of magnitude larger than at the Minimum Delay Point (MDP), which consumes an order of magnitude more energy.

The energy-delay (E-D) curve is quite flat around the MEP, so significant performance improvement can be achieved by a slight increase in the supply voltage above the MEP and moving into the near-V_{TH} region (see Fig. 6.2). In "near-V_{TH}" designs, all transistors are operated between the weak and moderate inversion regions. Since operation in the NT region presents a good tradeoff in terms of energy-performance leading to significant energy reduction with only a moderate

penalty in performance (as compared to the super-threshold region), it has become very attractive for many modern applications.

At first glance the above description suggests that a simple reduction in the power supply voltage of traditional circuits could yield reliable sub/near-threshold operation. Unfortunately, this is not the case since the power supply reduction is accompanied by a number of problems and significant challenges. The low voltage associated with frequency reduction is not suitable for all modes of operation and an adaptive voltage control mechanism may be required. Lower supply voltages also mean lower noise margins, reduced yield and increased vulnerability to process variations and temperature fluctuations. The characteristics of semiconductor behavior in sub/near-threshold are not well represented by standard transistors models and are different from those in super-threshold region, resulting in different device sizing and ratio optimizations.

Although some logic families, such as CMOS, are known to be fully operational in all voltage regions, many logic families such as the ratioed-logic and dynamic families have been shown to be inapplicable in low voltage regions. An alternative Dual-Mode-Logic (DML) is shown below to be fully functional at all voltage regions. Furthermore, as will be presented, the DML's capability to switch between different modes of operation (Static or Dynamic) makes it possible to achieve Energy efficient operation under wide range of supply voltages, while exhibiting significant improvement in performance as compared to conventional CMOS logic.

2. Introduction to Logic families for Digital circuit design.

Numerous circuit styles can implement a given logic function. These styles are called logic families. Every logic family inherently has its own advantages and drawbacks, and a designer's choice of an appropriate family depends on the application and its specifications such as energy, performance, area, temperature, reliability, etc.

The most common and well-known logic styles are Complementary Metal Oxide Semiconductor (CMOS) [33–35, 47], Pseudo nMOS [48], Pass Transistor Logic (PTL) [40–42, 49], Gate Diffusion Input (GDI) [50–54], Dynamic logic (i.e. Domino/CMOS NORA, etc.) [43–45], Current Mode Logic (CML) [22, 23], Sense Amplifier Based Logic (SABL) [55, 56]. This range of styles has shrunk over the last three decades to only a handful logic families which are widely used nowadays (mainly CMOS for non-custom designs). Modern designs consist of a limited selection of well-known and explored logic families for automate or semi-automate design flow. In the following sub-sections, we introduce the key features of the most common logic families. We primarily address logic styles that serve as the basic building blocks for DML.

(1) *Complementary Metal Oxide Semiconductor (CMOS)*

The most common design logic family today is CMOS. This family is based on the use of complementary MOS transistors to perform logic functions (see Fig. 6.3) with a very small consumption of static current. CMOS gates are based on the fundamental inverter circuit consisting of two transistors. Both transistors are

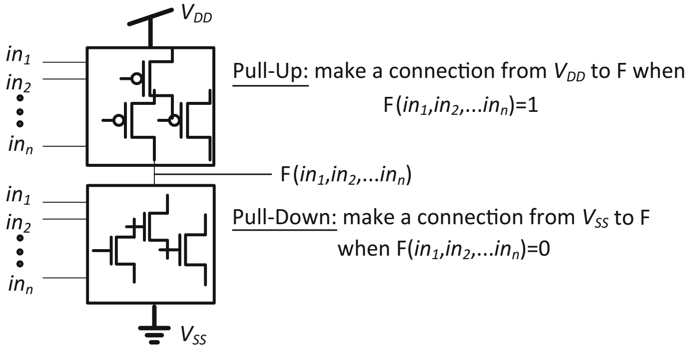


Fig. 6.3 Standard CMOS logic. Reproduced from [22]

MOSFETs, with one n-channel whose source is grounded, and one p-channel whose source is connected to the V_{DD}. Their gates are connected to form the input, and their drains are connected to form the output. The two MOSFETs are designed to have matching characteristics and hence are complementary to each other. When off, their resistance is effectively infinite; when on, their channel resistance is low. Since the gate is essentially an oxide isolated circuit, it draws no current in the steady state and the output voltage is equal to one of the strong power supply voltages, depending on which transistor is conducting [22].

The advantages of a conventional CMOS design methodology are well-known and explored, and partially detailed. These include strong on and off states, rail to rail logic levels, and until the advent of recent processes, it also featured very low static power consumption. The main drawbacks of CMOS are the large number of transistors (twice the number of inputs) which reflects very large input and output capacitances that increase the delay. Moreover, in most advanced nanoscale processes where the feature size is scaled to below 65 nm, the static leakage current increases dramatically as a result of the increment in the subthreshold slope factor [1]. This issue has become a significant obstacle to low voltage designs. The increased leakage results in a decreased *on/off* current ratio and therefore an increased delay. The decreased *on/off* ratio paves the way for high-probability failure mechanisms, especially under global process variations for high frequency applications if not designed correctly [57, 58]. More common sensitivities come into play with low process nodes such as Gate Induced Drain Leakage (GIDL), Drain Induced Barrier Lowering (DIBL) [58, 59], Punch Through [60], Gate tunneling [61] etc. CMOS performance under low voltage regimes is discussed in the next sub-section.

In order to limit energy consumption in current and future CMOS technology generations, the scaling of the threshold voltage needs to keep up with performance requirements. However, the techniques used to decrease the threshold voltage increase its variability. In CMOS, gates are generally designed with a stacked

network and a parallel network, whose leakage can be substantial. Thus, summing large leakage currents which are intensified by V_{TH} variations makes the CMOS design particularly sensitive to process variations in nanometer CMOS. V_{TH} sensitivity is caused by random dopant fluctuations, line edge roughness, oxide thickness variations, etc. These effects combine to yield an exponential change in the ON or leakage (OFF) currents under low voltage operation [62].

Utilization of low voltage CMOS gates is associated with a number of challenges [63], including low worst-case frequency (which is determined by the design's most critical path), relatively large leakage currents in modern processes, large area and capacitances. All these make low voltage CMOS design impractical in many applications where performance is important, leaving CMOS low voltage design for applications such as medical devices [64, 65] and portable applications with very low performance requirements [6, 66]. Several attempts to use modified CMOS logic to operate under low voltage with boosted performance were made by Soeleman and Roy [67]. These utilized a dynamically switching threshold voltage but were shown to be less stable for sub-threshold operation under process variation and temperature fluctuations.

Unlike the complementary CMOS which implements a symmetric *Voltage Transfer Characteristic (VTC)* and has robust high and low noise margins, Dynamic Logic [68–70] was proposed traditionally to counterbalance the symmetric nature of CMOS to provide high performance. Naturally this comes at the expense of energy consumption, as will be discussed next.

(2) *Dynamic logic*

Dynamic Logic [68–70] was developed for high performance digital circuits. Dynamic logic can be described as a technique to operate specially designed logic gates in two different phases: the pre-charge phase where the dynamic node is charged to logic level "1", and the evaluation phase which may lead to a discharge of the dynamic output node. The two phases are synchronized using a CLK signal. Figure 6.4 depicts the basic scheme of a dynamic gate.

In contrast to static families (constant low resistive connections to a constant supply of V_{DD} or G_{ND}, e.g. CMOS) a dynamic design has very high performance,

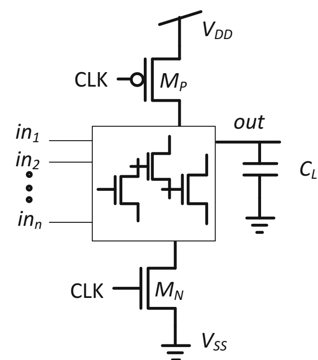


Fig. 6.4 Basic scheme of a dynamic gate. Reproduced from [69]

reduced area utilization and inherently eliminates short-circuit currents. Static gates can evolve to 'high' or 'low' logic levels with equal probability; therefore, they are sized for equal High-to-Low *(T_{PHL})* and Low-to-High *(T_{PLH})* times. A dynamic gate has only one transition during evaluation, and each output can only change once (in contrast to CMOS where sporadic changes may occur). Therefore, several charging's of the same network do not take place in dynamic operation and the gates can be sized to optimize only one transition.

Dynamic logic presents a number of design challenges, especially in modern nanoscaled technologies. The challenges are associated with drawbacks such as charge leakages, charge sharing, cross-talk sensitivity, signal integrity and restoration issues, high dynamic power consumption (at each cycle, even with clock gating), susceptibility to glitches with no data replenishment and the increased complexity of dynamic design clock distribution and control. These sensitivities and challenges are intensified by device dimensions, supply voltage scaling, process variations and temperature fluctuations. Furthermore, there are no standard design flow tools and libraries to design with dynamic families (unlike CMOS). The aforementioned challenges have significantly affected the attractiveness of dynamic logic in the last decade. In fact, dynamic logic is almost never used in modern state of the art designs. However, as will be shown in the next sections on DML, basic concepts of dynamic logic can be utilized along with other design styles to achieve fast and reliable operation.

(a) *The cascading challenge and dynamic logic topologies*

The cascading of several basic dynamic gates together using the same clock signal (*CLK*) will lead to a built-in race condition. To illustrate the race condition, two cascaded n-type dynamic inverters are shown in Fig. 6.5.

During the pre-charge phase (i.e., *CLK* = '0'), the outputs of both inverters are pre-charged to V_{DD}. On the rising edge of the clock, output out1 starts to discharge. The second output should remain in the pre-charged state of V_{DD} because its expected value is '1' (out1 transitions to '0' during evaluation). However, there is a

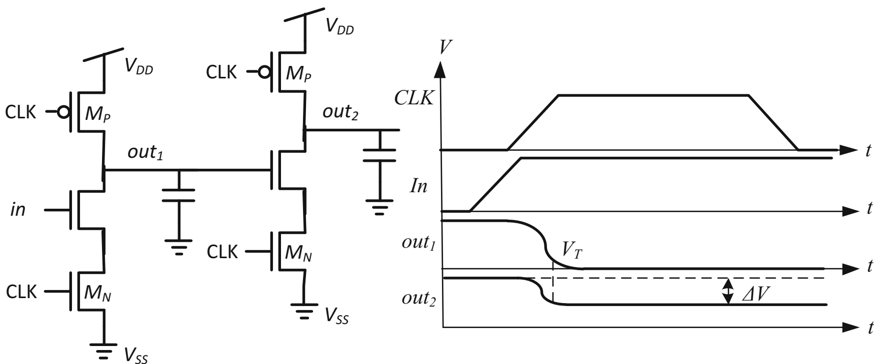


Fig. 6.5 Cascade of dynamic n-type blocks. Reproduced from [69]

finite propagation delay for the input to discharge out1 to *GND*. Thus, the second output also discharges. The conducting path is between out2 and *GND* (V_{ss}), and precious charge is lost at out2 until out1 reaches V_{THn} (the conducting path is only disabled once out1 reaches V_{THn}). This leaves out2 at an intermediate voltage level and the correct level will not be recharged. Clearly, dynamic gates rely on capacitive storage unlike static gates such as CMOS. The charge loss leads to reduced noise margins and potential malfunctioning in all dynamic families and topologies. Thus, for a correct-by-design implementation two major dynamic design families were introduced:

- Domino Logic
- NORA or np-CMOS Logic

The two proposed topologies to overcome these drawbacks are presented next.

Domino logic—this topology ensures '0' in all of the gates' output nodes immediately after the pre-charge period. The cascading issue is resolved by adding an inverter after each dynamic gate and therefore cutting off the next stage Pull-Down-Network during the pre-charge period. Figure 6.6 shows the general structure of a basic Domino gate. Figure 6.7 provides an example of a cascade connection of pre-charge high gates.

np-CMOS\ NORA Logic structure—the addition of inverters in domino logic to achieve correct logic functionality is also associated with a significant degradation in performance caused by the delay of these inverters. *np-CMOS topology* was introduced to improve performance and be able to use all of the evaluation time for necessary logic computation. This method provides an alternative approach to cascading dynamic logic by using two flavors (n-block and p-block types) of dynamic logic. In a p-block logic gate, pMOS devices are used to build a PUN, including a pMOS evaluation transistor (Fig 6.8). The nMOS is a pre-discharge transistor that drives the output low during pre-charge. The output conditionally makes a '0' → '1' transition during evaluation depending on the p-block gates. The n-block gates are controlled by CLK, and the p-block gates are controlled by

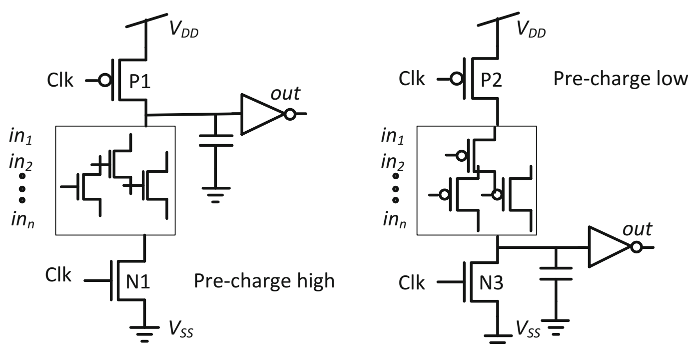


Fig. 6.6 Basic domino logic gate. Reproduced from [70]

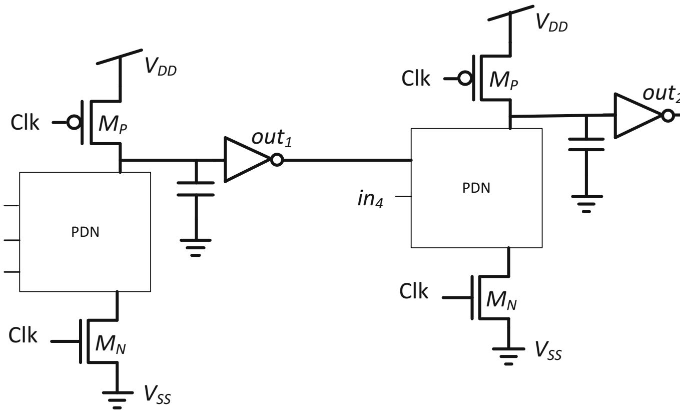


Fig. 6.7 Domino dynamic logic topology, cascaded gates

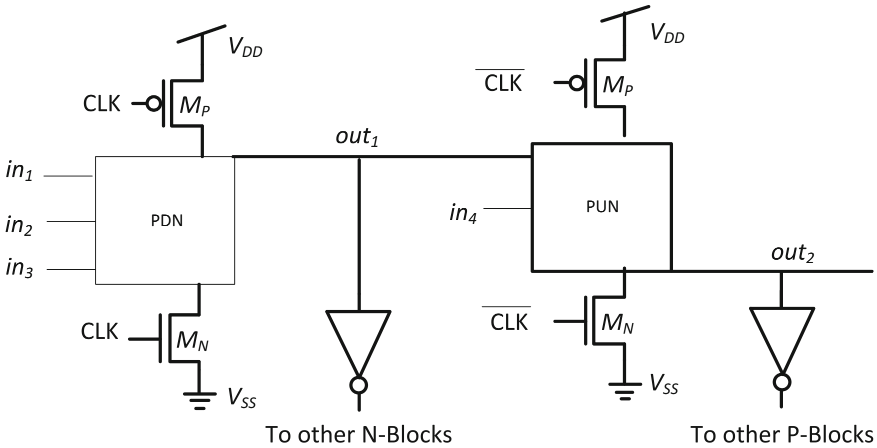


Fig. 6.8 np-CMOS dynamic logic topology

$\overline{CLK}$. n-blocks gates can directly push p-block gates, and vice versa. During the pre-charge phase (CLK = '0'), the output of the n-block gates, out1, is charged to V_{DD}, while the output of the p-block discharges to the ground ('0'). During evaluation, the output of the n-block gate can only switch via a '1' → '0' transition, conditionally turning on some transistors in the p-block. This ensures that no false glitches occur. One of the drawbacks of this logic is that half of the total amount of transistors are pMOSs. They are larger, slower (due to their low driving strength) and their ability to push current in comparison to nMOSs is smaller (due to mobility differences). Therefore, under optimization that targets propagation delays the PMOSs will consume a large area.

(b) *A Footer Implementation*

n-type dynamic gate inputs are low during pre-charge and therefore a designer may find it attractive to eliminate the evaluation transistor (footer). Elimination of the footer leads to a reduction in the sizes of the stacked PDN devices, a reduction of the clock tree distribution efforts and load and an increase in the pull-down drive strength. All these result in an improved evaluation delay. Unfortunately, this change prolongs the pre-charge period. Pre-charge will now ripple through the logic network whereas originally it ran as a parallel operation for all gates in a chain at the same time. An example of a dynamic chain without footer transistors is shown in Fig. 6.9.

An optional mitigation of the pre-charge ripple effect involves the insertion of a logic stage with a footer every few stages. This should be a reasonable compromise between the fast evaluation phase on one hand and a reasonable pre-charge phase duration on the other.

Another important negative effect of a rippling pre-charge is the extra power dissipation that occurs when both the pull-up and pull-down devices are on. In order to eliminate these short circuit currents, designers try to delay the arrival of the clock between the dynamic gates, as shown in Fig. 6.10.

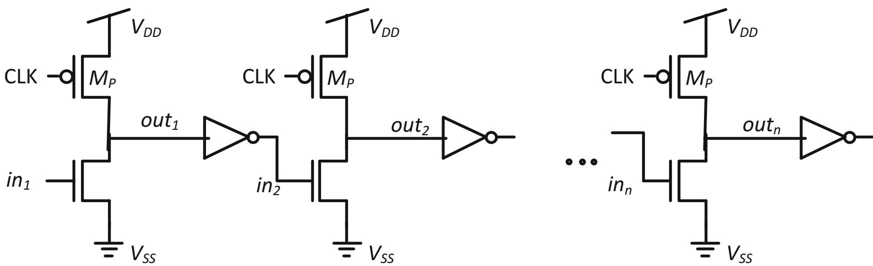


Fig. 6.9 Dynamic chain without footer transistors

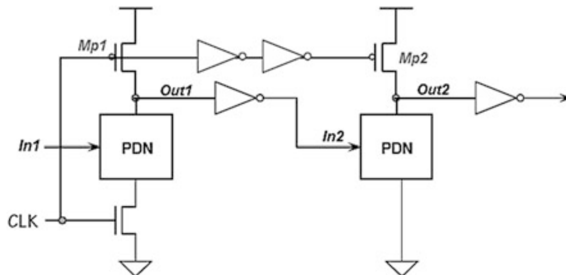


Fig. 6.10 Delay insertion for static power consumption mitigation

Thus, the existence of a footer constitutes a tradeoff between the evaluation phase duration and the subsequent pre-charge phase. Both periods form the operation cycle of a circuit, but typically the evaluation is much longer and the pre-charge phase can be concealed with multi-phasing clocks [71].

(c) *Low-Voltage Dynamic Logic*

The rationale for operating dynamic logic in the sub-threshold and near-threshold regions was to achieve an improved performance, as compared to conventional CMOS, while still dissipating significantly less energy than in the super-threshold region. Attempts to use low voltage dynamic logic were made by Soeleman et al. [72]. The conventional domino logic was used with an operation voltage below the transistor's threshold voltage. While these attempts produced very interesting results, they are less practical in modern commercial applications because of their high sensitivity to process variations in advanced technologies. The use of dynamic logic in recent versions was also abandoned because of the increased control and clock complexity. In addition, domino circuits have been shown to exhibit poor tolerance to device subthreshold leakage [73]. Another attempt to modify standard dynamic logic to allow reliable sub-V_{TH} operation was described in [74]. While showing high performance, the high-speed domino, shown in Fig 6.11, still suffered from high design complexity and area overhead.

Thus overall, dynamic logic operation with low voltage supply is very challenging, complex and presents high sensitivities to process variations. Typically, it can only be used in the sub-threshold and near-threshold designs if major modifications and extreme design efforts are made.

(3) *Other Design Styles*

This overview of existing logic families began with CMOS and dynamic logic families since they are the basic building blocks of a DML gate (as discussed in the next sub-section). For this reason, the two previous sub-chapters provided an in-depth presentation of their characteristics. However, many other design styles

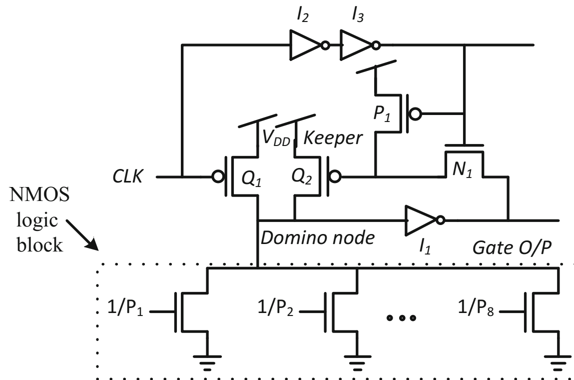


Fig. 6.11 An 8-input HS-domino OR gate. Reproduced from [74]

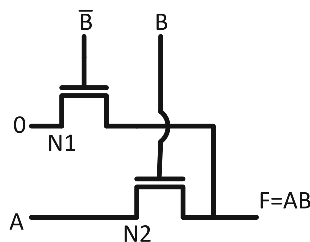


Fig. 6.12 PTL AND gate

have been suggested over the years in addition to CMOS- and dynamic- logics. In this section we provide examples of a few alternative logic families.

Pass Transistor Logic (PTL)

In contrast to CMOS, which only addresses the primary inputs that drive the gate terminals of MOSFETs, Pass Transistor Logic (PTL) logic allows the primary inputs to also drive source or drain terminals. The advantage of this family is that a one pass-transistor network (either nMOS or pMOS) is sufficient to perform a logic computation, which results in a reduced transistor count, smaller input and output loads, especially while using nMOS based PTL cells. The PTL advantages include low transistor count, small loads and, in some cases, improved delay. An example of an AND PTL gate is shown in Fig. 6.12. However, the threshold voltage drop (V_{TH}-drop) through the nMOS transistors while passing a logic "1" makes the swing (level) restoration at the gate outputs necessary to avoid static currents at the subsequent gate's output (say a CMOS).

In order to decouple gate inputs and outputs and to provide acceptable output driving capabilities, inverters are usually attached to the gate outputs. Special design emphasis should be given to the fact that in PTL, MOS networks are connected to the gate inputs rather than constant power lines. If two networks are open ("On") at the same time while one drives G_{ND} and the other drives V_{DD}, short circuit currents and power consumption will occur. This forbidden state is often called a "sneak path" and entails several logic restrictions when constructing PTL networks. Each PTL design must have a multiplexer general structure to avoid these "sneak paths". This issue is very restrictive and is one of the key obstacles to automate tools for PTL. For these reasons, certain restrictions need to be enforced when designing with PTL; i.e., two networks are required in addition to the swing restoration circuit and an additional buffer. In addition, some of the PTL circuits are need to be sized with a given device ratio (ratioed) which means that they are very sensitive to the sizing of their transistors. Hence numerous sizing restrictions are required to preserve their functionality. The swing restoration circuitry (also called a keeper) is also size-dependent and may not sufficiently mitigate the static power. Usually, PTL circuits require an increased design effort (mostly custom and not automate design).

While some PTL circuits are in use in super-threshold designs, mainly for area reduction, they were shown to be sensitive to voltage scaling [75]. Under low voltage, PTL gates acquire reduced noise margins and suffer from reliability issues with low yield.

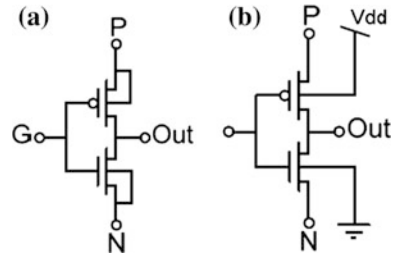


Fig. 6.13 Basic GDI cell; **a** original version. **b** standard CMOS process compatible. Reproduced from [52]

Gate Diffusion Input (GDI)
This logic family was introduced by A. Morgenstein et al. in 2002 [50, 51]. This family combines ideas from the physical structures of CMOS and PTL gates. The method is based on the use of a single simple cell, as shown in Fig 6.13a. At first glance, the basic cell is reminiscent of the standard CMOS inverter. However, the GDI cell contains three inputs: G (the common gate input of both the nMOS and the pMOS), P (the input to the source/drain of the pMOS), and N (the input to the source/drain of the nMOS). This simple cell can effectively implement a large variety of Boolean functions. The implementation of most of these functions is very complex in static CMOS. The GDI approach was originally proposed for fabrication in Silicon on Insulator (SOI) and twin-well CMOS processes. The GDI cell, shown in Fig 6.13b [52], is a modification of the basic cell depicted in Fig 6.13a, and is also compatible with standard CMOS fabrication processes. Various combinational and sequential circuits, such as adders, multipliers, comparators, Flip-Flops and counters, have been implemented in processes down to 65 nm, showing a power reduction of up to 40%.

The main disadvantage of GDI is the lack of a full swing in some logic functions and a finite input resistance. Recently, a solution to these problems was proposed by using a self-restoring (SR) transistor to regain the swing (two versions in Fig 6.14) [53, 54]. However, these proposals are still beyond the scope of automate design flows and have not been tested under low supply voltages.

Source Coupled Logic (SCL) or Current mode logic (CML)
The general structure of SCL (or CML) family gates [22, 23] was first implemented by bipolar transistors and extended to applications with MOS transistors. This Current Mode Logic (CML), dubbed MOS current mode logic (MCML), is a useful logic style for the implementation of high speed circuits. These circuits operate with a constant bias current for each gate and are suitable for markets requiring accurate high-speed mixed signal applications. The general structure of this type of gate is shown in Fig 6.15. In general, SCL circuits are differential and require a special analog layout periphery. These circuits were proposed mainly for mixed-signal applications where *dI/dt* is very important (such as D/A or A/D converters). Due to the almost constant current and their inherent small voltage swing they are highly

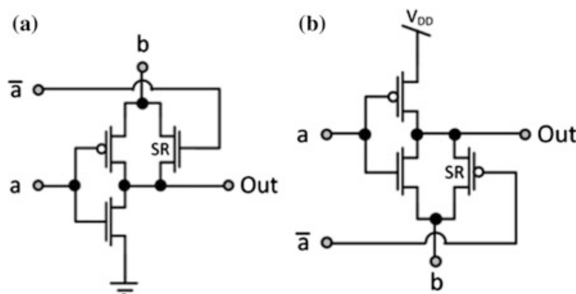


Fig. 6.14 **a** F1 with SR transistor. **b** F2 with SR transistor. Reproduced from [53, 54]

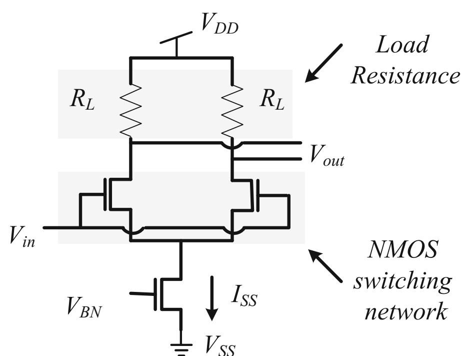


Fig. 6.15 A conventional SCL based Inverter/Buffer. Reproduced from [77]

immune to cross-talk noise incurred by switching current spikes (unlike CMOS) and are also advantageous for cryptographic circuits. While MCML dissipates constant static power (the main power dissipation source of MCML), it requires less dynamic power than conventional logic because of the smaller output swings, which are controlled and not sensitive to voltage scaling. The reduced output swing for each of the differential nodes is induced and controlled by the PMOS resistor and the saturated constant current NMOS. These are designed for a constant reduced swing for the whole circuit. The reduce swing makes the switching fast, which is the main performance boost in such circuits. Recently, Alioto et al. [76, 77] examined the ability of this logic to fit the sub-threshold regime. In this study, it was shown that the total energy consumed by a MCML circuit in sub-threshold (static power) might be smaller than the total energy (switching plus static) of a CMOS circuit. In [77] values of a 10 pA constant bias current were reported for a sub-threshold MCML circuit. While these results are very interesting and promising, many aspects of MCML implementation in large scale and advanced nodes still require study. In addition, the compatibility of the MCML to standard design flow and tools needs to be examined.

6.3 DML Basics

The DML logic gates family provides a very high level of energy delay (*E-D*) optimization flexibility [78–80]. DML allows an on-the-fly changeover between two operational modes at the gate level: the *Static* mode and the *Dynamic* mode. In the static mode, DML gates consume very low energy, with some performance degradation as compared to standard CMOS gates. Alternatively, dynamic DML gate operations obtain very high performance at the expense of increased energy dissipation. While DML gates have a very simple and intuitive structure, they require an unconventional sizing scheme to achieve the desired behavior [78–80].

In this section, we briefly present the scheme and the operating principles of a basic DML gate and discuss energy efficiency of DML in the static mode.

A. *DML overview*

A basic DML gate architecture is composed of an un-clocked static gate, e.g. CMOS, and an additional transistor *M1*, whose gate is connected to a global clock signal [78, 79]. We focus on DML gates where the static gate implementation is based on conventional CMOS. A DML gate implementation can be either *"Type A"* and *"Type B"*, as shown in Figs. 6.16a–d respectively. In the static DML mode of operation (Static mode), the M1 transistor is cut off by applying the high *Clk* signal for the *"Type A"* and the low *Clk_bar* for the *"Type B"* topology. Therefore, the gates of both topologies operate similarly to the static logic gate (CMOS in this case). For the dynamic operation of the gate (Dynamic mode), the *Clk* is enabled for toggling by providing two separate phases: pre-charge and evaluation. During the pre-charge phase, the output is charged to V_{DD} in *"Type A"* gates and discharged to G_{ND} in *"Type B"* gates. During evaluation, the output is evaluated according to the values at the gate inputs, in a way similar to NORA/np-CMOS implementations [43–45]. It was shown that DML gates exhibit very robust operation in both the static and dynamic modes under process variations (PVT) and at low supply voltages [78–80]. Dynamic mode robustness is mainly achieved by the intrinsic active restorer (pull-up in *"Type A"* and pull-down in *"Type B"*). This restorer is fairly insensitive to glitches, charge leakage and charge sharing. The unique sizing of the DML gate transistors is the key factor for achieving low energy consumption in the static DML mode (in which the topology of the gate is identical to the static gate). This sizing is also responsible for the reduction of all the gate capacitances. In a similar way, the unique transistor sizing enables evaluation through a low resistive network, thus achieving fast operation in the dynamic mode. An intuitive visualization of the tradeoff inherently related to DML is shown in Fig. 6.17. Energy efficiency is achieved in the static DML mode at the expense of slower operation (low energy and low performance). However, the dynamic mode is characterized by high performance, albeit with increased energy consumption (high energy and high performance). These tradeoffs allow a very high level of flexibility at the system level, as discussed in Sub-section B. Figure 6.18a, b show the sizing of CMOS

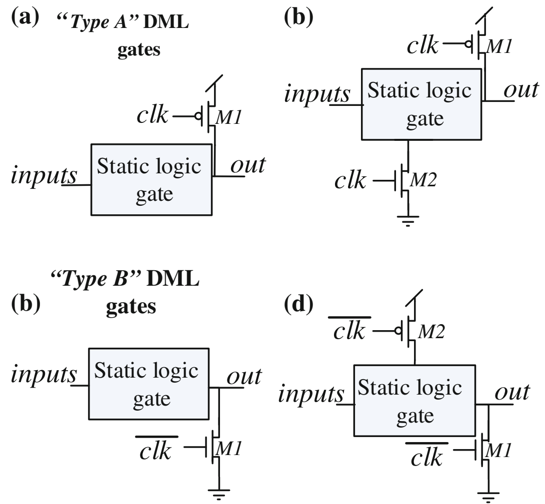


Fig. 6.16 **a** General un-footed "*Type A*" DML gate. **b** General footed "*Type A*" DML gate. **c** General un-headed "*Type B*" DML gate. **d** General headed "*Type B*" DML gate

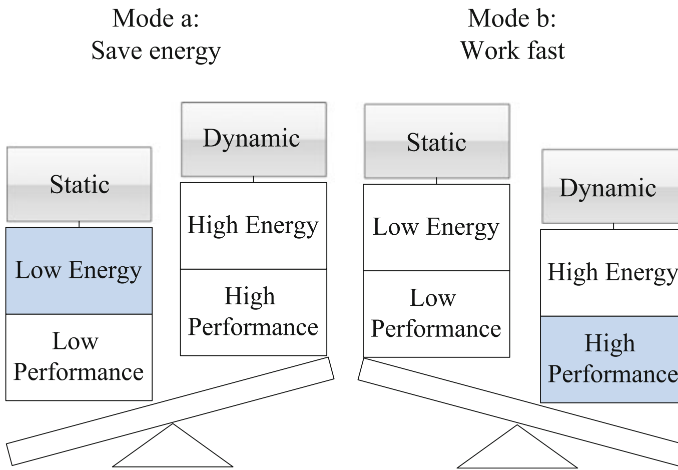


Fig. 6.17 DML gate tradeoffs in its different operating regimes

based DML gates in "*Type A*" and "*Type B*", respectively [81]. These are optimized for dynamic operation. Figure 6.18c shows the conventional sizing of a standard CMOS gate where W_{MIN} is a minimal transistor width, β is the PUN to the PDN inherent up-sizing factor and f is the gate's general up-sizing factor [80, 82, 83]. As can be seen, the in/out capacitances of DML gates are significantly reduced

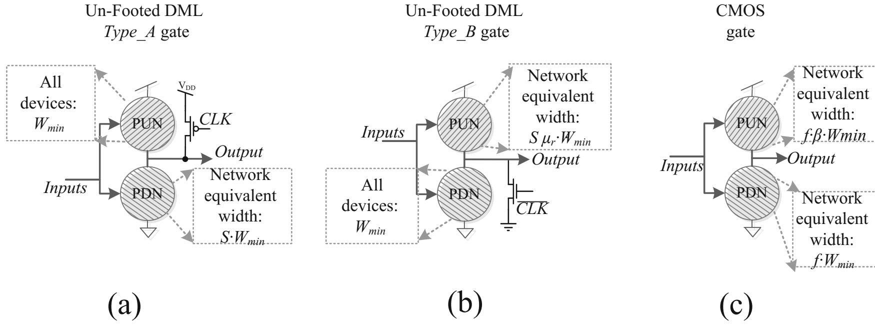


Fig. 6.18 **a** General un-footed "*Type A*" CMOS based DML gate with detailed sizing. **b** General un-headed "*Type B*" CMOS based DML gate with detailed sizing. **c** General CMOS gate with detailed sizing

compared to CMOS gates, due to the utilization of minimal width transistors in the pull-up of *"Type A"* or pull-down of *"Type B"* networks. The size of the pre-charge transistor is kept equal $S\ W_{MIN}$ to maintain a fast pre-charge period, despite the output load upsized gate, where S is the evaluation network upsizing factor. For more details, the reader is referred to [80].

Figure 6.16b, d show the footed *"Type A"* and the headed "*Type B*" DML gates, respectively. The use of these topologies is explained in detail in [80]. They allow a successful pre-charge for a cascaded topology of standard static gates/synchronous devices to a DML logic. Many aspects of DML gate sizing, as well as the preferred set of gates for "*Type A*" and "*Type B*" topologies have been analyzed and discussed. Optimization for network up-sizing parameters for load driving was carried out using the Logical Effort (LE) method [80]. The key DML achievement is that while presenting very high performance in the dynamic mode through the proposed sizing, the same topology also enables improved energy efficiency in the static mode as compared to a conventional CMOS.

B. *Static DML as a semi-energy-optimal CMOS*

The design parameter optimization of a conventional CMOS gate is mainly influenced by V_{TH}, transistor width, V_{DD}, channel length, oxide thickness and body voltage. The influence of these parameters on *E-D* plane optimization has received attention in the literature [3]. For the CMOS family, the symmetry of the gate (i.e. equal rise and fall times) is highly important. This is due to the fact that in a combinational system there is always some uncertainty regarding the transition type. As a result, the pull-up network (PUN) of CMOS gates, which is constructed by low mobility PMOS devices, is sized up by the β parameter [82]. When optimizing a CMOS gate's energy at the expense of its performance, the transistor's width is the main parameter used for reducing the energy consumption. Several factors account for this: (1) Switching energy is proportional to the load and quadratically dependent on V_{DD}. Under energy optimization, the symmetry of gate performance does not constitute a constraint, so transistor width can be reduced, as

well as β. This significantly lowers the load capacitances. (2) As a result of circuit V_{DD} lowering and technology scaling, leakage energy has become one of the key factors in static power dissipation. The leakage energy is caused by the numerous leakage currents of a device. The main leakage currents are the sub-threshold and gate leakage currents [1, 84]. These currents are linearly dependent on transistor width and under energy optimization they are considerably reduced.

CMOS based DML gates that are operated in a static mode with transistor sizes optimized for the dynamic mode are de facto a semi-energy-optimal CMOS structure with an additional negligible output capacitance for the *Clk* transistors (transistors M1 and M2). This is because when optimizing dimensions for the dynamic mode, the complementary network is composed of minimum sized gates, whereas in the static mode this minimum transistor area sizing induces minimal dynamic and static energy consumption. Static DML is still highly robust due to its complementary nature [78, 79] and withstands aggressive voltage scaling. This methodology can also be referred to as a stand-alone technique for reducing the energy consumption of digital circuits. The *E-D* tradeoff space in this approach is very broad, but this section limits the discussion to transistors sizing, such as the one shown in Fig. 6.18a, b for DML gates.

6.4 DML Utilization for Increased E-D Flexibility

The performance of most digital circuits and systems is determined by the delay of their critical paths (CP). Even though standard synthesis tools attempt to design logic blocks without CP [85–87] (i.e. equalized path delay), the slack from the targeted *Clk* (Clock) frequency always exists and needs to be remedied by designers. Many methods have been proposed to address these slacks. These methods include adaptive voltage scaling with a CP emulator circuit [81], multi oxide thickness driven threshold-voltages, multi-channel lengths for energy reduction in the non-CPs and performance boost in the CPs [88, 89]. Meijer et al. and Liu et al. applied a body bias on a non-CP to improve energy consumption and increase performance of the CPs, respectively [90, 91]. While these methods solve the critical path slack problem, in most cases they also result in a significant increase in energy consumption. In addition to these gate level approaches, higher-level approaches have been presented such as multi-mode logics, parameterized logic [92].

In [1], a concept of *E-D* space optimization as a function of different gate topology utilizations was proposed. Figure 6.19 shows that by constructing gates with different transistors sizing/topologies a different *E-D* curve can be achieved. This clearly shows the advantages of operating in MEP with one topology and in MDP with another topology. Until recently, this was only possible by constructing two different designs. The unique capabilities of the DML logic allow very high level flexibility at the gate level and actually extend the area of energy-delay optimization space. In fact, the utilization of DML makes it possible to extend the

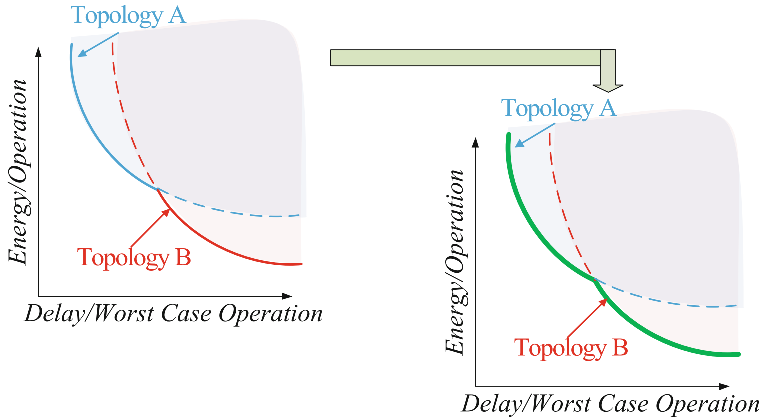


Fig. 6.19 E-D space for different topologies as a function of V_{DD}. Different topologies make it possible to achieve both MEP and MDP. Reproduced from [1]

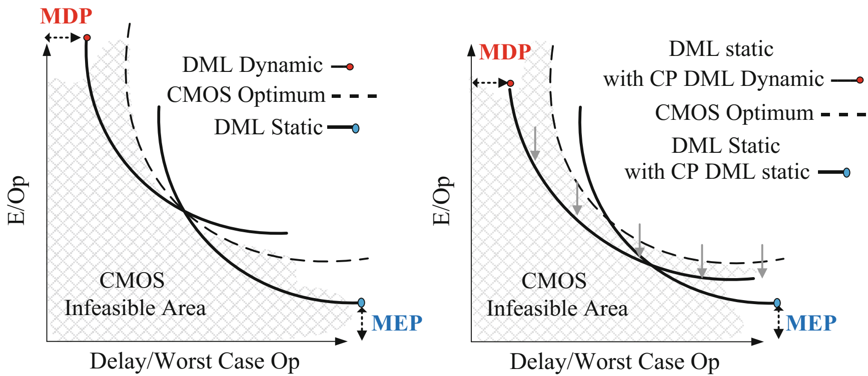


Fig. 6.20 E-D space for CMOS, DML Dynamic and DML static as a function of V_{DD}

E-D curve and operation in both MEP and MDP *with the same design* by switching from the *Static* mode to the *Dynamic* mode. This idea is schematically presented in Fig. 6.20.

In this section, we show how to utilize this flexibility to improve the energy efficiency and performance of combinatorial circuits by manipulating their critical and non-critical paths. An approach that locates the design's critical paths (CPs) and operates these paths in the boosted performance mode is described. In this approach, the non-critical paths are operated in the low energy DML mode, which does not affect the performance of the design, but affords a significant energy consumption reduction. This approach was analyzed on a 128bit Carry Skip Adder (CSA). Simulations carried out in a standard 40 nm digital CMOS process with V_{DD} = 400 mV show that this approach leads to a 2X performance improvement

along with a 2.5X reduction of energy consumption compared to a standard CMOS implementation. At $V_{DD} = 1.1$ V, improvements of 1.3X and 1.5X in performance and energy were achieved, respectively. This section describes the approach and provides a design example.

A. *Delay improvement of conventional CMOS by operating CPs in DML dynamic mode*

As mentioned above, the CPs of a CMOS design are automatically identified using standard design flow tools. By replacing these paths alone with DML gates and applying the dynamic mode on these paths, their delay can be significantly improved. The rest of the design is implemented using standard CMOS static logic. Of course, special design constraints should be enforced at all intersections between a static path and a dynamic one. In some of these cases, a footer/header should be applied [78, 79, 93, 94]. Figure 6.21 presents a in which the CPs were located and only those paths were allowed to toggle between the dynamic and static mode according to system requirements. If the system can withstand slower operation, the CP logic will operate in the static mode. If the system is required to meet the defined Clk period for all cycles, the CPs will operate in the dynamic mode. One such application is a smart phone that operates in two different modes: a slow frequency for the power-save/hibernate mode and a fast mode for video streaming. Nevertheless, it should be emphasized that low complexity systems can typically only handle one frequency for operation and the CPs will constantly operate in the dynamic mode. Normally, the number of gates on the CP is low as compared to the total number of gates in the design. Therefore, in most cases, the inherent dynamic-operation energy of these CPs will lead to a non-significant increase in the total energy consumption of the design.

B. *CP-DML Approaches to energy efficiency and High Performance*

This section elaborates on approaches for energy efficient and high performance design of combinatorial systems using DML gates. Theoretically, a general DML design can be controlled (input signal-driven control or external signal-driven

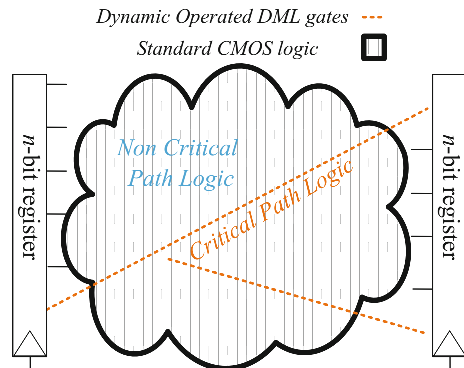


Fig. 6.21 A mapped circuit for CPs and non-CPs where the former operates in the dynamic mode and the latter in standard CMOS

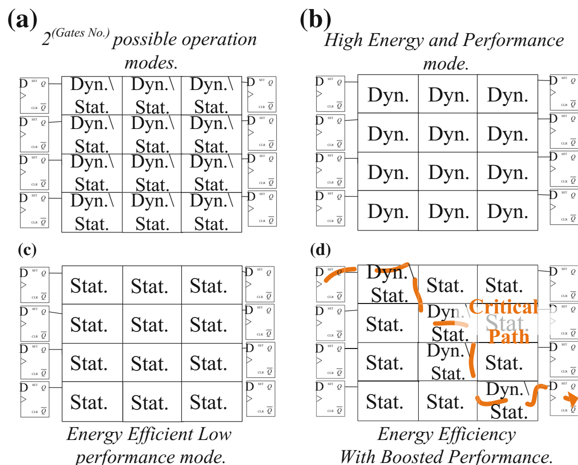


Fig. 6.22 **a** DML design optional operation modes. **b** DML design degenerated to the dynamic mode. **c** DML design degenerated to the static mode. **d** DML design where only the CPs' are dynamically operated while the rest of the design operates in the low energy static mode. Dyn. and stat. stand for dynamic and static respectively

control) to operate each gate in one of two modes: static or dynamic. This means that a general design, implemented with DML gates, can be operated in $2^{(Gates\ Number)}$ different ways, each leading to a different operating point on the *E-D* space of the design. Figure 6.22a visualizes this modularity. The degenerated approaches for operating all the gates in one of the two modes, similar to a sole gate, are shown in Fig. 6.22b, c. Switching between these two modes leads to a distinct tradeoff; in other words, the design is optimized either to achieve maximum performance or minimum energy consumption.

Figure 6.22d presents a more interesting approach, which allows both performance improvement and energy reduction. Sub-section A presented an approach to improve the delay of the conventional CMOS design where the CPs were mapped and operated in the dynamic DML mode while the rest of the circuit was assumed to preserve a standard CMOS logic gate topology. In fact, this approach was proposed to solve CP timing violations at the expense of a small degradation in energy consumption, as compared to a complete CMOS design. Here, all the components of the design that are not part of the CPs are mapped to static mode DML gates (similar to the semi-energy optimized CMOS gates described in the previous section). In most designs, these non-CPs are not time constrained and thus the asymmetric behavior of their transitions and consequently their performance degradation can withstand the *Clk* period. The use of the static DML mode for the vast majority of gates in the design leads to a significant reduction in the total dynamic and static energy consumption. Figure 6.23 visualizes this approach.

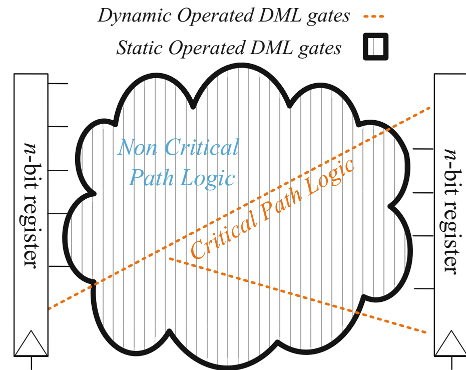


Fig. 6.23 A mapped circuit for CPs and non-CPs where the former operates in the dynamic mode and the latter in the static mode

C. *Modular benchmark*

A Carry Skip Adder (CSA, also called carry bypass adder) was chosen as a benchmark to demonstrate and evaluate the aforementioned concepts. The CP of the CSA increases as a function of the number of inputs, making it possible to examine the *E-D* trends as a function of the CP lengths. It is important to note that these methods can apply to any combinational circuit and CSA was chosen as a benchmark solely for its modularity and simplicity.

Three different scenarios are simulated and presented to demonstrate the DML CP approaches:

(1) A CP accelerator, as described in sub-section A, which has 2 operation modes:

- "DML Carry Path-Dynamic"—The DML CPs are activated in the dynamic mode.
- "DML Carry Path-Static"—The DML CPs are activated in the static mode.

In both of these modes the rest of the non-CP portions of the system are designed with standard CMOS.

(2) A CP accelerator with energy efficient non-CPs, as described in sub-section B, which has 2 operation modes:

- "DML Carry Path-Dynamic with low energy non-CPs-Static"—The DML CPs are activated in the dynamic mode, while the rest of the system operates in the DML static mode.
- "DML Carry Path- Static with low energy non-CPs- Static"—The DML CPs are activated in the DML static mode, similar to the rest of the system.

(3) CMOS equivalent design.

(1) *CMOS CSA design*

A conventional CSA is composed of a set of Ripple Carry Adder (RCA) blocks. They essentially utilize a carry propagation signal to skip the carry from one RCA to the next RCA block. It is possible to predict the propagation of the carry by a simple XOR gate [95]. This prediction mechanism can substantially reduce the delay [96]. The CP in CSA occurs when the carry ripples at the first block, and then skips the rest of the blocks and then ripples again at the last block. This is the longest possible route in the CSA. Lehman et al. researched CSAs with non-uniform sized distributed RCA blocks [97]. Majerski presented a multi-level carry-skip propagation architecture [98]. Guyot et al. and Oklobdzija et al. proposed algorithms for choosing optimized block sizes [99, 100]. Here, a simple CMOS CSA design with a fixed size of 4-bit blocks was implemented, as shown in Fig. 6.24. Clearly, the methods presented in this section can be generalized to any CSA block size constant or variable and any multi or single level carry path. The general equations to represent a single-bit Full Adder (FA) are:

$$S = A \oplus B \oplus C_{in} \tag{6.1}$$

$$C_{out} = AB + C_{in} \cdot (A + B) \tag{6.2}$$

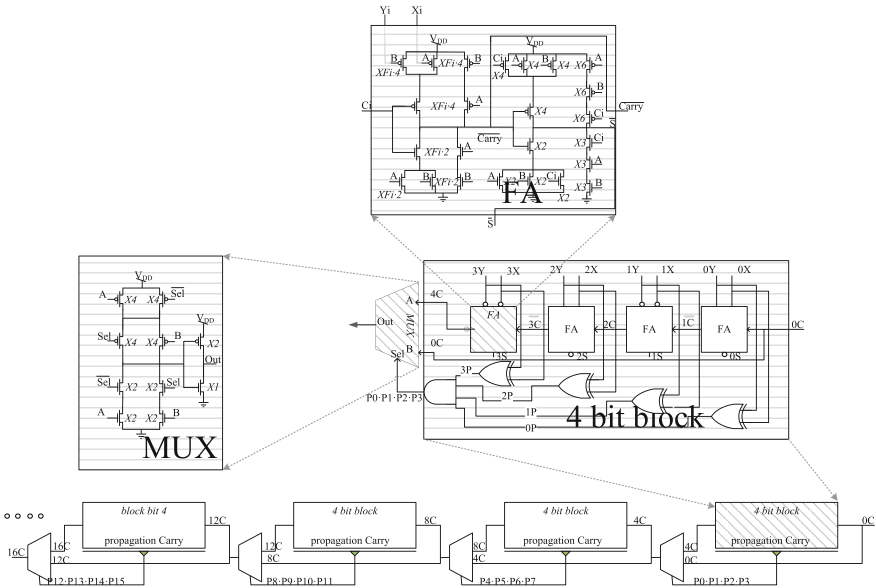


Fig. 6.24 Testbench circuit: CMOS designed CSA where the non-elementary CMOS gates such as the mirror FA and MUX are presented with transistor sizes. Transistor sizes are marked by an X 'Number' where X represents the multiplication in *Wmin*—technology minimum transistor width

$$P = A \oplus B \tag{6.3}$$

where $\oplus$ is the conventional XOR symbol. For a RCA, C_{out} is an input to the next FA. For the CP, the carry propagates through all FAs. Due to the fact that C_{out} is on the CP for each RCA, the mirror circuit for computing C_{out} is used [96], as shown in Fig. 6.25. This circuit calculates the inverted value of C_{out} and when serially chained, it reduces the circuitry on the CP (i.e., eliminates one inverter for each FA). Furthermore, the use of the mirror adders makes it obligatory to have inverting inputs for all odd FAs and inverting outputs for all even FAs [95], as shown in Fig. 6.24. All the logical gates presented in the figure are constructed with standard CMOS. A standard sizing optimization for the RCA of mirror FAs using Logical Effort [83] yields the sizing factor F_i (as shown in Fig. 6.24 for all the carry path gates) for all i's which are a multiple of 4, $F_i = 1$ and for all the rest $F_i = 3.5$.

(2) *DML Critical Path design*

Figure 6.25 shows the DML implementation of the CSA CP. The CP flows through the first NOR (assuming that the carry in of the whole design is '0') and through all the MUXs of the design. The gate level implementation of the CP can be constructed with various topologies of DML: DML NOR gates are more efficiently implemented by the *"Type A"* topologies and NAND gates by *"Type B"*, as discussed in [78, 79]. Boolean logic does not allow for an efficient implementation of a MUX with a NOR following a NAND or vice versa, which is the preferred topology for DML logic design. Therefore, in the chosen topology, the CP is

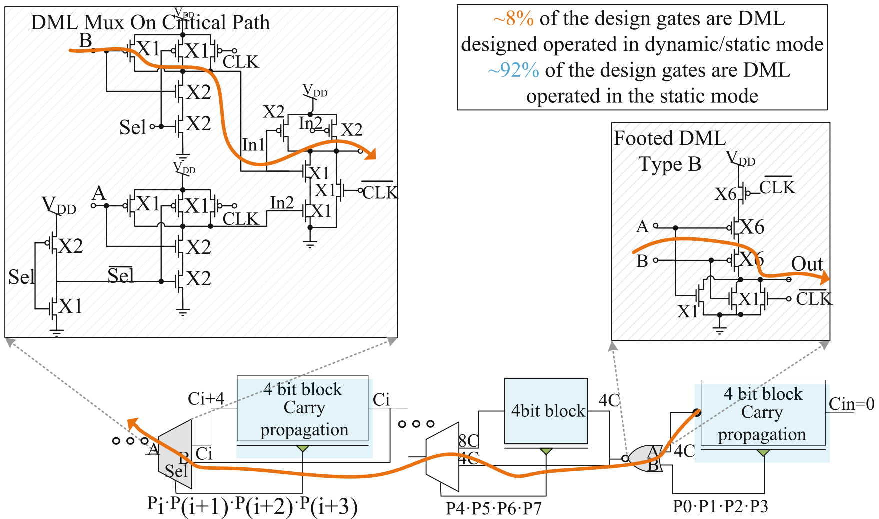


Fig. 6.25 DML critical path design for the selected CSA. The figure shows transistor widths for the gates of interest

composed solely of NANDs (where one gate is implemented using efficient *"Type B"* and the other has a less optimal *"Type A"* structure). The last inverter in each RCA block is a headed *"Type B"* inverter which maintains the correct Pre-Charge phase for the CP. The sizes of the transistors in terms of minimal transistor width are shown in Fig. 6.25. In the design, when implemented in this way, only 8% of the transistors (optionally) operate dynamically, while the remaining 92% of the transistors are kept in the low energy static mode. This modular design preserves the same complexity and the same dynamic-to-static-gates-ratio as a function of the input vector's length, *N* [bits].

D. *Simulation Results*

The modular benchmarks circuits described in the previous section were simulated in a standard 40 nm CMOS process on a Cadence Virtuoso simulator [101] environment. Implementations of these methods on the benchmark CSAs were mainly examined over the *E-D* plane and as a function of the operating voltage and the CP length. Note that the naming conventions for the different designs and operating modes can be found at the start of sub-section *C*. All energy and delay measurements are per-operation.

(1) *The E-D Plane as* $f(V_{DD})$

Each design was carefully measured as a function of the supply voltage. To demonstrate the efficiency of DML both for low voltage (sub-threshold and near-threshold) and strong inversion (super-threshold) operations, measurements were carried out with supply voltages varying from 0.4 V to 1.1 V.

The E-D curves for all designs of a 128 bit CSA are plotted in Fig. 6.26a. The order of the curves from top to bottom is: (1) CMOS, (2) CMOS design with a CP in the Dynamic DML mode, (3) CMOS design with a CP in the static DML mode, (4) low energy non-CP design with a CP in the Dynamic DML mode, and (5) Low energy non-CP design with a CP in the static DML mode. The latter two curves are presented in the gray enhanced region at the bottom of the graph. This region represents the low energy area of the E-D plane achieved by implementing all non-CPs with the low energy DML static mode (which, as described above, can also be referred to as a "low energy CMOS"). The two areas of interest are circled at the edges of Fig. 6.26a and are enlarged in Fig. 6.26b, c. Figure 6.26b shows the tradeoff area for a 400 mV operating voltage for all designs. Figure 6.26c presents that same tradeoff for 1.1 V. These two extremities clearly show that these designs are highly flexible in terms of energy consumption and performance for the whole range of voltages.

The 0.4 V low-voltage (Fig. 6.26b) results show a twofold improvement in performance of the DML enhanced CP design (second and third curves) as compared to the CMOS plot (first curve). This achievement, however, comes at the price of a 16% increase in energy consumption. If the system is such that two operational frequencies are allowed, when a low-power operation is required, the static mode (with a low frequency) can be applied, yielding a 2.5X energy

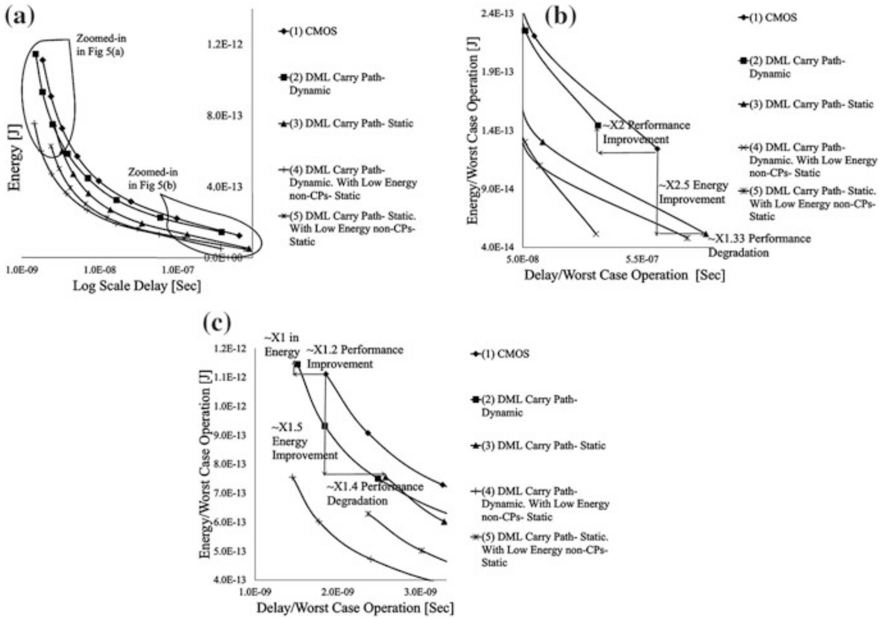


Fig. 6.26 E-D measurements for all 128 bit designs **a** E-log(D) plots for all supply voltages from 0.4 to 1.1 V. **b** Enlarged *E-D* plots for 0.4−0.5 V supply voltages. **c** Enlarged *E-D* plots for 1 −1.1 V supply voltages

improvement at the expense of a drop in performance of 1.3X. This ability to change the operating point on the *E-D* plane on-the-fly is a feature that can be easily utilized to improve system flexibility and *E-D* efficiency.

The 1.1 V high voltage supply (Fig. 6.26c), shows a CP performance boost of 20% with an energy consumption increase of only 3%. Again, if the system is such that two operational frequencies are allowed, when a low-power operation is required, the static mode can be instigated yielding a 1.5X energy improvement at the cost of a drop in performance of 1.4X. These results reveal that a low-voltage operation magnifies the differences between the different modes. There are a a few reasons for this trend. First, the performance advantage of DML circuits in the dynamic mode over standard CMOS intensifies with a lowering of the supply voltage [78, 79]. The second less dominant factor is the reduced sensitivity of DML circuits to increased leakage currents at low supply voltages [78, 79].

Inspection of the performance-optimized DML CP with low energy non-CP plots (the two lowermost curves), clarifies that the total energy is reduced two to three fold (gray region) for all voltage regions—which is substantial. In addition, the improvement in CP performance of 1.3X and 2.1X are achieved for the 1.1 V

and 400 mV supplies, respectively. The results for the CP are quite similar to the results achieved for operating without the low energy non-CP gates. This is due to the fact that the CPs themselves have not changed. Thus overall, the flexibility of the DML design leads to a significant improvement in both energy and performance.

6.5 Summary

In this chapter, we presented general energy-performance tradeoffs of digital designs and explained principles of operation in different regions. We briefly described low voltage operation in the sub-threshold and near-threshold regions and presented the advantages and challenges of low voltage circuits. Then, we briefly discussed existing logical families for digital design and elaborated on their characteristics. The recently proposed dual mode logic (DML) gate family, which enables a very high level of energy-delay optimization flexibility, was presented as an alternative to standard CMOS design. It was explained that the flexibility of DML can be utilized to dramatically improve both the energy efficiency and the performance of combinatorial circuits by manipulating their critical and noncritical paths. It was shown that by using DML logic, the performance of the 40 nm CSA benchmark circuit was improved by 2X, while also achieving a reduction in energy consumption of 2.5X. Since the CSA circuit is not optimal for DML implementations, these improvements should be even more significant for other designs.

DML logic opens up new opportunities for optimization at higher abstraction levels, such as the algorithmic and system levels, in addition to the inherent gate level opportunities which are not possible with a CMOS based design.

References

1. J.M. Rabaey, *Low power design essentials* (Springer Science and Business Media, 2009)
2. K. Roy, S. Prasad, *Low-power CMOS VLSI circuit design* (Wiley India, 2009)
3. J. Rabaey, M. Pedram, P. Landman, *Low Power Design Methodologies* (Kluwer Academic Publishers, Boston, 1995)
4. M. Keating, D. Flynn, R. Aitken, A. Gibbons, K. Shi, *Low power methodology manual for system-on-chip design* pp. 33–100
5. B. Calhoun, Y. Cao, X. Li, K. Mai, L. Pileggi, R. Rutenbar, K. Shepard, Digital circuit design challenges and opportunities in the era of nanoscale CMOS. Proc. IEEE **96**(2), 343–365 (2008)
6. B. Calhoun, J. Bolus, S. Khanna, A. Jurik, A. Weaver, T. Blalock, Sub-threshold operation and cross-hierarchy design for ultra low power wearable sensors. in *ISCAS '09*, pp. 1437–1440, May (2009)
7. M. Pedram, S. Nazarian, Thermal modeling, analysis, and management in VLSI circuits: principles and methods. Proc. IEEE **94**(8), 1487–1501 (2006)

8. A. Wang, A. Chandrakasan, A 180 mV subthreshold FFT processor using a minimum energy design methodology. IEEE J. Solid State Circuits **40**(1), 310–319 (2005)
9. L. Alarcon, T. Liu, M. Pierson, J. Rabaey, Exploring very low-energy logic: a case study. J. Low Power Electron. **3**(3), 22–233 (2007)
10. L. Benini, G. De Micheli, State assignment for low power dissipation. IEEE J. Solid-State Circuits **30**, 258–268 (1995)
11. Y. Xia, A.E.A. Almaini, Genetic algorithm based state assignment for power and area optimization. IEE Proc.Comput. Digital Tech. **149**, 128–133 (2002)
12. L. Xie, P. Qiu, Q. Qiu, Partitioned bus coding for energy reduction in, in *Proceedings of the ASP-DAC Design Automation Conference*, vol.2 (Asia and South Pacific, 2005), pp. 1280 −1283
13. B.H. Calhoun, A.P. Chandrakasan, Ultra-dynamic voltage scaling (UDVS) using sub-threshold operation and local voltage dithering. IEEE J. Solid-State Circuits **41**, 238–245 (2006)
14. B. Zhai, D. Blaauw, D. Sylvester, K. Flautner, The limit of dynamic voltage scaling and insomniac dynamic voltage scaling, very large scale integration (VLSI) systems. IEEE Trans. **13**, 1239–1252 (2005)
15. J. Shinde, S.S. Salankar, Clock gating—A power optimizing technique for VLSI circuits, in *Annual IEEE India Conference,* (INDICON, 2011), pp. 1−4
16. L. Li, W. Wang, K. Choi, S. Park, M.K. Chung, SeSCG: selective sequential clock gating for ultra-low-power multimedia mobile processor design, in *IEEE International Conference on Electro/Information Technology*, (EIT), pp. 1−6
17. W. Shen, Y. Cai, X. Hong, J. Hu, An effective gated clock tree design based on activity and register aware placement, very large scale integration (VLSI) systems. IEEE Trans. **18**, 1639–1648 (2010)
18. H. Mahmoodi, V. Tirumalashetty, M. Cooke, K. Roy, Ultra low-power clocking scheme using energy recovery and clock gating, very large scale integration (VLSI) systems. IEEE Trans. **17**, 33–44 (2009)
19. R. Bhutada, Y. Manoli, Complex clock gating with integrated clock gating logic cell, in *DTIS International Conference on Design & Technology of Integrated Systems in Nanoscale Era*, pp. 164−169 (2007)
20. P.K. Pal, R.S. Rathore, A.K. Rana, G. Saini, New low-power techniques: Leakage Feedback with Stack & Sleep Stack with Keeper, in *2010 International Conference on Computer and Communication Technology (ICCCT)*, pp. 296−301 (2010)
21. S. Dropsho, E.G. Friedman, Managing static leakage energy in microprocessor functional units, in *Proceedings 35th Annual IEEE/ACM International Symposium onMicroarchitecture, 2002.(MICRO-35)*, pp. 321–332 (2002)
22. S. Badel, Y. Leblebici, Breaking the Power-Delay Tradeoff: Design of Low-Power High-Speed MOS Current-Mode Logic Circuits Operating with Reduced Supply Voltage, in *IEEE International Symposium on Circuits and Systems, 2007. ISCAS,* pp. 1871−1874
23. A. Inoue, V.G. Dklobdzija, W.W. Walker, M. Kai, T. Izawa, A low power SOI adder using reduced-swing charge recycling circuits, in *2001 IEEE International Solid-State Circuits Conference, Digest of Technical Papers. ISSCC*. pp. 316−317, 459 (2001)
24. H. Keil, M. Momeni, A. Guntoro, A.G. Ortiz, M. Glesner, A novel leakage-estimation method for input-vector control, in *IEEE Asia Pacific Conference on Circuits and Systems, APCCAS*, pp. 570−573 (2008)
25. S. Mukhopadhyay, C. Neau, R. T. Cakici, A. Agarwal, C. H. Kim, K. Roy, Gate leakage reduction for scaled devices using transistor stacking. IEEE Trans. on Very Large Scale Integr. (VLSI) Syst. **11**(4) (2003)
26. J.P. Halter and F.N. Najm, A gate-level leakage power reduction method for ultra-low-power CMOS circuits, in *IEEE Curctom Integrated circuits conference*, pp. 475−478 (1997)
27. J. Kuo, H. Wang, A 24 GHz CMOS power amplifier using reversed body bias technique to improve linearity and power added efficiency, in *2012 IEEE MTT-S International Microwave Symposium Digest* (MTT), pp. 1−3 (2012)

28. L. Xiao, C. Liu, Y. Sun, A novel adaptive reverse body bias technique to minimize standby leakage power and compensate process and temperature variations. Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC) **2011**, 1565–1568 (2011)
29. K. K. Kim, Y. B. Kim, Optimal Body Biasing for Minimum Leakage Power in Standby Mode, in *IEEE International SymposiumCircuits and Systems, ISCAS*, pp. 1161−1164 May (2007)
30. H. Jeon, Y. B. Kim, M. Choi, A novel technique to minimize standby leakage power in nanoscale CMOS VLSI, in *I2MTC '09*, pp. 1372−1375 (2009)
31. H.S. Won, K.S. Kim, K.O. Jeong, K.T. Park, K.M. Choi, J.T. Kong, An MTCMOS design methodology and its application to mobile computing, in *ISLPED '03*, pp. 110−115 (2003)
32. Z. Liu, V. Kursun, Characterization of wake-up delay versus sleep mode power consumption and sleep/active mode transition energy overhead tradeoffs in MTCMOS circuits, in *MWSCAS '08*, pp. 362−365, Aug (2008)
33. A.P. Chandrakasan, S. Sheng, R.W. Brodersen, Low-power CMOS digital design. IEEE J. Solid-State Circuits **27**, 473–484 (1992)
34. G. Schrom, S. Selberherr, Ultra-low-power CMOS technologies. in *International Semiconductor Conference (CAS) Digest of Technical Papers*, pp. 237−246 (1996)
35. A.P. Chandrakasan, R.W. Brodersen, Minimizing power consumption in digital CMOS circuits. Proc. IEEE **83**, 498–523 (1995)
36. S. Kang, Elements of low power design for integrated systems, in *ISLPED '03*, August (2003)
37. A. Wang, B. H. Calhoun, A. Chandrakasan, *Sub-threshold design for ultra low-power systems* (Springer publishers, 2005)
38. N. Verma, A. Chandrakasan, A 256 kb 65 nm 8T subthreshold SRAM employing sense-amplifier redundancy. IEEE J. Solid State Circuits **43**(1), 141–149 (2008)
39. D. Bol, D. Kamel, D.S Flandre, J.D. Legat, Nanometer MOSFET Effects on the Minimum-Energy Point of 45 nm Subthreshold Logic, in *ISLPED '09*, pp. 3−8, Aug (2009)
40. R. Zimmermann, W. Fichtner, Low-power logic styles: CMOS vs pass-transistor logic. IEEE J. Solid-State Circuits **32**, 1079–1090 (1997)
41. M. Anis, M. Allam, and M. Elmasry, *Impact of technology scaling on CMOS logic styles*, IEEE Trans. Circuits Syst. II Analog Digital Sig. Process. **49**(8), pp. 577,588 (2002)
42. S. Hsiao, M. Tsai, C. Wen, Transistor sizing and layout merging of basic cells in pass transistor logic cell library, in *IEEE International Symposium on VLSI Design, Automation and Test, 2008. VLSI-DAT 2008*, pp. 89−92, 23−25 April (2008)
43. M.W. Allam, M.H. Anis, M.I. Elmasry, High-speed dynamic logic styles for scaled-down CMOS and MTCMOS technologies, in *2000 IEEE International Symposium on Low Power Electronics and Design, ISLPED '00*, pp. 155−160, (2000)
44. N.F. Goncalves, H. De Man, NORA: a racefree dynamic CMOS technique for pipelined logic structures. IEEE J. Solid-State Circuits **18**, 261–266 (1983)
45. H. Razak, *High performance ASIC design: using synthesizable domino logic in an ASIC flow* (Cambridge University Press, 2008)
46. R.G. Dreslinski, M. Wieckowski, D. Blaauw, D. Sylvester, T. Mudge, Near-threshold computing: reclaiming moore's law through energy efficient integrated circuits. Proc. IEEE **98**, 253–266 (2010)
47. N. Weste, K. Eshraghian, *Principles of CMOS digital design*, (Addison-Wesley, pp. 304 −307)
48. S. Hendrawan, K. Roy, B.C. Paul, Robust subthreshold logic for ultra-low power operation. IEEE Trans. Very Large Scale Integr. VLSI Syst. **9**(1), 90–99 (2001)
49. K. Yano, Y. Sasaki, K. Rikino, K. Seki, Top-down pass-transistor logic design. IEEE J. Solid-State Circuits **31**(6), pp. 792−803 (1996)
50. A. Morgenshtein, A. Fish, A. Wagner, Gate-diffusion input (GDI)-a novel power efficient method for digital circuits: a design methodology, in *Proceedings 14th Annual IEEE International ASIC/SOC Conference, 2001*, pp. 39−43 (2001)

51. A. Morgenshtein, A. Fish, I.A. Wagner, Gate-diffusion input (GDI)—a technique for low power design of digital circuits: analysis and characterization, in *ISCAS 2002 IEEE International Symposium on Circuits and Systems 2002*, **1**,pp. I–477-I–480 (2002)
52. A. Morgenshtein, I. Shwartz, and A. Fish. Gate diffusion input (GDI) logic in standard CMOS nanoscale process.in 2010 *IEEE 26th Convention of. IEEE Electrical and Electronics Engineers in Israel (IEEEI)* (2010)
53. V. Sze, A.P. Chandrakasan, A 0.4-V UWB baseband processor, in *2007 ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)*, pp. 262–267 (2007)
54. A. Morgenshtein et al., Full-swing gate diffusion input logic—Case-study of low-power CLA adder design. Integr. VLSI J. **47**(1), 62–70 (2014)
55. B. Nikolic et al., Improved sense-amplifier-based flip-flop: design and measurements. IEEE J. Solid-State Circuits **35**(6), 876–884 (2000)
56. K. Tiri, M. Akmal, I. Verbauwhede. A dynamic and differential CMOS logic with signal independent power consumption to withstand differential power analysis on smart cards. in *Proceedings of the 28th European IEEE, Solid-State Circuits Conference, 2002. ESSCIRC* (2002)
57. G. Chen, M. Fojtik, D. Kim, D. Fick, J. Park, M. Seok, M. Chen, Z. Foo, D. Sylvester, D. Blaauw, Millimeter-scale nearly perpetual sensor system with stacked battery and solar cells, in *2010 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 288–289 (2010)
58. A. Asenov, A.R. Brown, J.H. Davies, S. Kaya, G. Slavcheva, Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs. Electron Devices, IEEE Transactions on **50**, 1837–1852 (2003)
59. K. Roy, S. Mukhopadhyay, H. Mahmoodi-Meimand, Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. Proc. IEEE **91**, 305–327 (2003)
60. A. Agarwal, S. Mukhopadhyay, A. Raychowdhury, K. Roy, C.H. Kim, Leakage power analysis and reduction for nanoscale circuits. Micro IEEE **26**, 68–80 (2006)
61. J. Kao, S. Narendra, A. Chandrakasan, Subthreshold leakage modeling and reduction techniques, in *IEEE/ACM International Conference on Computer Aided Design, 2002. ICCAD 2002*, pp. 141–148 (2002)
62. B. Zhai, S. Hanson, D. Blaauw, D. Sylvester, Analysis and mitigation of variability in subthreshold design, in *Proceedings of the 2005 International Symposium on Low Power Electronics and Design, 2005. ISLPED '05*, pp. 20–25 (2005)
63. D. Bol, R. Ambroise, D. Flandre, J. Legat, Analysis and minimization of practical energy in 45 nm subthreshold logic circuits, in *IEEE International Conference on Computer Design, 2008. ICCD 2008*, pp. 294–300 (2008)
64. H. Kim, K. Roy, Ultra-low power DLMS adaptive filter for hearing aid applications. Low power electronics and design, international Symposium on **2001**, 352–357 (2001)
65. H. Kaul, M.A. Anders, S.K. Mathew, S.K. Hsu, A. Agarwal, R.K. Krishnamurthy, S. Borkar, A 320 mV 56 μW 411 GOPS/Watt ultra-low voltage motion estimation accelerator in 65 nm CMOS. IEEE J. Solid-State Circuits **44**, 107–114 (2009)
66. Y. Lin, D. Sylvester, D. Blaauw, A sub-pW timer using gate leakage for ultra low-power sub-Hz monitoring systems, in *IEEE Custom Integrated Circuits Conference, 2007. CICC '07*, pp. 397–400 (2007)
67. H. Soeleman, K. Roy, B. Paul, Robust ultra-low power sub-threshold DTMOS logic, in *Proceedings of the 2000 International Symposium on Low Power Electronics and Design, 2000. ISLPED '00*, pp. 25–30 (2000)
68. M. Penney, L. Lau, *MOS Integrated Circuits*, pp. 260–182 (1972)
69. J.M. Rabaey, A.P. Chandrakasan, B. Nikolic, *Digital integrated circuits: a design perspective*. ch. 4 (Upper Saddle River, N.J.: Pearson Education, 2003), p. 222

70. H. Soeleman, K. Roy, B. Paul, Sub-domino logic: ultra-low power dynamic sub-threshold digital logic, in *Proceedings of the 2000 International Symposium on VLSI Design, 2001*, pp. 211−214 (2001)
71. D. Harris, M.A. Horowitz, Skew-tolerant domino circuits. IEEE J. Solid-State Circuits **32**, 1702–1711 (1997)
72. H. Soeleman, K. Roy, Ultra-low power digital subthreshold logic circuits, in *Proceedings 1999 International Symposium on Low Power Electronics and Design 1999*, pp. 94−96 (1999)
73. S. Thompson, I. Young, J. Greason, M. Bohr, Dual threshold voltages and substrate bias: keys to high performance, low power, 0.1 & mu;m logic designs, in *1997 Symposium on VLSI Technology, 1997 Digest of Technical Papers*. pp. 69−70 (1997)
74. M.W. Allam, M.H. Anis, M.I. Elmasry, High-speed dynamic logic styles for scaled-down CMOS and MTCMOS technologies, in *Proceedings of the 2000 International Symposium on Low Power Electronics and Design, 2000. ISLPED '00*. pp. 155−160 (2000)
75. K. Yano, Y. Sasaki, K. Rikino, K. Seki, Top-down pass-transistor logic design. IEEE J. Solid-State Circuits **31**, 792–803 (1996)
76. M. Alioto, G. Palumbo, Design strategies for source coupled logic gates. Circuits and Systems I: Fundamental Theory and App. IEEE Trans. **50**, 640–654 (2003)
77. A. Tajalli, E.J. Brauer, Y. Leblebici, E. Vittoz, Subthreshold source-coupled logic circuits for ultra-low-power applications. IEEE J. Solid-State Circuits **43**, 1699–1710 (2008)
78. A. Kaizerman, S. Fisher, A. Fish, Subthreshold dual mode logic, very large scale integration (VLSI) systems. IEEE Trans. **21**, 979–983 (2013)
79. I. Levi, A. Kaizerman, A. Fish, Low voltage dual mode logic: model analysis and parameter extraction. Microelectronics Journal, Elsevier **44**(6), 553–560 (2013)
80. I. Levi, A. Belenky, A. Fish, Logical effort for CMOS based dual mode logic (DML) gates. IEEE Trans. VLSI Syst. **22**(5), 1042–1053 (2014)
81. M. Elgebaly, M. Sachdev, Efficient adaptive voltage scaling system through on-chip critical path emulation, in *Proceedings of the 2004 International Symposium on Low Power Electronics and Design, 2004. ISLPED '04*, pp. 375−380 (2004)
82. J. M. Rabaey, A.P. Chandrakasan, B. Nikolic, *Digital integrated circuits: a design perspective,* ch. 4, (Upper Saddle River, N.J.: Pearson Education, 2003), p. 222
83. I.E. Sutherland, B. Sproull, D. Harris, *Logical Effort—Designing Fast CMOS Circuits* (Morgan Kaufmann, San Mateo, CA, 1999)
84. S.P. Mohanty, N. Ranganathan, E. Kougianos, P. Patra, *Low-power high-level synthesis for nanoscale CMOS circuits*, (Springer, 2008)
85. T. Sasao, *Switching theory for logic synthesis*, (Kluwer Academic Publishers, 1999)
86. Y. Kukimoto, M. Berkelaar, K. Sakallah, *Static timing analysis*, *Logic Synthesis and Verification*, pp. 373−401 (2002)
87. J.J. Zasio, K.C. Choy, D.R. Parham, *Static timing analysis of semiconductor digital circuits* (1990)
88. H.L.A. Chen, E.K.W. Loo, J.B. Kuo, M.J. Syrzycki, Triple-Threshold Static Power Minimization Technique in High-Level Synthesis for Designing High-Speed Low-Power SOC Applications Using 90 nm MTCMOS Technology, in *Canadian Conference on Electrical and Computer Engineering, 2007. CCECE 2007*, pp. 1671−1674 (2007)
89. N. Sirisantana, L. Wei, K. Roy, High-performance low-power CMOS circuits using multiple channel length and multiple oxide thickness, in *Proceedings 2000 International Conference on Computer Design, 2000*, pp. 227−232 (2000)
90. M. Meijer, J.P. de Gyvez, Body-bias-driven design strategy for area-and performance-efficient CMOS circuits,very large scale integration (VLSI) systems. IEEE Trans. **20**, 42–51 (2012)
91. X. Liu, S. Mourad, Performance of submicron CMOS devices and gates with substrate biasing, in *Proceedings ISCAS 2000 Geneva. The 2000 IEEE International Symposium on Circuits and Systems, 2000*, vol. 4, pp. 9−12 (2000)

92. A. Mokhov, D. Sokolov, A. Yakovlev, Adapting asynchronous circuits to operating conditions by logic parametrization, in *2012 18th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC),* pp. 17−24 (2012)
93. I. Levi, O. Bass, A. Kaizerman, A. Belenky, A. Fish, High speed dual mode logic carry look ahead adder, in *Proceedings IEEE International Symposium on Circuits and Systems* (Seoul, Korea, May 2012) pp. 3037−3040
94. I. Levi, A. Albeck, A. Fish, S. Wimer, A low energy and high performance DM^2 adder. IEEE Trans. Circuits Syst. I Regul. Pap. **61**(11), 3175–3183 (2014)
95. I. Koren, *Computer arithmetic algorithms* (Universities Press, 2002)
96. A.T. Tran and B.M. Baas, Design of an energy-efficient 32-bit adder operating at subthreshold voltages in 45-nm CMOS, in *2010 Third International Conference on Communications and Electronics (ICCE)*, pp. 87−91 (2010)
97. M. Lehman, N. Burla, *Skip techniques for high-speed carry-propagation in binary arithmetic Units*, IRE Transactions on Electronic Computers, vol. EC-10, pp. 691−698 (1961)
98. S. Majerski, *On determination of optimal distributions of carry skips in adders*, IEEE Transactions on Electronic Computers, pp. 45−58 (1967)
99. A. Guyot, B. Hochet, J.M. Muller, A way to build efficient carry-skip adders. Comut. IEEE Trans. **100**, 1144–1152 (1987)
100. V.G. Oklobdzija and E.R. Barnes, Some optimal schemes for ALU implementation in VLSI technology, in *1985 IEEE 7th Symposium on Computer Arithmetic (ARITH)*, pp. 2−8 (1985)
101. K.S. Kundert, P. Foreword By-Gray, *The designer's guide to SPICE and SPECTRE* (Kluwer Academic Publishers, 1995)
102. D. Markovic, C.C. Wang, L.P. Alarcon, T. Liu, J.M. Rabaey, Ultralow-power design in near-threshold region. Proc. IEEE **98**, 237–252 (2010)

Chapter 7
Secure Power Management and Delivery Within Intelligent Power Networks on-Chip

Inna Partin-Vaisband and Eby G. Friedman

Abstract A critical challenge in multifunctional *heterogeneous* systems-on-chip is efficiently delivering and intelligently managing high quality dynamically controlled secure power to support power efficient and portable systems. To achieve efficient real-time multi-voltage power delivery and management, a systematic, scalable, and secure system is required. A fine grain power management framework comprising a variety of circuits, algorithms, and architectures is described in this chapter to control power routing and switching, while optimally allocating power among a variety of different power domains at run time. Stability, security, and design complexity are included within this framework.

Technology scaling has been driving the exponential growth of the semiconductor market over the last four decades, increasing integration density, speed, compactness, and functionality while decreasing power consumption and cost per function. Recently, the future of the More Moore (MM) virtuous circle has been challenged by a decrease in revenue growth and increase in the cost of research, development, and engineering (RD&E) of the worldwide semiconductor market [1, 2]. In addition, novel market segments such as intelligent transportation, innovative health care, sophisticated security systems, internet of things (IoT), and smart energy applications have emerged, requiring increasingly diverse functionality such as RF, power control, passive components, sensors/actuators, biochips, optical communication, and MEMs. These non-digital heterogeneous functionalities do not naturally scale (as shown in Fig. 7.1), yet provide significant value to the end customer. To support future societal needs, functional diversification (More-than-Moore (MtM)) has been recognized as a primary objective of semiconductor RD&E [2]. Integration of non-digital functionalities from the system board-level into system platforms such as systems-in-package (SiP) [3], systems-on-chip (SoC) [4], and TSV-based 3-D

I. Partin-Vaisband (✉) · E.G. Friedman
Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL 60607, USA
e-mail: vaisband@uic.edu

E.G. Friedman
e-mail: friedman@ece.rochester.edu

© Springer International Publishing AG 2017

G. Eisenstein and D. Bimberg (eds.), *Green Photonics and Electronics*,
NanoScience and Technology, DOI 10.1007/978-3-319-67002-7_7

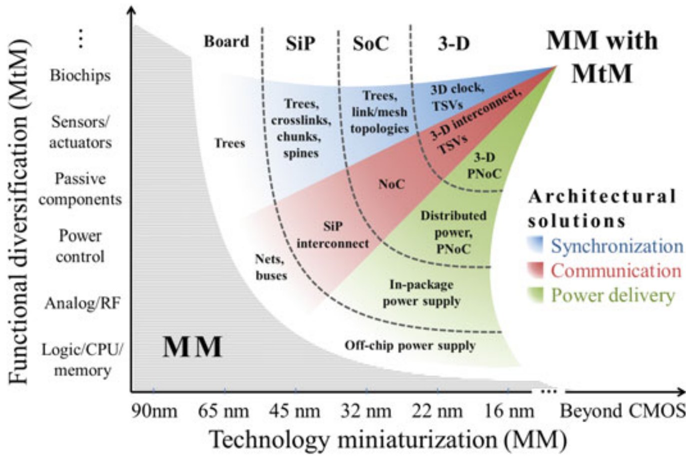


Fig. 7.1 Evolution of functional communication, synchronization, and power delivery based on functional diversification

systems [5, 6] is a primary near and long term challenge of our semiconductor-based society.

Communication, synchronization, and power delivery are three essential components of synchronous integrated circuits. To facilitate the integration of these diverse functions, architectural, circuit, device, and material level solutions are required for each of these components. The evolution of signaling, clock distribution, and power delivery is illustrated in Fig. 7.1, exhibiting a roadmap for architectural solutions that scale while providing functional diversification. The primary foci of this chapter are to provide secure, power efficient, scalable power delivery systems, and a systematic methodology for intelligent on-chip power management.

With existing techniques for leakage current analysis, the security of the on-chip information has become a growing concern. The secure delivery of high quality power to the on-chip circuitry with minimum energy loss is a fundamental requirement of modern integrated circuits (ICs). The system-wide quality of the power supply can be efficiently addressed with distributed point-of-load (POL) power delivery [7–9], and includes the on-chip integration of multiple power supplies. Distributed power delivery requires the co-design of 100s of ultra-small power converters with many thousands of decoupling capacitors and billions of nonlinear current loads within multiple power domains, significantly increasing the design complexity of existing power delivery systems. Systems with 100s of on-chip power domains and 10s of different power supply voltages have recently been reported, and thousand core ICs are being considered [10–15]. The number of transistors purposely being under utilized ("dark silicon") in future multicore ICs is expected to increase rapidly; 21% dark silicon at the 22 nm technology node, and more than 50% dark silicon at the 8 nm technology node have recently been reported [16], requiring methodologies to manage power on-chip [17–20]. Per core dynamic voltage scaling (DVS) and

dynamic voltage and frequency scaling (DVFS) are primary design techniques to efficiently manage the power budget within 100s of power domains and thousands of cores [21–23]. These systems require the on-chip integration of compact controllers, further increasing the design complexity of these power management and delivery systems.

Centralized dynamic power management has recently been used to increase power efficiency in modern heterogeneous systems [24, 25]. Centralized power delivery systems are often characterized by a long feedback path from the individual power domains to a single power management controller, dissipating additional power due to unnecessary data transport, and limiting real-time control over the energy budget due to the slow feedback response. In addition, leakage current in centralized power delivery systems can be efficiently used to evaluate the actual current flowing to the load and extract encrypted information from the system. Security attacks on power delivery systems therefore pose a significant threat on encrypted information. It has recently been demonstrated that power attacks are effective and computationally feasible for a limited number of instances under attack (for example, up to 8 bits under attack [26]). A centralized approach that places the power management functions in one or a few locations not only affects system scalability, but potentially makes the system more prone to power attacks. While distributed power delivery may be preferable for preventing power attacks, communicating energy information in a distributed power management scheme increases the vulnerability of a system to power attacks. To protect on-chip information, specialized circuits and design techniques need to be developed that are resilient to security attacks, possibly trading off performance and power consumption for system security. Alternatively, to efficiently secure on-chip information, the detection of system intrusion followed by a protective action should also be considered. To efficiently and securely manage power while delivering high quality current to complex next generation systems, enhanced methodologies, architectures, and management policies for designing scalable, intelligent power management and delivery systems with fine granularity of dynamically controlled voltage levels are required.

A general platform for delivering power is described here, which is called a power network on-chip (or PNoC), to provide systematic, scalable, and secure management of the power delivery process in heterogeneous integrated circuits. PNoC-based power delivery and management simultaneously enhances the quality of power, performance, scalability, and cost. Since power delivery and management are fundamental to all nanoelectronic systems, specialized algorithms, architectures, and approaches will greatly accelerate the development of next generation heterogeneous systems.

The rest of the chapter is organized as follows. The principles of the PNoC design methodology, architecture, and challenges are described in Sect. 7.1. Routing, stability, security, and design complexity and system scalability are discussed, respectively, in Sects. 7.2, 7.3, 7.4, and 7.5. Some concluding remarks are offered in Sect. 7.6.

7.1 Power Network on-Chip for Distributed Power Delivery and Management

Power is traditionally managed off-chip with energy efficient power converters, delivering high quality DC voltage and current to the on-chip electrical grid which reliably distributes the on-chip power [7]. The supply voltage, current density, and parasitic impedances, however, scale aggressively with each technology generation, degrading the quality of the power delivered from the off-chip power supplies to the on-chip load circuitry. To mitigate the parasitic impedances of the board, the power supply-in-package (PSiP) approach with partially off-chip yet in package power supplies has recently been considered as an intermediate power supply technology with respect to cost, complexity, and performance [27]. To lower the parasitic impedance of both the board and package, the power is regulated on-chip. To fully integrate a power converter on-chip, advanced passive components, packaging technologies, and circuit topologies are essential. Recently, several power converters suitable for on-chip integration have been fabricated [8, 9, 28–42]. Based on these power converters, a power supply system with multiple on-chip power converters is now possible.

The integration of on-chip power regulators is an important cornerstone to the power delivery process. A single on-chip power converter is however not capable of supplying sufficient, high quality regulated current to the billions of current loads within the 10s of on-chip power domains. To maintain a high quality power supply despite increasing the on-chip parasitic impedances, 100s of ultra-small power regulators should ultimately be integrated on-chip, close to the loads within the individual multiple power domains [8, 9, 18]. Integrated power converters however suffer from low power efficiency. Thus, area and power efficiency are primary concerns with a POL approach. To increase the power efficiency of a POL power delivery system, a system-wide methodology is necessary. The power is primarily converted with several efficient power converters, delivered to 10–100s of ultra-small on-chip LDOs, and regulated by those LDOs close to the points-of-load. To support DVS and DVFS, scalable power delivery systems are required that are capable of *dynamically* redistributing high quality regulated power to billions of loads in multiple, functionally diverse power domains. To satisfy power limits at each technology node, the increasing amount of “dark silicon” should be exploited [16]. To switch transistors with greater granularity in real-time, a distributed architecture and a systematic dynamically controllable power delivery and management methodology are required.

The key concept in systemizing the design of power delivery is to convert the power off-chip, in-package, and/or on-chip with multiple power efficient but large switching power supplies, deliver the power to on-chip voltage clusters, and regulate the power with 100s of linear low dropout (LDO) regulators at the point-of-load [43]. The principles of the PNoC design methodology are described in Sect. 7.1.1. In Sect. 7.1.2, the performance of the PNoC architecture is compared with existing approaches based on the evaluation of several test cases. Challenges of distributed power delivery and management are discussed in Sect. 7.1.3.

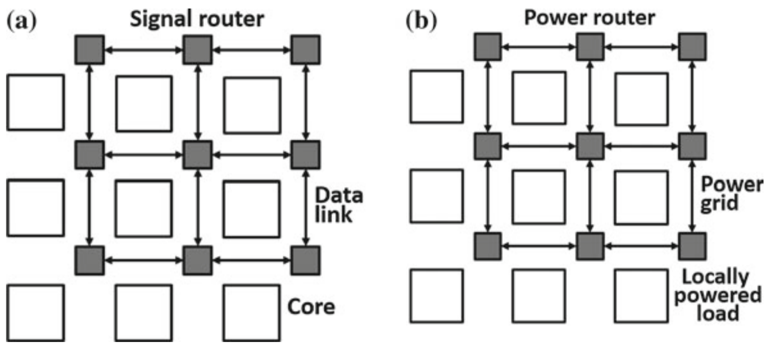


Fig. 7.2 On-chip networks **a** network-on-chip (NoC), and **b** power network-on-chip (PNoC)

7.1.1 Concept of Power Network-on-Chip

The concept of a power network on-chip is introduced here as a vehicle for distributed on-chip power management. The analogy between a NoC and PNoC is illustrated in Fig. 7.2 with simplified NoC and PNoC models. An NoC separates computation from communication, enhancing the performance, scalability, and control of quality of service (QoS), while increasing engineering productivity in the development of heterogeneous integrated systems. Similar to a network-on-chip, the PNoC approach provides a scalable platform for efficient power delivery and management. In addition, enhanced control over the quality of power (QoP), real-time voltage scaling, and integration of emerging technologies while mitigating invasive and non-invasive power attacks within the proposed framework are possible with PNoC.

7.1.2 Power Network-on-Chip Architecture

SoC platforms will consist of 100s of heterogeneous hardware agents (application processors, DSP cores, fixed function accelerators, sensors, and actuators), each with unique power, performance, and QoS requirements [4, 44, 45]. While the quality of the power supply can be efficiently addressed with a point-of-load power delivery system [9], distributed power delivery is impractical with existing *ad hoc* approaches. Maintaining a secure and stable power delivery system is an additional concern. To cope with the design complexity of distributed on-chip power delivery, a novel and scalable power delivery platform is required that relies on the reuse of components, secure architectures, secure methodologies, and applications. A power network on-chip utilizes a specific architecture and design methodology to achieve intelligent power management and scalable, secure power delivery in heterogeneous ICs.

Centralized dynamic power management is commonly used to increase power efficiency in modern heterogeneous systems. A typical power delivery scheme is

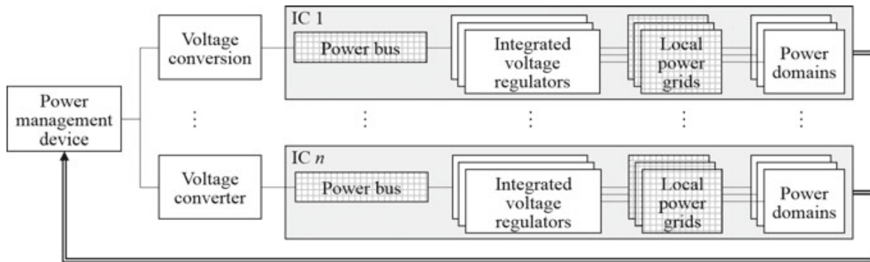


Fig. 7.3 Centralized power management of SoC with n modules

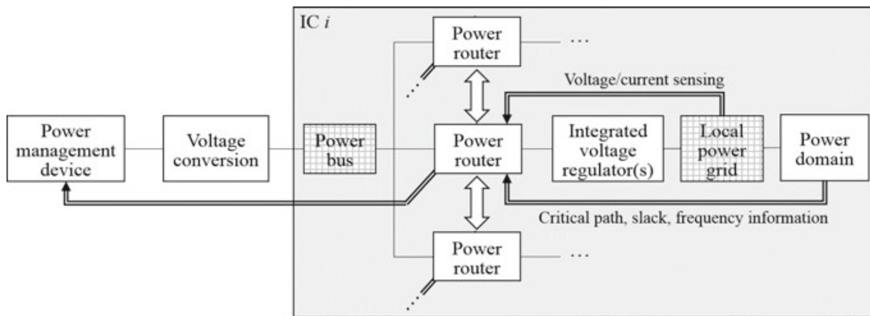


Fig. 7.4 Distributed PNoC scheme for on-chip system-wide power management

illustrated in Fig. 7.3 for an SoC with multiple power domains within each of n individual modules. In this configuration, the power is managed by a centralized power management system. These centralized power delivery systems are often characterized by a long feedback path from the individual power domains to a single power management controller, dissipating extra power due to unnecessary data transport, and limiting real-time control over the energy budget due to the slow feedback response. In addition, a centralized approach places all of the power management functions in one or a few locations, affecting system scalability.

A distributed power delivery system based on the PNoC approach is shown in Fig. 7.4. In this scheme, the energy budget is managed locally with intelligent distributed power routers, considering essential information such as timing slacks, voltages, currents, and temperatures that are sensed within the individual power domains. Real-time fine grain power management is possible by shortening the power domain-to-router feedback path (as compared to the centralized approach). Power management adjustments therefore redistribute the overall energy budget in a near real-time manner by communicating local power management decisions among the individual power routers within a PNoC. A centralized power management mechanism can also be integrated to support predetermined DVS and DVFS operations. The management functions in this configuration are coordinated with specialized algorithms and controllers among the intelligent, distributed, and centralized power management mechanisms.

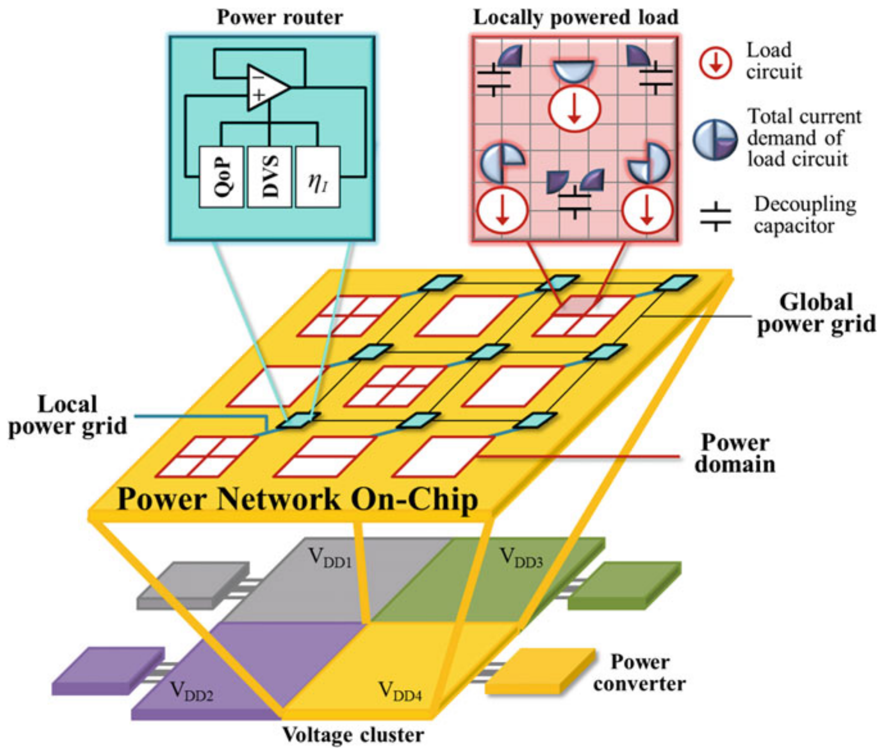


Fig. 7.5 Power network on-chip with power routers distributing current throughout the power grid to the local current loads

A PNoC composed of power routers connected to global power grids and locally powered loads is shown in Fig. 7.5. The global power from the converters is managed by the power routers, delivered to power domains, and regulated within locally powered loads. A locally powered load combines all of the current loads and decoupling capacitors physically located within a specific on-chip region within a power domain. This system senses the voltages and currents throughout the system, and manages the POL regulators with power switches. Based on these sensed voltages and currents, a programmable controller makes real-time decisions to apply a new set of configurations to the routers per time slot, dynamically managing the on-chip power delivery process. The power delivery policies are dynamically customized with specialized algorithms through a microcontroller that routes the power. These algorithms satisfy real-time power and performance requirements. Functionally diverse heterogeneous systems require additional degrees of power control and management of resources. In these complex heterogeneous systems, each power router is connected to a router-to-router communication channel to dynamically manage the shared energy budget of all of the power domains within a voltage cluster, exploiting the built-in modularity and local intelligence of PNoC to support scalable power delivery systems.

7.1.3 Challenges in Distributed Power Delivery

The power in PNoC based SoCs is virtually managed through specialized power routers, switches, and programmable control logic, while supporting scalable power delivery in heterogeneous ICs. A PNoC is comprised of physical links and routers that provide both virtual and physical power routing. This system senses the voltages and currents throughout the system, and manages the POL regulators through power switches. Based on the sensed voltages and currents, a programmable unit makes real-time decisions to apply a new set of configurations to the routers per time slot, dynamically managing the on-chip power delivery process. The challenge of efficiently routing the power in real-time is discussed in Sect. 7.2, and a power routing approach is evaluated.

A distributed system with multiple parallel connected LDO regulators delivering power to a single grid exhibits degraded stability due to complex interactions among the LDO regulators, power distribution network, and current loads. Thus, the stability of these parallel connected voltage regulators is an additional concern and requires accurate evaluation. To design a stable closed loop regulator, sufficient phase margin in the open loop transfer function is required. Phase margin is a necessary and sufficient parameter for determining the stability of a single LDO. Evaluating the open loop characteristics is however not practical with parallel LDO regulators due to the multiple regulation loops. Evaluating the stability of a distributed power delivery system is therefore not possible with the traditional phase margin criterion. An alternative stability criterion is proposed in Sect. 7.3 for evaluating the stability of parallel voltage regulators driving a single power grid.

Securing information in modern ICs is another significant challenge that needs to be addressed. Non-invasive and invasive power attack methods [26, 46–48] are widely used today to decode encrypted information or reduce the performance and/or lifetime of computing devices. A common countermeasure for power analysis is masking the effective power consumption of an IC by injecting additional power or modifying power profiles with DVS techniques [49]. These approaches however degrade the overall power efficiency of the system. Alternatively, securing information and performance with traditional algorithmic encryption techniques lack scalability and often do not exhibit sufficient security [50]. To efficiently mitigate power analysis attacks, specialized design flows and circuits that disrupt the correlation between the internal leakage power and overall power consumption are necessary. Accurate evaluation of power efficiency and security levels is required. Secure on-chip power networks are currently under development to efficiently manage power in modern ICs. The sensitivity of PNoC to power attacks is discussed in Sect. 7.4.

The complexity and scalability of a distributed nonlinear dynamically controllable POL power supply system is a significant design issue. Automated modeling, optimization, and synthesis techniques for complex analog circuits have recently been considered [51], and automated tools for analog circuits with 100s of components have become commercially available [51]. To automate the design of a power

delivery system, accurate methods to evaluate performance metrics (e.g., quality of transient response, stability, and power) are required. Automating the design of power delivery system is discussed in Sect. 7.5.

7.2 Power Routing in SoCs

The PNoC architecture is a mesh of power routers and locally powered loads, as depicted in Fig. 7.2. The power routers are connected through power switches, distributing current via power grids to those local loads with similar voltage requirements. A PNoC is illustrated in Fig. 7.6 for a single voltage cluster with nine locally powered loads and three different supply voltages, $V_{DD,1}$, $V_{DD,2}$, and $V_{DD,3}$. The power network configuration is shown at two different times, t_1 and t_2.

The power routers, local current loads, and power grids are described, respectively, in Sects. 7.2.1, 7.2.2, and 7.2.3. Different PNoC topologies and specific design objectives are also considered. A PNoC with four power routers is evaluated in Sect. 7.2.4.

7.2.1 Power Routers

The efficient management of the energy budget is dynamically maintained by the power routers. Each power domain is controlled by a single power router. A router topology ranges from a simple linear voltage regulator, shown in Fig. 7.7a, to a complex power delivery system, as depicted in Fig. 7.7b, with sensors, dynamically

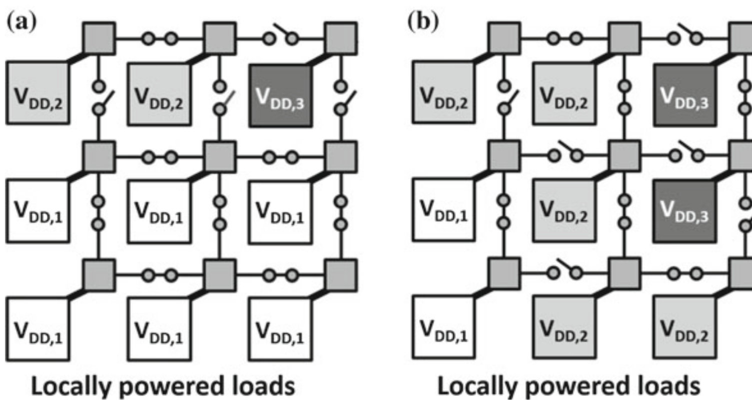


Fig. 7.6 On-chip power network with multiple locally powered loads and three supply voltage levels **a** PNoC configuration at time t_1, and **b** PNoC configuration at time t_2

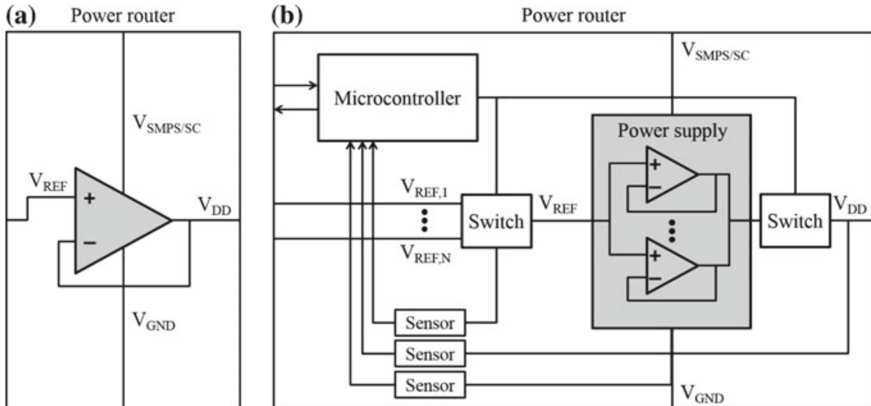


Fig. 7.7 Power routers for PNoC **a** Simple topology with linear voltage regulator. **b** Advanced topology with dynamically adaptable voltage regulator and microcontroller

adaptable power supplies, switches, and a microcontroller. This structure features real-time voltage/frequency scaling, adaptable energy allocation, and precise control over the on-chip QoP. With PNoC routers, the power is managed locally based on specific local current and voltage demands, decreasing the dependence on the remotely located loads and power supplies. Thus, the scalability of the power delivery process is greatly enhanced with the PNoC approach.

7.2.2 Locally Powered Loads

Locally powered loads with different current demands and power budgets can be efficiently managed with PNoC. The local power grids provide a specific voltage to the nearby load circuits. The highly complex interactions among the multiple power supplies, decoupling capacitors, and load circuits are important, where the interactions among the nearby components are typically more significant. The effective region for a point-of-load power supply is the overlap of the effective regions of the surrounding decoupling capacitors [9, 52]. Loads within the same effective region are combined into a single equivalent locally powered load regulated by a dedicated LDO. All of the LDO regulators within a power domain are controlled by a single power router.

7.2.3 Power Grid

A shared power grid with multiple LDO sources and dedicated power grids each driven by a separate LDO is considered here. These topologies are illustrated in Fig. 7.8 with two power routers, and shared and dedicated local power grids. The

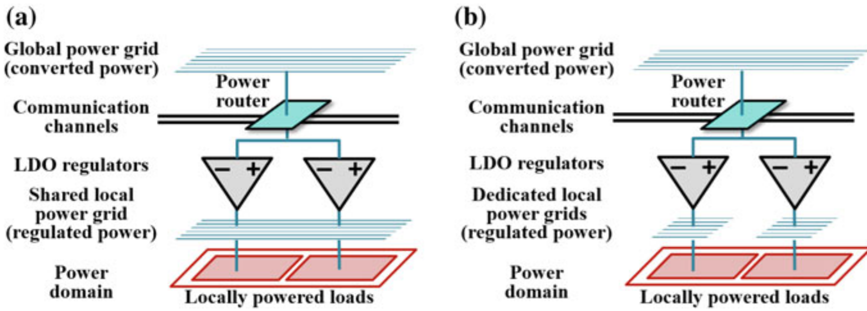


Fig. 7.8 PNoC power router with two locally powered loads, and **a** shared local power grid, and **b** dedicated local power grids

dedicated power grids require fine grain distribution of the local power current. Alternatively, a shared power grid with multiple LDO sources is prone to stability issues due to current sharing, and process, voltage, and temperature variations. Specialized adaptive mechanisms are included within the power routers to stabilize a power delivery system within a multi-source shared power grid.

7.2.4 Case Study

To evaluate the performance of a power router, a PNoC with four power routers is considered, supplying power to four power domains. IBM power grid benchmarks [53] model the behavior of the individual power domains. To simulate a dynamic power supply in PNoC, the original IBM voltage profiles are scaled to generate target power supplies between 0.5 and 0.8 V. Target voltage profiles with four voltage levels (0.8, 0.75, 0.7, and 0.65 V) within a PNoC are illustrated in Fig. 7.9. The number of power domains with each of the four power supply voltages changes dynamically based on the transient power requirements of the power domains.

A schematic of a PNoC with four power domains (I, II, III, and IV) and four power routers (PR_I, PR_{II}, PR_{III}, and PR_{IV}) is presented in Fig. 7.10. Each of the power routers is composed of an LDO with four switch controlled reference voltages to support dynamic voltage scaling. In addition, the power routers feature adaptive RC compensation and current boost networks controlled by load sensors to provide quality of power control and optimization, as shown in Fig. 7.11. The adaptive RC compensation network is comprised of a capacitive block connected in series with two resistive blocks, all digitally controlled. These RC impedances are digitally configured to stabilize the LDO within the power routers under a wide range of process variations. The current boost circuit is composed of a sensor block that follows the output voltage at the drain of transistor M_P (see Fig. 7.11), and a current boost block that controls the current through the differential pair within the LDO. When a high

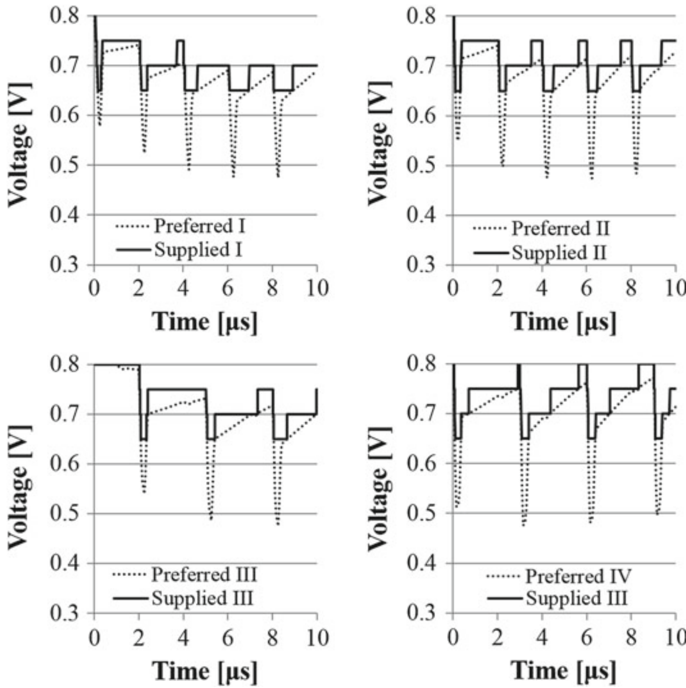


Fig. 7.9 Preferred and supplied voltage levels in PNoC with four power domains

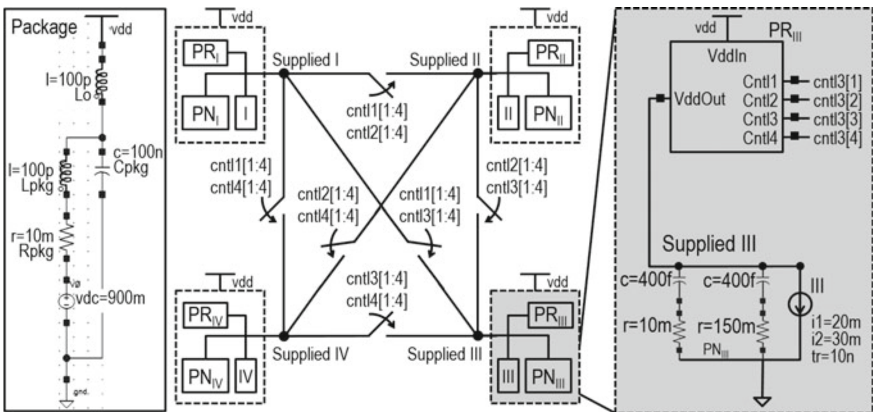


Fig. 7.10 PNoC with four power domains and four power routers connected with control switches

slew rate transition at the output of the LDO occurs, the boost mode is activated, raising the tail current of the differential pair within the LDO regulator. Alternatively, during regular operation, no additional current flows into the differential pair,

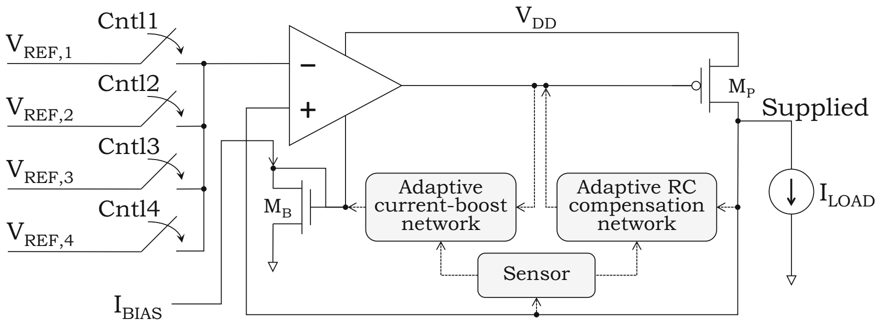


Fig. 7.11 Power router with voltage regulator, load sensor, and adaptive networks

Table 7.1 Output load in PNoC with four power domains

Domain	I	II	III	IV
Minimum output current (mA)	10	75	20	20
Maximum output current (mA)	50	10	30	20
Output transition time (ns)	50	50	10	N/A

enhancing the power efficiency of the LDO. A description of some configurations of the load sensor and adaptive networks can be found in [18].

The power routers are connected with controlled switches to mitigate load transitions in domains with similar supply voltages. To model the RLC parasitic impedances of the package and power network, non-ideal LDO input and output impedances (PN_I, PN_{II}, PN_{III}, and PN_{IV}) are considered, as shown in Fig. 7.10. The load current characteristics are listed in Table 7.1 for each of the four domains. The PNoC is simulated in SPICE. The simulation results are presented in Fig. 7.12, exhibiting a maximum error of 0.35%, 2.0%, and 2.7% for, respectively, the steady state dropout voltage, load regulation due to the output current switching, and load regulation due to the dynamic PNoC reconfigurations. Good correlation with the required power supply in Fig. 7.9 is demonstrated. The power savings in each of the power domains range between 21.0 and 31.6% as compared to a system without dynamic voltage scaling. These average power savings demonstrate that the PNoC architecture can control power supplies in real-time, optimizing the power efficiency of the overall power delivery system [54–56].

7.3 Stable Distributed Power Delivery Systems

Understanding the effects of the frequency domain parameters on the time domain characteristics provides significant insight into the transient behavior of complex systems [57, 58]. Traditionally, the phase margin of the output response determines the

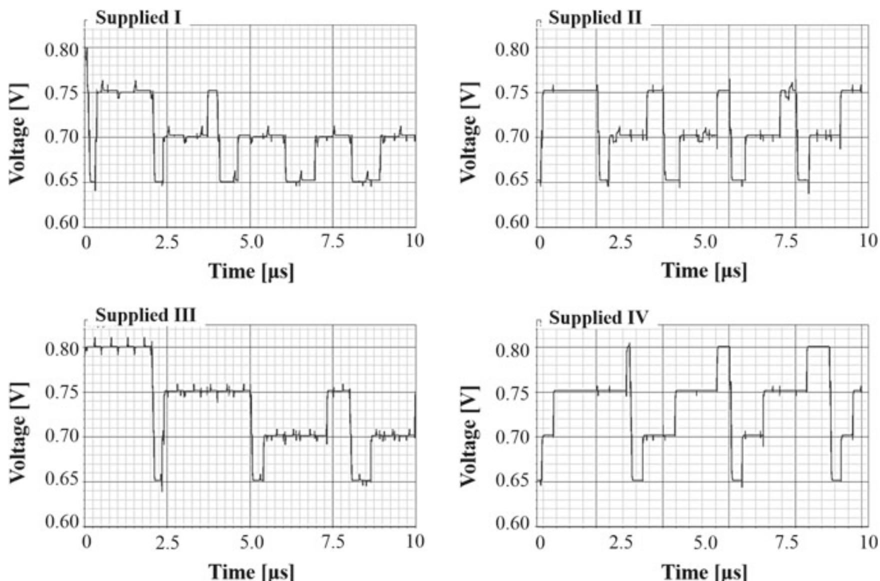


Fig. 7.12 Voltage levels in PNoC with four power domains

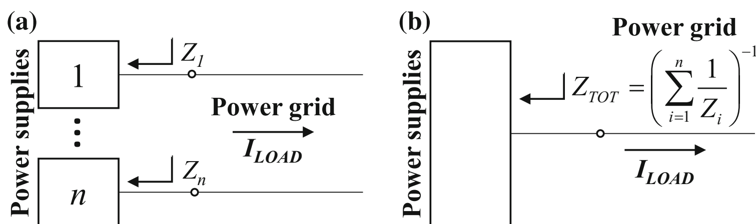


Fig. 7.13 Power delivery system **a** with two or more distributed power supplies, and **b** reduced single port network

stability of a single LDO regulator. Similarly, a criterion is required for determining the stability of a *distributed* power delivery system.

A distributed power delivery system with two or more power supplies driving a single power grid is depicted in Fig. 7.13a. In this distributed system, the power supplies can be combined into a single power delivery system, yielding an equivalent single port network, as shown in Fig. 7.13b. Note that the output impedance of the equivalent single port network, Z_{TOT} in Fig. 7.13b, is the parallel combination of all of the output impedances $Z_i, i = 1, \ldots, N$ of the individual power supplies shown in Fig. 7.13a. The output impedance of a distributed power delivery system is based on the individual output impedance of the parallel connected elements. Alternatively, a method is needed to identify those loops that cause instability in a system with multiple interacting feedback paths. The open loop transfer function, traditionally used to determine the stability of a lumped power delivery system, cannot be applied

to a distributed power delivery system with multiple control loops [59]. A criterion for evaluating the stability of a multi-feedback path system composed of distributed power regulators is therefore needed.

Necessary and sufficient conditions for a stable distributed power delivery system are described in this section. These conditions are based on the observation, proven in [60], that a linear, time-invariant (LTI) system is stable when coupled to an arbitrary passive environment if and only if the driving point impedance is a passive system. A distributed power delivery system is therefore stable if and only if the equivalent output impedance Z_{TOT} satisfies passivity requirements. The passivity of an LTI system is described here in terms of frequency domain parameters.

An LTI system is passive if the system can only absorb energy, yielding, in mathematical terms [61],

$$\int_{-\infty}^{T} v(t)i(t)dt \geq 0, \forall\, T, \tag{7.1}$$

where $v(t)$ and $i(t)$ are, respectively, the voltage across the system and the current flowing through the system. The total energy delivered to a passive system is determined from (7.1) based on the Parseval Theorem, exhibiting, for all positive currents,

$$\frac{1}{\pi}\int_{0}^{+\infty} Re[Z(j\omega)]|I(j\omega)|^2 d\omega \geq 0, \tag{7.2}$$

where $Z(s) = V(s)/I(s)$ is the system impedance, and $V(s)$ and $I(s)$ are, respectively, the phasor voltage and current of the system. The passivity condition based on (7.2), $\{Re[Z(\sigma + j\omega)] \geq 0, \forall \sigma > 0\}$, can be simplified based on [62], yielding the following sufficient conditions for passivity of an LTI system:

- $Z(\omega)$ has no right half plane (RHP) poles.
- The phase of $Z(\omega)$ is within the $(-90°, 90°)$ range.

A distributed system is, therefore, exponentially stable (converges within an exponential envelope) if and only if the impedance of the system satisfies these passivity requirements, marginally stable (oscillates with constant amplitude) if and only if the voltage and current phasors are shifted by precisely 90°, and unstable otherwise. The phase of the output impedance is an efficient alternative to determine the stability of these distributed systems, since the traditional phase margin approach is not practical due to the multiple control loops.

7.3.1 Experimental Evaluation of Stability Criterion

To demonstrate the concept of a stable distributed power delivery system based on the passivity-based stability criterion, a distributed power delivery system with six

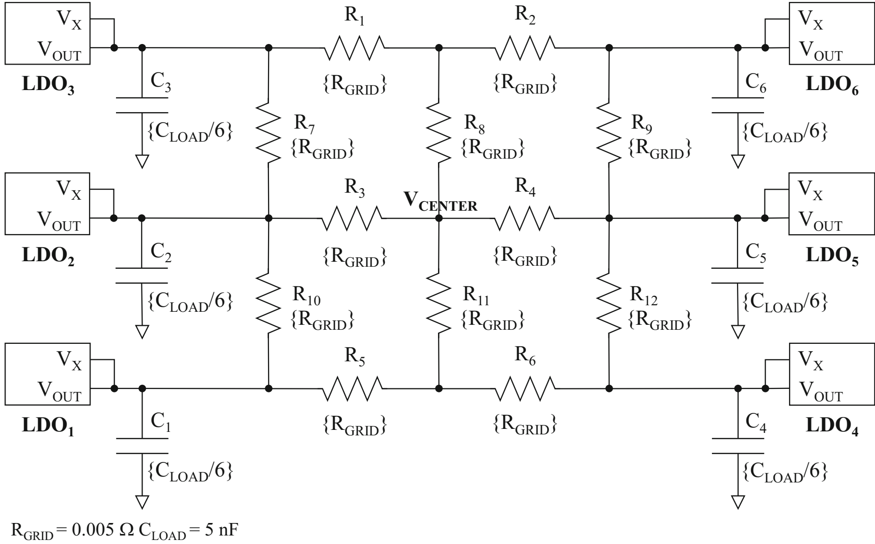


Fig. 7.14 Model of distributed LDO and power distribution network

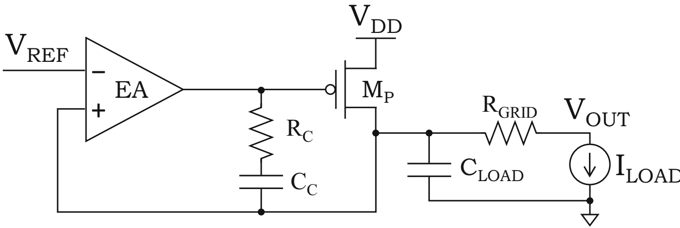


Fig. 7.15 LDO topology [59]

LDO regulators is evaluated. A model of the power delivery system with the six LDO regulators and a distributed power delivery network is shown in Fig. 7.14. Each power supply in the power delivery system is an LDO regulator [63] composed of an error amplifier (EA), output device (M_P), and compensation network $R_C C_C$, as depicted in Fig. 7.15. A three current mirror operational transconductance amplfier (OTA) topology [59] is used within each error amplifier.

The output impedance of the parallel connected voltage regulators is a primary factor in determining the stability of a distributed power delivery system, and can be characterized by the poles and zeros of the individual LDO regulators. To maintain stability in a distributed power delivery system with n LDO regulators, the poles of the output impedance $Z_{OUT}^{TOT}(s)$ must all be left plane poles and the phase of $Z_{OUT}^{TOT}(s)$ must be within the $(-90°, 90°)$ range $\forall\ s$.

The stability of the proposed power delivery system is demonstrated on an example system assuming a total current load of 300 mA. Load sharing among the LDO

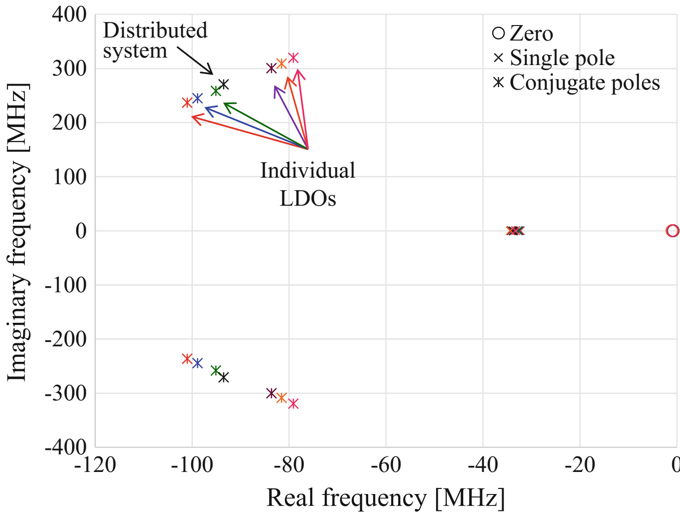


Fig. 7.16 Poles and zeros of output impedance of individual LDO regulators loaded by different currents (between 20 and 100 mA), and the poles and zeros of the system output impedance

regulators in the system exhibits a wide range of LDO currents (between 20 and 100 mA for an individual LDO regulator). The LDO in closest proximity with the current load supplies the largest portion (100 mA) of the total current requirements, which is higher by a factor of two than the average current load (52 mA) supplied by a single LDO. Alternatively, remote LDOs supply significantly less current (down to 20 mA), only half of the average LDO load current. The output impedance of the system under this load sharing scenario is evaluated for each of the LDO regulators and the overall distributed power delivery system. The poles and zeros within the range of interest are shown in Fig. 7.16. Note that the poles of the system output impedance are limited by the frequency range of the individual LDO poles. Thus, a distributed power delivery system with individually stable LDO regulators *under all feasible load currents* exhibits no right half plane poles (RHP). The stability of a multi-feedback system with individually stable power supplies is therefore limited by the phase of the system output impedance. To demonstrate the effect of the phase of the output impedance on the stability of a distributed system, the transient response and output impedance phase $\angle Z_{OUT}$ of the distributed system with six LDO regulators is shown in Fig. 7.17. In agreement with the passivity-based stability criterion, the output response diverges (oscillates with increasing amplitude), and converges within an exponential envelope for, respectively, $|\angle Z_{OUT}| > 90°$ and $|\angle Z_{OUT}| < 90°$. Note that the system with $C_C = 0.5$ pF and $\max_{\forall f}\{\angle Z_{OUT}\} = 89°$ slowly converges to the steady-state solution, exhibiting an inappropriate underdamped response to effectively regulate the voltage in a power delivery system. Alternatively, a system with $C_C = 5$ pF and $\max_{\forall f}\{\angle Z_{OUT}\} = 70°$ exhibits an overdamped response with a significant margin of stability. A strong correlation therefore exists between the

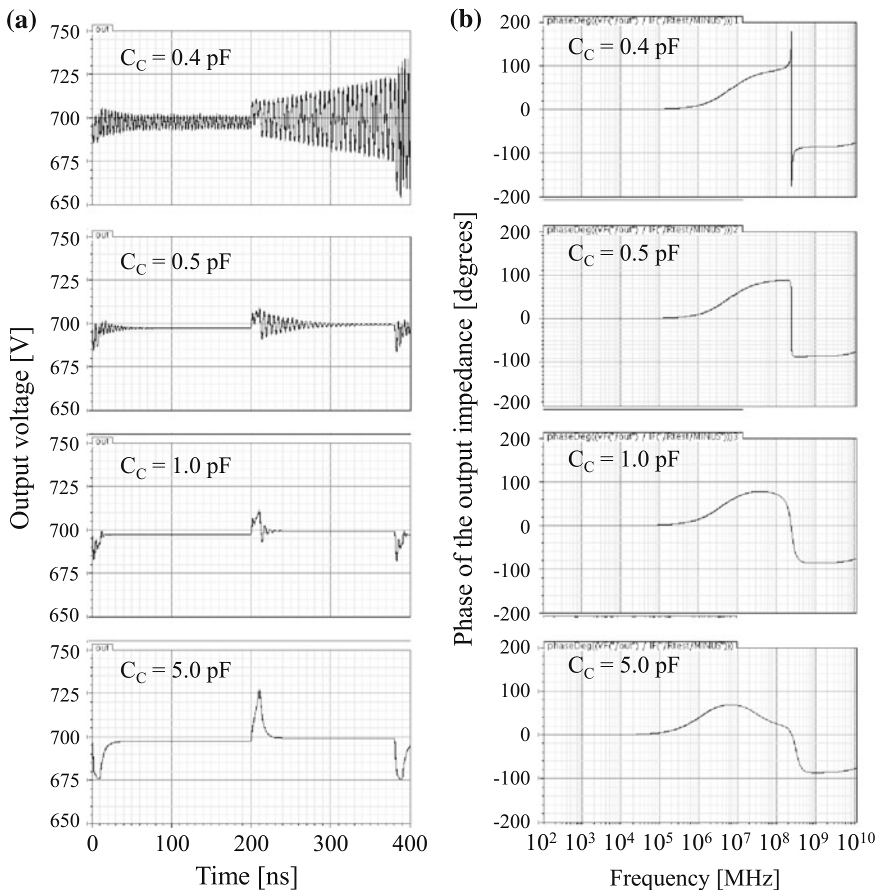


Fig. 7.17 Output response of a distributed power delivery system with different levels of compensation ($C_C = 0.4$ pF, $C_C = 0.5$ pF, $C_C = 1$ pF, and $C_C = 5$ pF), showing the correlation between the **a** transient response, and **b** phase of the output impedance

phase shift of the output voltage and load current, and the effective stability margin of the system. Based on this observation, the phase margin of the output impedance for a distributed power delivery system is

$$PM(Z_{out}) = 90^\circ - \max_{\forall f}\{\angle Z_{OUT}\}. \tag{7.3}$$

A distributed power delivery system is therefore unstable, stable, or marginally stable if the phase margin of the output impedance is, respectively, negative, positive, or zero. A safe phase margin of the output impedance should satisfy specific design criteria while avoiding excessively underdamped and overdamped voltage regulated systems.

7.4 Secure Power Delivery and Management

Non-invasive methods, such as side channel power attacks [26, 46, 47], are widely used to decode encrypted information based on the co-analysis of stored data and leakage current profiles. Alternatively, invasive power attacks (e.g., battery draining by sleep deprivation) [48] can be initiated to reduce the performance and/or lifetime of computing devices. Sensing leakage current in power delivery systems can be exploited to evaluate a specific current at the loads during cyber attacks. Several methods to mitigate power analysis attacks have recently been proposed at the architectural, algorithmic, and circuit levels [49, 50, 64]. These approaches usually mask the effective power consumption of an IC, degrading the overall power efficiency of the system. Circuits and design approaches to efficiently prevent, detect, and mitigate cyber attacks are therefore required.

The nature of power delivery circuits significantly affects the security of the system-wide power delivery system. For example, the LDO output current is proportional to the current dissipated within a linear regulator. This sensed excess current can be exploited to determine the actual current delivered by a regulator to a load. These power supplies with correlated leakage and effective currents are more susceptible to power security attacks. Alternatively, the distributed nature of the PNoC-based power delivery scheme increases the granularity of the power sources while decreasing the ability to observe system-wide power profiles based on the leakage current of an individual power supply. To efficiently mitigate power analysis attacks, existing voltage converters and regulators should be evaluated in terms of security requirements, and power supply circuits with strongly correlated internal leakage current and overall power dissipation should be avoided.

7.5 Automated Design of Stable Power Delivery Systems

To address the issues of power delivery complexity, a power delivery system should provide a scalable modular architecture that supports the integration of additional functional blocks and power features (e.g., DVS, DVFS, QoS and stability control, security mechanisms, and adaptive mechanisms to cope with PVT and enhance system efficiency) without requiring the re-design of the power delivery system. The architecture should also support heterogeneous circuits and technologies. An important aspect of systemizing the power delivery design process is to automate the design of certain complex analog circuits, further reducing the design complexity of power delivery systems. A process for automating the design of stable power delivery systems is described in this section.

Existing automated techniques for designing analog circuits are based on numerical optimization and evaluation engines [51]. Parametric models characterize the performance of an analog circuit (e.g., gain, bandwidth (BW), slew rate (SR), or phase margin (PM)) based on circuit design variables (e.g., transconductance and

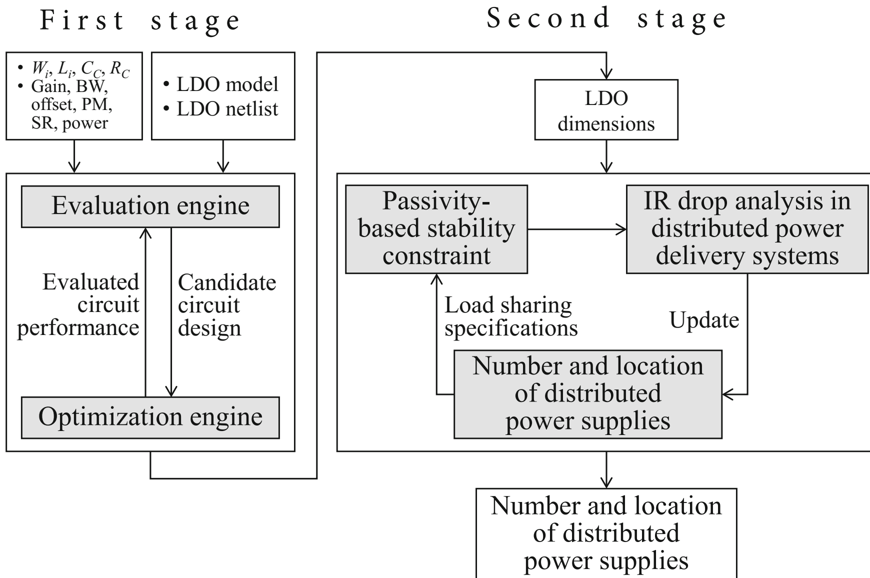


Fig. 7.18 Automated PBSC-based design flow for distributed power delivery systems

voltage bias) [51]. The performance of an individual power supply is typically characterized by a set of parameters, such as the DC gain, phase margin, DC offset, slew rate, and power dissipation. Alternatively, a distributed power delivery system is evaluated based on both the performance of the individual power regulators and additional performance metrics of the overall system such as the *phase margin of the output impedance*. The design complexity and stability of distributed on-chip power delivery systems can both be addressed by integrating the passivity-based stability criterion within existing automated design methodologies. An automated flow for designing a stable distributed power delivery system is shown in Fig. 7.18. The first stage of the flow is based on a standard parametric performance modeling technique [51]. During this stage, an LDO regulator is synthesized based on the specific LDO topology and design objectives. The output of the first stage is applied during the second stage to determine the number and location of the parallel connected power supplies in the distributed power delivery system. During this second stage, a distributed power delivery system with a different number and location of voltage regulators is iteratively evaluated based on the passivity-based stability criterion and distributed power supply placement algorithms [9, 65]. During each iteration, the worst case load sharing scenario is determined for a specific power delivery system. The passivity-based stability of the distributed system is evaluated based on the individual current loads. If required, the number and location of the power supplies are updated. The number and location of the parallel connected power supplies that satisfies the quality of power and stability requirements of the distributed power delivery system are determined.

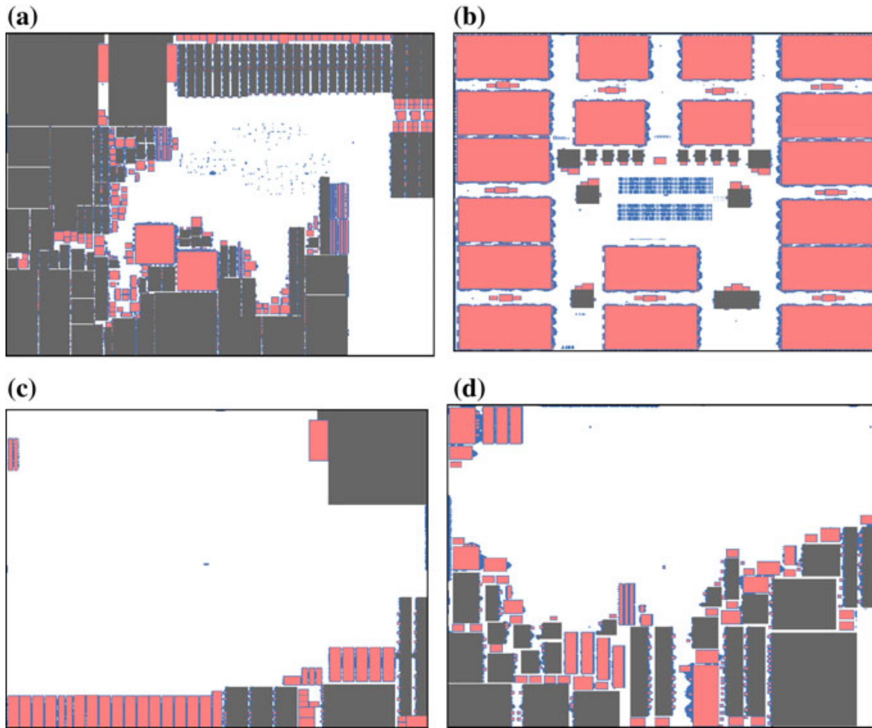


Fig. 7.19 Floorplan of ISPD'11 circuits [66] **a** superblue5, **b** superblue10, **c** superblue12, and **d** superblue18

The operation of the second stage of the automated PBSC-based design flow is demonstrated here based on the ISPD'11 placement benchmark suite circuits [66]. The floorplan of the superblue5, superblue10, superblue12, and superblue18 circuits is illustrated in Fig. 7.19. Each of the circuits is composed of 1000s of fine grain rectangular shapes. To reduce the complexity of the circuit evaluation process, the fine grain shapes are combined into larger rectangular blocks. Of the combined blocks, only the largest blocks are considered, producing a reduced floorplan. The magnitude of the distributed current loads is proportional to the size of these blocks with, for this example, a total load current of 1 A. The location of each of the current loads is at the center of the corresponding rectangular block. The number of fine grain shapes, large blocks, coverage of reduced floorplan, and power grid characteristics are listed in Table 7.2. Note that the blocks in the reduced floorplan occupy more than 85% of the total active circuit area.

A constant voltage is ideally distributed to all of the current loads within a circuit. Practically, the quality of power is degraded due to on-chip parasitic interconnect impedances. A map of the voltage drops of the superblue5 circuit without on-chip power supplies is shown in Fig. 7.20, yielding a maximum voltage drop of 23.4%

Table 7.2 Properties of ISPD benchmark circuits

Benchmark circuits	Number of fine grain shapes	Number of large combined blocks	Coverage of reduced floorplan (%)	Power grid size	Number of blocks within power grid
superblue5	29,736	129	85.0	774×713	551,862
superblue10	2318	30	89.2	638×968	617,584
superblue12	3578	15	98.4	444×518	229,992
superblue18	6776	71	99.5	381×404	153,924

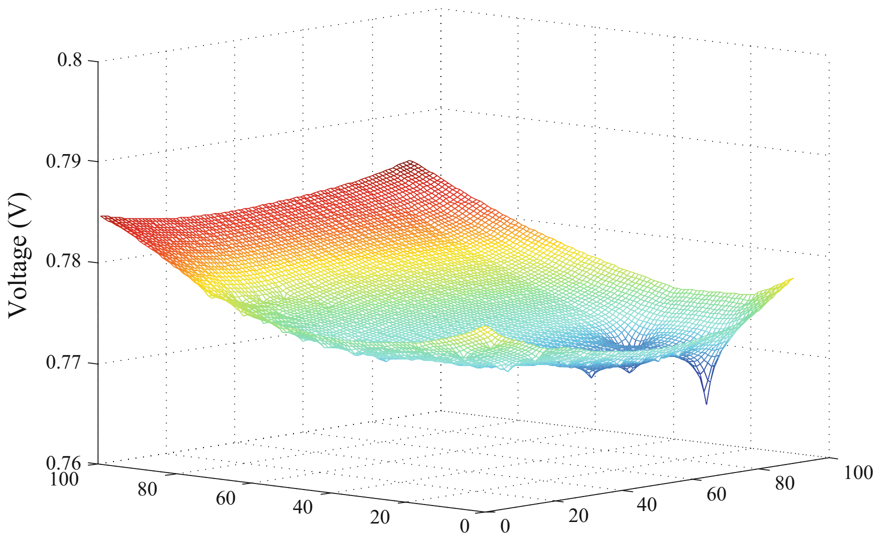


Fig. 7.20 Voltage drop map of the superblue5 circuit

assuming an off-chip voltage supply of 1 V. To address the quality of the on-chip power, power delivery systems with a single on-chip power supply (case 1), three on-chip power supplies (case 2), and twelve on-chip power supplies (case 3) are considered. For each of the three cases, the *IR* drop of a distributed power delivery system is evaluated based on the *IR* drop algorithm for a power grid with multiple power supplies and current loads [67]. The location of the power supplies in cases 1 and 2 is modeled as a mixed integer nonlinear programming problem [9], and optimized based on a general algebraic modeling system (GAMS) [68]. In case 3, the power supplies are uniformly distributed on-chip. The stability is evaluated for each of the three cases based on the passivity-based criterion. A map of the voltage drops and phase of the output impedance within superblue5 for a different number of on-chip power supplies is shown in Fig. 7.21. The maximum voltage drop decreases with an increasing number of power supplies, exhibiting a reduction in the

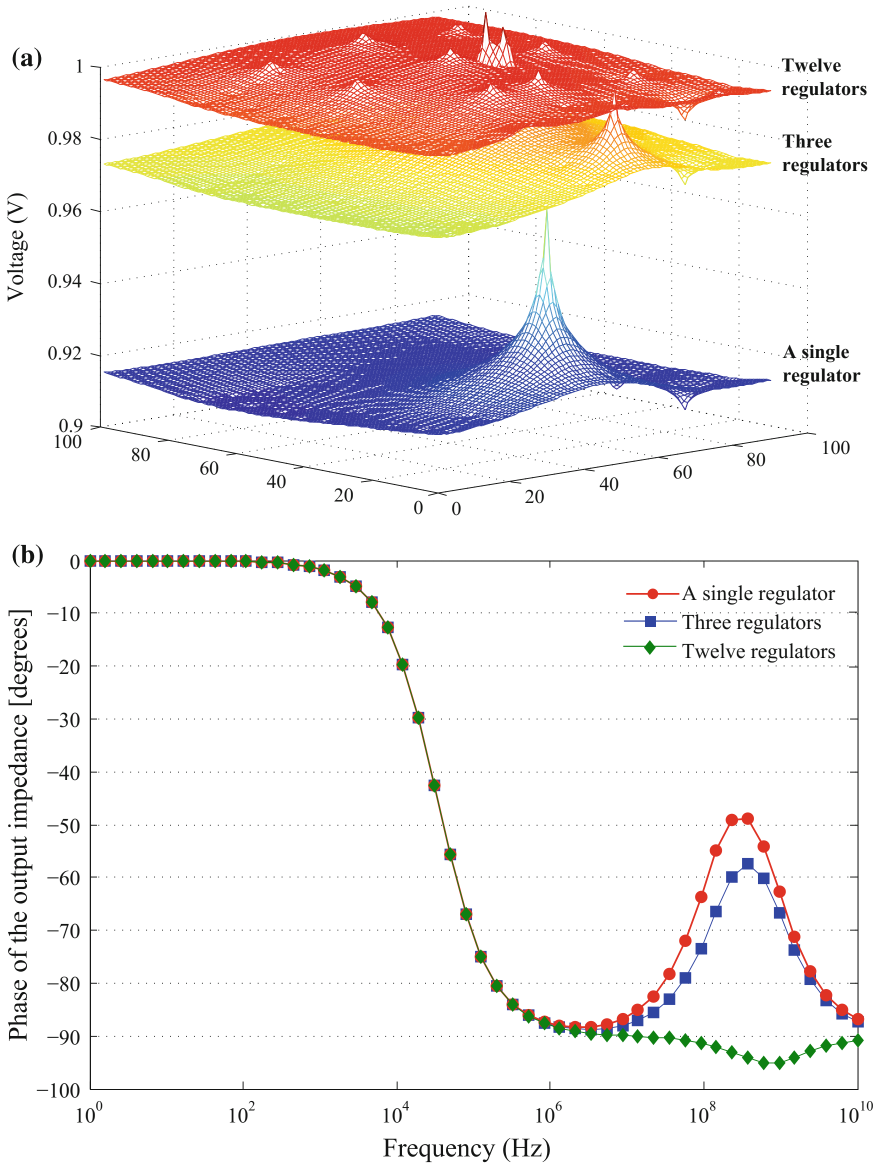


Fig. 7.21 Superblue5 circuit with a single, three, and twelve on-chip power supplies, **a** map of voltage drops, and **b** phase of output impedance

maximum voltage drop of 14.2%, 20.3%, and 22.3% with, respectively, a single, three, and twelve on-chip power supplies. Alternatively, the output current of the individual regulators is dependent on the number of power supplies, affecting both the phase of the output impedance and the stability characteristics of the distributed

Table 7.3 *IR* drop and stability characteristics of ISPD benchmark circuits

Circuit	No regulators	A single regulator		Three regulators		Twelve regulators	
	Max. IR drop (%)	Max. IR drop (%)	Stability	Max. IR drop (%)	Stability	Max. IR drop (%)	Stability
superblue5	23.4	9.17	Stable	3.11	Stable	1.11	Unstable
superblue10	23.0	10.8	Stable	4.75	Stable	2.43	Unstable
superblue12	23.6	10.7	Stable	6.39	Stable	4.88	Unstable
superblue18	22.7	10.7	Stable	4.43	Stable	1.54	Unstable

system. Based on the stability criterion, the superblue5 circuit with the power delivery system is stable with a single power supply and three power supplies (the phase of the output impedance is within the $(-90°, 90°)$ range), and unstable with twelve power supplies (the minimum phase of the output impedance is $-95.1° < -90°$). While the distributed power delivery system with twelve power supplies exhibits higher quality of power than the systems with fewer power supplies, this system is shown to be unstable under a step load. Thus, a system with fewer power supplies is preferable to deliver power within the superblue5 circuit when considering both quality of power and system stability.

The second stage of the automated PBSC-based design flow, shown in Fig. 7.18, is implemented in Matlab. Pseudo-code of the Matlab algorithm is summarized in Algorithm 1. An LDO model [59] describes the small signal response of the on-chip power supplies, and is used to evaluate the output impedance of the power supplies and overall power delivery system. The power delivery systems for the ISPD'11 benchmark circuits, superblue5, superblue10, superblue12, and superblue18, are evaluated based on this algorithm. The maximum *IR* drop and stability characteristics are listed in Table 7.3. For these benchmark circuits, the maximum voltage drop is significantly less with increasing number of on-chip power supplies. Alternatively, the stability of the distributed power delivery system is a function of the specific load distribution and affected by the output impedance of the POL power delivery system. The automated PBSC-based design flow generates a distributed power delivery system that addresses both the quality of power and stability requirements.

7.6 Summary

Delivering high quality, dynamically controlled power to support power efficient and portable systems is a critical challenge in *heterogeneous* systems-on-chip. Efficiently providing multiple point-of-load on-chip regulated voltages requires fundamental changes to the power delivery and management process. To cope with the increasing design complexity of heterogeneous, distributed, and dynamically controllable

Algorithm 1 Automated PBSC-based design flow.

1: // A typical LDO model [69]
2: *LDOModel*;
3: // Supply voltages, load currents and locations
4: *CircInfo*;
5: // All nodes in the evaluated circuit
6: *CircNodes*;
7: // List of numbers of LDOs to evaluate
8: *NumRegsList*;
9: // Number of LDOs in the preferred power delivery system
10: *PreferredNumRegs* ← 0;
11: // Location of the LDOs in the preferred power delivery system
12: *PreferredLocs* ← N/A;
13: // Maximum *IR* drop of the preferred power delivery system
14: *PreferredIRDrop* ← 1;
15:
16: **for all** *NumRegs* ← *NumRegsList* **do**
17:   **for all** *RefNode* ← *CircuitNodes* **do**
18:     // Find optimal LDO locations [68]
19:     *OptLocs* ← OPT_LOC(*CircInfo*, *NumRegs*);
20:
21:     // Analyze *IR* drop in a power grid [67]
22:     *IRDrop*(*RefNode*) ← CALC_IRDROP(*CircInfo*, ...
23:                 *OptLocs*, ...
24:                 *RefNode*);
25:
26:     // Calculate the output impedance of the LDOs
27:     **for all** *LDO* ← *LDOs* **do**
28:       *ISupply* ← Current delivered by the LDO;
29:       $Z_{OUT}^{LDO}(LDO)$ ← CALC_ZOUT(*LDOModel*, ...
30:                 *ISupply*);
31:     **end for**
32:
33:     // Calculate the output impedance of the system
34:     $Z_{OUT}^{SYS} \leftarrow \left(\sum_{LDOs} \frac{1}{Z_{OUT}^{LDO}(LDO)} \right)^{-1}$;
35:
36:     **if** $\max\{|\angle Z_{OUT}^{SYS}|\} < 90°$ **then**
37:       **if** $\max\{IRDrop\} < PreferredIRDrop$ **then**
38:         *PreferredIRDrop* ← $\max\{IRDrop\}$;
39:         *PreferredNumRegs* ← *NumRegs*;
40:         *PreferredLocs* ← *OptLocs*;
41:       **end if**
42:     **end if**
43:   **end for**
44: **end for**

power delivery systems while enhancing the scalability of both static and dynamic power delivery systems, a systematization of the power delivery and management process is required.

The concept of intelligent on-chip networks for delivering power is introduced here as a means to address the objectives of scalable, systematic power delivery and management in high performance heterogeneous integrated systems. Primary design and performance challenges, such as physical and virtual power routing, stability, security, complexity, and scalability of PNoC based systems are reviewed. The PNoC architecture is evaluated, exhibiting significant power savings as compared to traditional power delivery systems.

A passivity-based stability criterion is described. Based on this criterion, a distributed power delivery system is stable if the total output impedance of the parallel connected LDOs exhibits no right half plane poles and a phase between $-90°$ and $+90°$. Similar to a single voltage regulator, the phase margin of the output impedance (the difference between the maximum phase and 90°) determines the stability of a distributed power delivery system. The integration of emerging techniques for mitigating power attacks within the proposed PNoC platform is discussed. To efficiently mitigate power analysis attacks, power supply circuits with strongly correlated internal leakage power and overall power dissipation should be avoided. Distributed power management functions are preferable to make the system less prone to power attacks.

Reducing the design complexity while increasing the scalability of the power delivery design process with on-chip power networks is described. To further reduce the design complexity, a process for automating the design of power delivery systems is proposed. The automated design flow is evaluated with ISPD benchmarks, exhibiting a distributed power delivery system that addresses both the quality of power and stability requirements. Distributed intelligence should be added to on-chip power networks to provide a secure and systematic power delivery and management process while reducing design complexity and increasing scalability.

References

1. Semiconductor Industry Association, The International Technology Roadmap for Semiconductors (2012), http://www.itrs.net
2. W. Arden, M. Brillout, P. Cogez, M. Graef, B. Huizing, R. Mahnkopf, More-than-Moore (2010), http://www.itrs.net
3. P. Cauvet, S. Bernard, M. Renovell, System-in-package, a combination of challenges and solutions, in *Proceedings of the IEEE VLSI Test Symposium*, pp. 193–199, May 2007
4. W. Wolf, A.A. Jerraya, G. Martin, Multiprocessor system-on-chip (MPSoC) technology. IEEE Trans. Comput. Aided Design Integr. Circuits Syst. **27**(10), 1701–1713 (2008)
5. V.F. Pavlidis, E.G. Friedman, Interconnect-based design methodologies for three-dimensional integrated circuits. Proc. IEEE **97**, 123–140 (2009)
6. V.F. Pavlidis, E.G. Friedman, *Three-Dimensional Integrated Circuit Design*, Morgan Kaufmann (2009)
7. R. Jakushokas, M. Popovich, A.V. Mezhiba, S. Kose, E.G. Friedman, *Power Distribution Networks with On-Chip Decoupling Capacitors*, 2nd edn. (Springer, 2011)
8. S. Kose, S. Tam, S. Pinzon, B. McDermott, E.G. Friedman, Active filter based hybrid on-chip DC-DC converters for point-of-load voltage regulation. IEEE Trans. Very Large Scale Integr. (VLSI) Circuits, **21**(4), 680–691 (2013)

9. S. Kose, E.G. Friedman, Distributed on-chip power delivery. IEEE J. Emerg. Sel. Topics Circuits Syst. **2**(4), 704–713 (2012)
10. Y. Kanno et al., Hierarchical power distribution with 20 power domains in 90-nm low-power multi-CPU processor, in *Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 540–541, February (2006)
11. T. Hattori et al., A power management scheme controlling 20 power domains for a single-chip mobile processor, in *Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 542–543, February (2006)
12. Y. Kikuchi et al., A 40 nm 222 mW H.264 full-HD decoding, 25 power domains, 14-core application processor with x512b stacked DRAM. IEEE J. Solid-State Circuits **46**(1), 32–41 (2011)
13. D.N. Truong, W.H. Cheng, T. Mohsenin, Z. Yu, A.T. Jacobson, G. Landge, M.J. Meeuwsen, C. Watnik, A.T. Tran, Z. Xiao, E.W. Work, J.W. Webb, P.V. Mejia, B.M. Baas, A 167-processor computational platform in 65 nm CMOS. IEEE J. Solid-State Circuits **44**(4), 1130–1144 (2009)
14. S.Y. Borkar, Thousand core chips—a technology perspective, in *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 746–749, June 2007
15. D. Yeh, L.-S. Peh, S. Borkar, J. Darringer, A. Agarwal, W.-M. Hwu, Thousand-core chips [Roundtable]. IEEE Design Test Comput. **25**(3), 272–278 (2008)
16. H. Esmaeilzadeh, E. Blem, R. St. Amant, K. Sankaralingam, D. Burger, Dark silicon and the end of multicore scaling, in *Proceedings of the ACM International Symposium on Computer Architecture*, pp. 365–376, June 2011
17. I.P.-Vaisband, Power Delivery and Management in Nanoscale ICs, in *Ph.D. Thesis* (University of Rochester, May 2015)
18. I. Vaisband, B. Price, S. Kose, Y. Kolla, E.G. Friedman, J. Fischer, Distributed LDO regulators in a 28 nm power delivery system. Analog Integr. Circuits Signal Proces. **83**(3), 295–309 (2015)
19. S. Gangopadhyay, S.B. Nasir, A. Raychowdhury, Integrated power management in IoT devices under wide dynamic ranges of operation, in *Proceedings of the IEEE/ACM Design Automation Conference*, May 2015
20. I. Vaisband, M. Azhar, E.G. Friedman, S. Kose, Digitally controlled pulse width modulator for on-chip power management. IEEE Trans. Very Large Scale Integr. (VLSI) Circuits **22**(12), 2527–2534 (2014)
21. P. Macken, M. Degrauwe, M.V. Paemel, H. Oguey, A voltage reduction technique for digital systems, in*Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 238–239, February 1990
22. Q. Wu, P. Juang, M. Martonosi, D.W. Clark, Voltage and frequency control with adaptive reaction time in multiple-clock domain processors, in *Proceedings of the IEEE International Symposium on High Performance Computer Architecture*, pp. 178–189, February 2005
23. C. Isci, A. Buyuktosunoglu, C.-Y. Cher, P. Bose, M. Martonosi, An analysis of efficient multi-core global power management policies: maximizing performance for a given power budget, in *Proceedings of the Annual IEEE/ACM International Symposium on Microarchitecture*, pp. 347–358, December 2006
24. M. Ware, K. Rajamani, M. Floyd, B. Brock, J.C. Rubio, F. Rawson, and J.B. Carter, architecting for power management: the IBM POWER7^{TM} approach, in *Proceedings of the IEEE International Symposium on High Performance Computer Architecture*, pp. 1–11, January 2010
25. J. Friedrich et al., The Power8^{TM} processor: designed for big data, analytics, and cloud environments, in *Proceedings of the IEEE International IC Design & Technology Conference*, pp. 1–4, May 2014
26. M. Alioto, L. Giancane, G. Scotti, A. Trifiletti, Leakage power analysis attacks: a novel class of attacks to nanometer cryptographic circuits. IEEE Trans. Circuits Syst. I Regul. Pap. **57**(2), 355–367 (2010)
27. F. Waldron, J. Slowey, A. Alderman, B. Narveson, and S.C. O'Mathuna, Technology roadmapping for power supply in package (PSiP) and power supply on chip (PwrSoC), in *Proceedings of the IEEE International Applied Power Electronics Conference and Exposition*, pp. 525–532, February 2010

28. H.-P. Le, S.R. Sanders, E. Alon, Design techniques for fully integrated switched-capacitor DC-DC converters. IEEE J. Solid-State Circuits **46**(9), 2120–2131 (2011)
29. Y. Ramadass, A. Fayed, A. Chandrakasan, A fully-integrated switched-capacitor step-down DC-DC converter with digital capacitance modulation in 45 nm CMOS. IEEE J. Solid-State Circuits **45**(12), 2557–2565 (2010)
30. M. Bathily, B. Allard, F. Hasbani, A 200-MHz integrated buck converter with resonant gate drivers for an RF power amplifier. IEEE Trans. Pow. Electron. **27**(2), 610–613 (2012)
31. Y. Ahn, H. Nam, J. Roh, A 50-MHz fully integrated low-swing buck converter using packaging inductors. IEEE Trans. Pow. Electron. **27**(10), 4347–4356 (2012)
32. Y.-H. Lee, S.-C. Huang, S.-W. Wang, W.-C. Wu, P.-C. Huang, H.-H. Ho, Y.-T. Lai, K.-H. Chen, Power-tracking embedded buckboost converter with fast dynamic voltage scaling for the SoC system. IEEE Trans. Pow. Electron. **27**(3), 1271–1282 (2012)
33. H. Jia, J. Lu, X. Wang, K. Padmanabhan, Z.J. Shen, Integration of a monolithic buck converter power IC and bondwire inductors with ferrite epoxy glob cores. IEEE Trans. Pow. Electron. **26**(6), 1627–1630 (2011)
34. W. Yan, W. Li, R. Liu, A noise-shaped buck DCDC converter with improved light-load efficiency and fast transient response. IEEE Trans. Pow. Electron. **26**(12), 3908–3924 (2011)
35. L. Wang, Y. Pei, X. Yang, Y. Qin, Z. Wang, Improving Light and intermediate load efficiencies of buck converters with planar nonlinear inductors and variable on time control. IEEE Trans. Pow. Electron. **27**(1), 342–353 (2012)
36. H. Nam, Y. Ahn, J. Roh, 5-V buck converter using 3.3-V standard CMOS process with adaptive power transistor driver increasing efficiency and maximum load capacity. IEEE Trans. Pow. Electron. **27**(1), 463–471 (2012)
37. M. Wens, M.S.J. Steyaert, A fully integrated CMOS 800-mW fourphase semiconstant ON/OFF-Time step-down converter. IEEE Trans. Pow. Electron. **26**(2), 326–333 (2011)
38. J. Guo, K.N. Leung, A 6-μW chip-area-efficient output-capacitorless LDO in 90-nm CMOS technology. IEEE J. Solid-State Circuits **45**(9), 1896–1905 (2010)
39. P. Hazucha et al., Area-efficient linear regulator with ultra-fast load regulation. IEEE J. Solid-State Circuits **40**(4), 933–940 (2005)
40. T.Y. Man, K.N. Leung, C.Y. Leung, P.K.T. Mok, M. Chan, Development of single-transistor-control LDO based on flipped voltage follower for SoC. IEEE Trans. Circuits Syst. I Regul. Pap. **55**(5), 1392–1401 (2008)
41. M. Al-Shyoukh, H. Lee, R. Perez, A Transient-enhanced low-quiescent current low-dropout regulator with buffer impedance attenuation. IEEE J. Solid-State Circuits **42**(8), 1732–1742 (2007)
42. V. Kursun, S.G. Narendra, V.K. De, E.G. Friedman, Analysis of buck converters for on-chip integration with a dual supply voltage microprocessor. IEEE Trans. Very Large Scale Integr. (VLSI) Circuits **11**(3), 514–522 (2003)
43. I. Vaisband, E.G. Friedman, Heterogeneous methodology for energy efficient distribution of on-chip power supplies. IEEE Trans. Pow. Electron. **28**(9), 4267–4280 (2013)
44. D.E. Lackey, P.S. Zuchowski, T.R. Bednar, D.W. Stout, S.W. Gould, J.M. Cohn, Managing power and performance for system-on-chip designs using voltage islands, in *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 195–202, November 2002
45. P. Magarshack, P.G. Paulin, System-on-chip beyond the nanometer Wall, in *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 419–424, June 2003
46. P. Kocher, J. Jaffe, B. Jun, Differential Power Analysis, in *Proceedings of the International Conference on Advances in Cryptology*, pp. 388–397, August 1999
47. L. Lin, W. Burleson, Leakage-based differential power analysis (LDPA) on sub-90 nm CMOS cryptosystems, in *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 252–255, May 2008
48. T. Martin, M. Hsiao, D.S. Ha, J. Krishnaswami, Denial-of-service attacks on battery-powered mobile computers, in *Proceedings of the IEEE Annual Conference on Pervasive Computing and Communications*, pp. 309–318, March 2004

49. H. Vahedi, R. Muresan, S. Gregori, On-chip current flattening circuit with dynamic voltage scaling, in *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 4277–4280, May 2006
50. K. Tiri, I. Verbauwhede, A digital design flow for secure integrated circuits. IEEE Trans. Comput. Aided Design Integr. Circuits Syst. **25**(7), 1197–1208 (2006)
51. R.A. Rutenbar, G.G.E. Gielen, J. Roychowdhury, Hierarchical modeling, optimization, and synthesis for system-level analog and RF designs. Proc. IEEE **95**(3), 640–669 (2007)
52. M. Popovich, M. Sotman, A. Kolodny, E.G. Friedman, Effective radii of on-chip decoupling capacitors. IEEE Trans. Very Large Scale Integr. (VLSI) Circuits, **16**(7), 894–907 (2008)
53. S.R. Nassif, Power grid analysis benchmarks, in *Proceedings of the IEEE/ACM Asia and South Pacific Design Automation Conference*, pp. 376–381, January 2008
54. I. Vaisband, E.G. Friedman, Power network on-chip for scalable power delivery, United States Patent (pending)
55. I. Vaisband, E.G. Friedman, Power network on-chip for scalable power delivery, in *Proceedings of the Workshop on System Level Interconnect Prediction*, pp. 1–5, June 2014
56. I. Vaisband, E.G. Friedman, Dynamic power management with power network-on-chip, in *Proceeding of the IEEE International Conference on New Circuits and Systems*, pp. 225–228, June 2014
57. J.H. Mulligan Jr., The effect of pole and zero locations on the transient response of linear dynamic systems. Proc. Inst. Radio Eng. **37**(5), 516–529 (1949)
58. A. Riccobono, E. Santi, A novel passivity-based stability criterion (PBSC) for switching converter DC distribution systems, in *Proceedings of the IEEE International Applied Power Electronics Conference and Exposition*, pp. 2560–2567, February 2012
59. D.A. Johns, K. Martin, *Analog Integrated Circuit Design* (Wiley, 1997)
60. J.E. Colgate, The control of dynamically interacting systems, in *Ph.D. Thesis*, Massachusetts Institute of Technology, August 1988
61. J.L. Wyatt, L.O. Chua Jr., J. Goknar, I. Gannett, D. Green, Energy concepts in the state-space theory of nonlinear n-ports: part i-passivity. IEEE Trans. Circuits Syst. **28**(1), 48–61 (1981)
62. O. Brune, Synthesis of a finite two-terminal network whose driving-point impedance is a prescribed function of frequency, in *Ph.D. Thesis*, Massachusetts Institute of Technology, August 1931
63. V. Kursun, E.G. Friedman, *Multi-Voltage CMOS Circuit Design* (Wiley, 2006)
64. O.A. Uzun, S. Kose, Converter-gating: a power efficient and secure on-chip power delivery system. IEEE J. Emerg. Sel. Top. Circuits Syst. **4**(2), 169–179 (2014)
65. S. Kose, E.G. Friedman, Efficient algorithms for fast IR drop analysis exploiting locality. Integr. VLSI J. **45**(2), 149–161 (2012)
66. N. Viswanathan et al., The ISPD-2011 routability-driven placement contest and benchmark suite, in *Proceedings of the ACM International Symposium on Physical Design*, pp. 141–146, March 2011
67. S. Kose, E.G. Friedman, Fast algorithms for IR voltage drop analysis exploiting locality, in *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 996–1001, June 2011
68. A. Brooke, D. Kendrick, A. Meeraus, *GAMS: A User's Guide* (The Scientific Press, 1992)
69. G.A. Rincon-Mora, Current efficient, low voltage, low dropout regulators, in *Ph.D. Thesis*, Georgia Institute of Technology, November 1996

Chapter 8
Energy Efficient System Architectures

Avinoam Kolodny

Abstract Energy consumption by data centers is growing at an exponential rate. Similar growth in power consumption had occurred in smaller-scale microprocessor-based computing. In this chapter, the evolution of microprocessor systems is described, highlighting principles and examples of power-saving strategies. The same principles can be applied to large-scale computing centers, in order to address the upcoming power crisis in cloud computing.

8.1 Power Issues in Computing Systems

Cloud computing, enabled by the combined capabilities of communication networks and advanced computer technology, created exponential growth in demand for data processing. This growth is reflected in energy consumption of large data centers. A data center, which may be viewed as a modern large-scale computer, is typically a large industrial building (more than 1 million square meters), consuming more than 100 MW of electrical power. The total yearly world-wide electrical energy consumed by data centers is shown in Fig. 8.1 [1]. As can be seen, the appetite for computing power grows steadily, and is projected to become within a few years a significant fraction of the global energy consumption, which is intolerably high.

Clearly, steps must be taken to change the trend towards higher and higher energy consumption in modern cloud computing systems. The challenge of power reduction in computing systems is not new. In fact, power efficiency had been improved in the past during most generation of computer technology, but the current rate of improvement is insufficient. The goal of this chapter is to examine the past development of smaller scale computing systems, particularly microprocessor systems, and to identify architectural principles for power saving, which may

A. Kolodny (✉)
Electrical Engineering department, Technion, Haifa, Israel
e-mail: kolodny@ee.technion.ac.il

© Springer International Publishing AG 2017

G. Eisenstein and D. Bimberg (eds.), *Green Photonics and Electronics*,
NanoScience and Technology, DOI 10.1007/978-3-319-67002-7_8

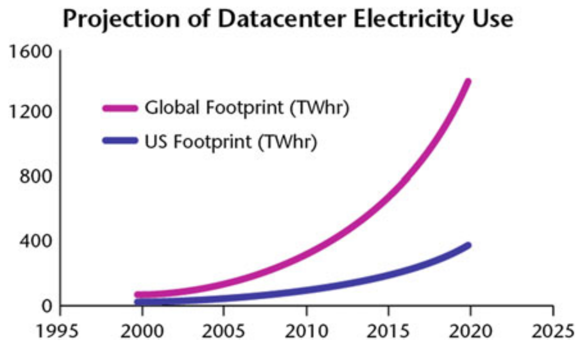


Fig. 8.1 Exponential growth in energy used by U.S. and world-wide data centers. *Source* http://www.computerworld.com/article/2598562/data-center/data-centers-are-the-new-polluters.html

be applied in turn for restraining the large scale energy requirements of cloud computers.

8.2 Characteristics of the Power Reduction Problem

Power dissipation, i.e. conversion of electrical energy into heat, is an inevitable side-effect of all hardware activities. The total energy consumed by a system is the sum of many small amounts, corresponding to all activities in the system. Spending of energy in a system is thus similar to spending of money in a large organization, where each activity is associated with some financial cost and the total cost is the sum of many small expenses. The cumulative nature of spending energy (or money) in a system defines certain characteristics for the task of power reduction, which is carried out by people who design and operate the system. The most important and obvious principle is that every little saving counts, just like a coin added to a savings box. In other system improvement tasks, such as system speed optimization, there is some critical sub-problem (such as a critical speed path) which masks all others, so that people can focus only on solving the critical issue, while other issues can be ignored. When dealing with power reduction, people still tend to focus on the biggest sub-problems, which are the highest power consuming components. However, numerous small sources of energy waste become significant as they accumulate together.

8.2.1 The Disruption Principle

When a problematic trend can go unnoticed, because the magnitude of the problem is still negligible, the trend often tends to continue until the problem becomes acute (causing too much pain or too much cost). This is primarily because people are accustomed to work only on the most burning issues. An example of this

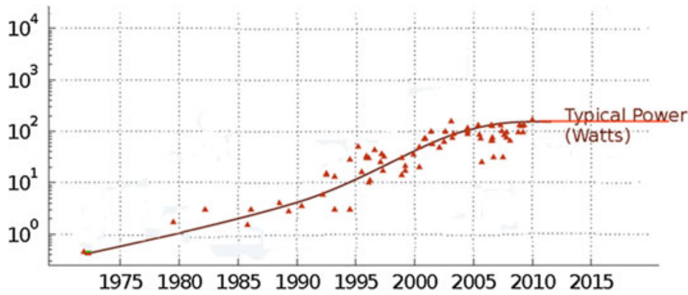


Fig. 8.2 Power dissipation in processor chips versus years (each point represent a processor product). *Source* http://cpudb.stanford.edu

phenomenon can be seen by plotting the progression of power consumption in microprocessor chips over many years (Fig. 8.2).

Since the introduction of the first microprocessors, their capabilities were developed according to Moore's law, which predicted doubling of the number of transistors in the fixed area (about 1 cm^2) of a silicon chip every two years due to device miniaturization. According to classical scaling theory [2], downscaling of all geometrical dimensions of devices in VLSI technology should lead to fixed power density, such that the power dissipated should stay unchanged in a chip, if the number of transistors is doubled but their size is halved, such that the chip area remains the same. As can be seen from the datapoints for the first 20 years, microprocessors indeed consumed about 5 W despite the exponential growth in their complexity during those years, as predicted by the theory. However, this was no longer true in the mid-90s. During these years, major deviations from classical scaling theory and wasteful spending of power (while optimizing circuits for high performance) have led to quick growth of power dissipation in microprocessors, and the bad trend was not changed until power levels hit about 100 W. This value, reached at about 2005, made the problem critical since air-cooling using fans could no longer remove the heat, and other cooling methods were impractically expensive. In addition, mobile devices using microprocessors became important, and low power was required for long battery life in such devices. At this point, various actions by product developers stopped the growth and caused the power curve to reach a plateau, while the number of transistors and the total chip performance continued to grow (this means that power efficiency has been steadily improved since then, as seen in Fig. 8.3).

8.2.2 The Locality Principle

The continuous improvement in power efficiency of processor chips, which started when power dissipation was recognized as a critical disruptive problem, has been

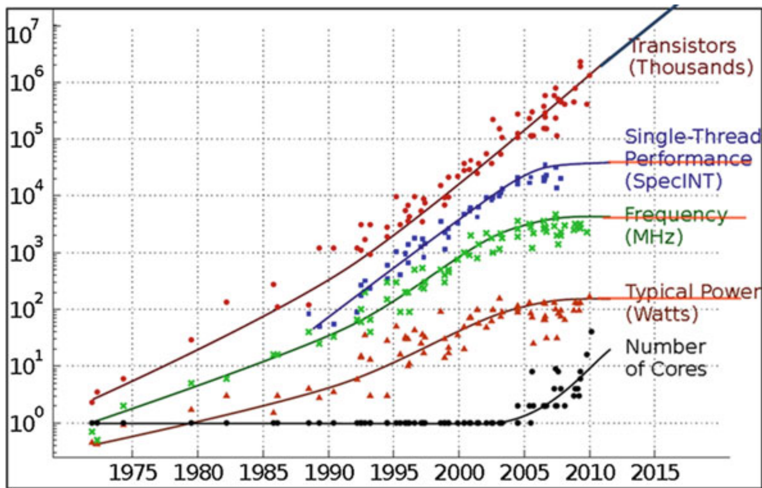


Fig. 8.3 Main architectural changes in microprocessors. The move to multi-core chips and no more growth in clock frequency stem from power considerations. *Source* http://cpudb.stanford.edu

achieved by combined efforts of design engineers working at many different levels. Like cost-cutting programs in big organizations, power-saving redesign activities were initiated in order to eliminate energy waste in the layout, at the circuit level, the logic design and the system architecture of processor chips. Two salient changes in processor architecture are clearly evident in Fig. 8.3: the industry stopped making larger and larger uniprocessors (it moved to chips containing multiple small processor cores), and the clock frequency within each core stopped growing (the quick growth in operating frequency was a major contributor to the problematic growth of power dissipation since the early 90s). Clearly, the most fundamental action taken by the industry was the move to multi-core chips. The basic principle behind this change is the locality principle: each core performs a computation locally, while the total performance of the chip is the sum of local performances (assuming that the computation task can be divided to portions which can be performed independently by different cores in parallel). This is illustrated in Fig. 8.4.

In order to understand the rationale and the importance of the locality principle, it is interesting to note Pollack's empirical rule [3]. Pollack tried to compare the architectural effectiveness of different uniprocessor architectures, assuming that all these processors could be manufactured in the same process technology. He plotted the performance of several generations of Intel uniprocessors versus the silicon area that they would occupy using this fixed technology. The result, shown in Fig. 8.5, is a sub-linear curve approximated by a square-root function. Assuming that dissipated power is proportional to chip area, it is clear that building a big uniprocessor by doubling the area of a small processor, is expected to double the power but provide only about 1.4X (sqrt(2)) in performance. Using the same area by two small

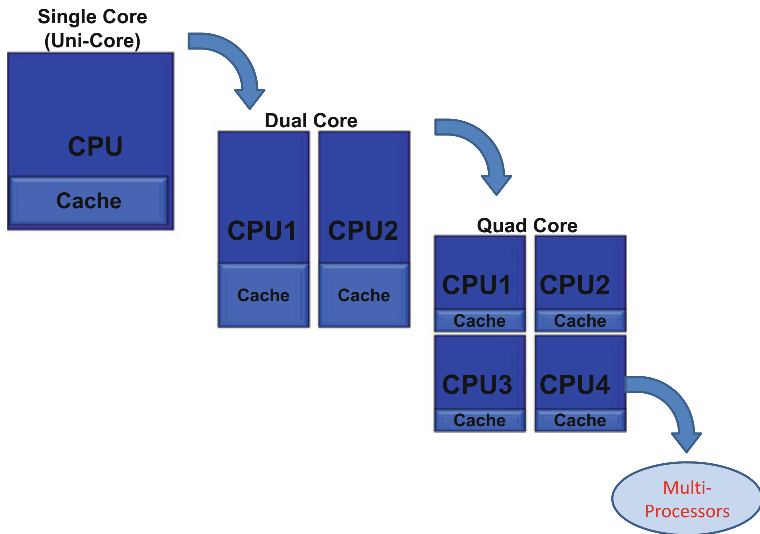
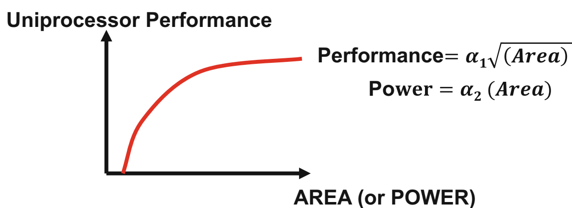


Fig. 8.4 Evolution of chip multi-processors (CMPs)

Fig. 8.5 Pollack's empirical rule: when different processor architectures are built in the same process technology, power grows linearly with chip area while performance grows sub-linearly



processors which operate in parallel can provide twice the performance. Continuing this line of reasoning, it makes sense to divide the computational task to many small local computations in parallel.

What is the reason for Pollack's rule? Basically, the more complicated architectures involve complex mechanisms which produce marginal returns in performance, including even some speculative execution which sometimes fails to provide any performance gain. However, the underlying physical explanation for this rule is related to the nature of power dissipation in CMOS technology: The logical state of variables is represented by electric charge, and each signal transition involves charging or discharging of some capacitance. In each cycle of charging and discharging, the charge $Q = CV$ causes conversion of energy $QV = CV^2$ into heat. The total power consumed by a chip during its operation, called active dynamic power, is due to the sum of activities of all signals in the chip. It can be expressed by the expression

$$P = \alpha C_{tot} V^2 f$$

where the total capacitance of all elements on the chip is multiplied by alpha (the average activity factor per clock cycle) and f is the clock frequency in cycles per second. The total capacitance is composed of device capacitances, and interconnect wire capacitances. In recent technology generations, device capacitances were scaled down while interconnect capacitances could not be scaled effectively [4]. Hence, wire capacitances have become the dominant contributors to on-chip power dissipation [5]. In fact, moving data on the chip requires more energy than performing a typical computational instruction on this data, by about 1–2 orders of magnitude). The wires have also become limiters of circuit speed [6]. Large uniprocessors contain long-distance wires, and the number of such wires grows with the complexity of the architecture. Basically, the inefficiency of large complex uniprocessors is caused by long-distance signal propagation on the chip. The energy dissipation and the delays associated with long-distance wires are avoided by using small simple cores which perform local computations in parallel.

8.2.3 The Challenge of Parallelism

Although parallel computing is required as a corollary of the locality principle, parallel algorithms and parallel programming are largely unsolved challenges for the computer science and engineering research communities. In the past, it had been often easier to use a faster processor instead of converting a serial program into an efficient parallel program. Nowadays, the current power crisis in computing may lead to innovative solutions driven by necessity [7].

Many applications which are written for processing in a multicore environment, allowing them to invoke numerous worker threads, suffer from inherent inefficiencies which stem from the algorithm. It is useful to characterize the inherent behavior of such software programs by simulating their performance as a function of the number of worker threads on an ideal machine (perfect fast-access memory with no shared resources and no architectural constraints). Plots like those in Fig. 8.6 reveal some “embarrassingly parallel” applications where the performance grows exactly as the number of threads, some programs where the performance growth lags behind the number of threads (because some threads work part-time only), and some poorly scalable programs which cannot utilize more than a certain number of threads. These types of program behavior are depicted in Fig. 8.7. This is an outcome of the well-known Amdahl’s principle, which says that serial code within a program (ordered instructions which must be executed sequentially) limits the speedup which may be achieved by parallel processing of the program. If more processors are invoked to serve such programs, most of the added processors would just waste power by being idle in operative mode while waiting for other threads to finish their tasks. In such cases, from the viewpoint of energy efficiency, it makes

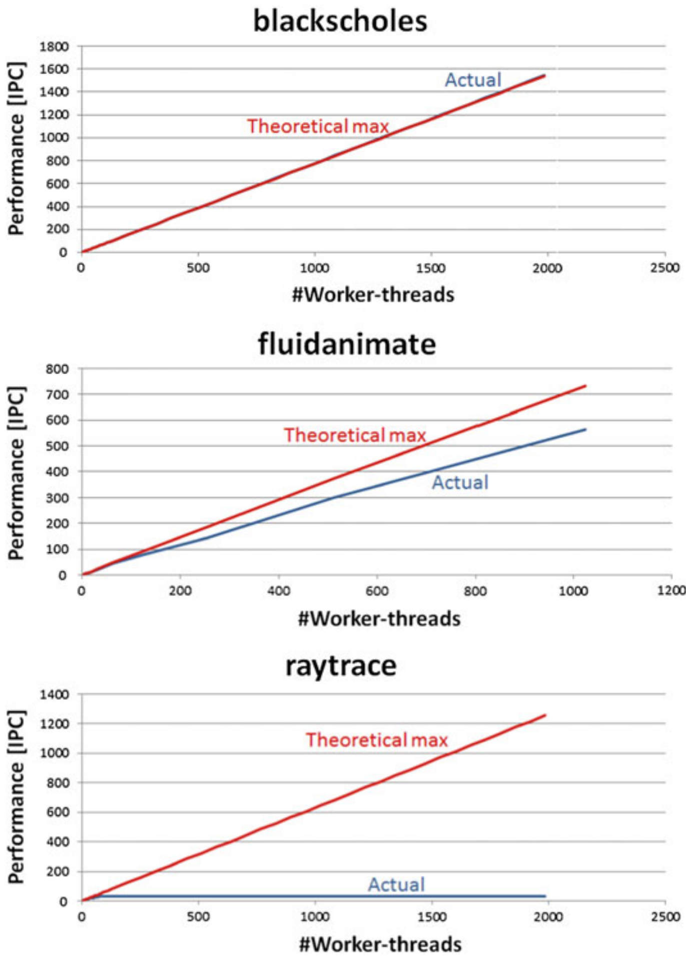


Fig. 8.6 Theoretical and actual performance of three software application programs versus number of allowed worker-threads, simulated on an ideal architecture. The gap between theoretical and actual is inherent in the programs

sense to hurry up and shut down idle cores, instead of wasting power (on leakage, clock operation, or other system components) while waiting.

In certain systems, it may be possible to bypass the challenge of parallel programming by assigning independent, unrelated tasks to different cores. In a data center environment, this happens when each core is assigned to serve a different client.

However, even if there are no intrinsic dependencies among computing tasks, they might still affect each other due to architectural constraints. A common reason for undesired interaction is contention on common shared resources, such as a memory bus: when different software threads need the same bus at the same time,

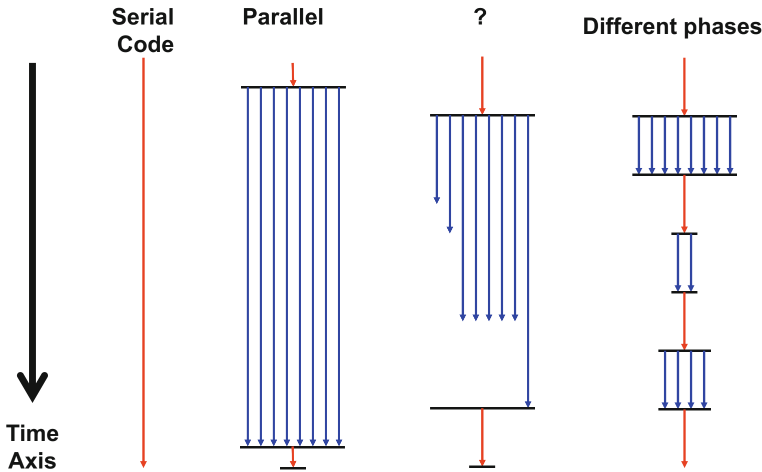


Fig. 8.7 Diagrams of parallelism in different programs, containing serial code segments (*red*), and parallel code (*blue*). Some threads may be idle until syncronization points are reached

only one of them gets the service while others are waiting. As explained above, waiting in active mode involves a waste of energy. Another example of undesired interaction is destructive interference, such as eviction of data from the shared cache because the cache location is needed for data from another software thread. System architects need to minimize such interferences by providing ample shared resources, or by limiting the number of competing threads, balancing the system such that no resource bottlenecks will be active. Any waiting for resources incurs wasting of power.

8.2.4 A Unified Machine Model

Computer architectures are based on processing units and external memory units. The communication channels between them have been a system bottleneck since the early days of computing machines. Many different approaches have been tried in order to improve performance despite this bottleneck, typically involving addition of some short-term memory into the processing unit. Two main approaches are: adding a shared cache memory for data from all the running threads in a multicore machine, and adding scratchpad memory as in a GPU, to support many more threads (such that many of them can have their state flushed to scratchpad memory while they are waiting for data from the external memory). Exploration of a unified machine model, allowing both these mechanisms, shows a valley of low performance when plotting the throughput of such a machine versus the number of active threads, as shown in Fig. 8.8 [8]. When the number of running threads is small, the

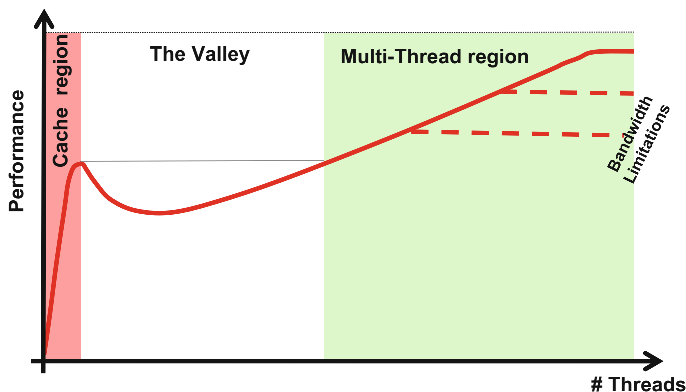


Fig. 8.8 Performance regions in a unified architecture. Caches are efficient up to some number of threads, and they lose their effectiveness if too many threads compete for cache space (the valley). Multi-thread machines employ a very large number of threads to keep the processors busy while data is fetched from external memory

shared cache memory can effectively hold the useful data locally near the computing cores, allowing them to work busily in parallel. However, when more software threads are added, the shared cache loses effectiveness since data from different threads compete for the same cache locations. The performance can be improved at this point by adding many more threads, to keep the cores busy while most of these threads wait on the side for data from external memory. Although this approach is useful for increasing performance, it might be impractical from a power viewpoint, since it violates the locality principle: accessing data from external memory costs much more than on-chip data access (by about two orders of magnitude).

8.2.5 Memory Intensive Systems

Adding local memory in close proximity to the processing cores still seems to be the solution for eliminating the so-called "external memory wall" bottleneck. However, while in the past such memory (cache or scratchpad) had been added in order to improve performance, local memory will be added into future systems in order to improve power efficiency.

Let's define a parameter called memory intensity, which is the throughput of a computer (number of instructions executed per second) divided by the bandwidth (bits per second) of its communication with external memory while running a given software task. By plotting this parameter versus bandwidth, we see that a constant throughput can be obtained at any point on a hyperbolic curve such as those shown in Fig. 8.9. Since data movement is expensive in terms of power, improvements in

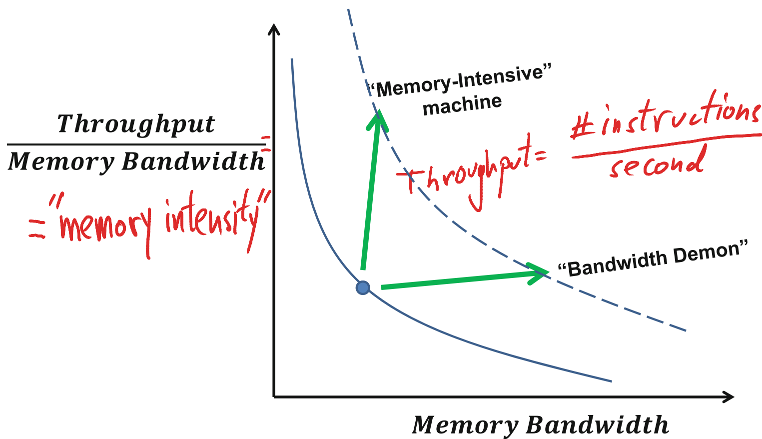


Fig. 8.9 Constant-throughput curve have a hyperbolic shape in the plane of memory-intensity versus memory bandwidth. An existing architecture may be upgraded to a higher throughput by increasing memory bandwidth (which is wasteful in power) or by increasing memory intensity (the amount of memory embedded within the computational hardware)

energy efficiency can be obtained only by moving towards the left on the shown curves. This means that energy efficient machines must have high memory intensity, or in other words they must have a lot of memory embedded in their execution units and thus avoid numerous reads and writes to external memory.

Memory intensive systems are likely to utilize new emerging device technologies, such as resistive memory devices [9], which offer denser and faster memories as compared to SRAM and DRAM. Architecturally, these memories are likely to be specialized function-specific structures using unique addressing schemes. It is highly likely that such special memories will be implemented within special-purpose functional blocks, also known as accelerators, which are optimized for performing tasks such as image processing. For example, in image processing accelerators, pieces of an image are stored in special frame buffer memories to support many local repeated accesses. Another example is a multistate pipeline register (MPR), which can be used to support fast context switching in multi-threading processors. An MPR is a flip-flop which stores an array of state bits belonging to several threads, where the active thread is selected and addressed by its thread number.

An interesting alternative for architecting memory-intensive system is to build some processing capabilities within storage systems. Such logic-within-memory capabilities would eliminate many unnecessary data transfer which occur in present systems and cause major energy waste. However, this kind of architecture requires major revolutions in the structure and organization of software.

8.2.6 Applying the Principles in Large Data Center Computing

Computing technology is still far away from fundamental physical limits to its energy efficiency [10]. The same basic principles, described above for small scale computing systems, can be applied for reducing the energy consumption of large data centers in cloud computing. The trend shown in Fig. 8.1 is very likely to cause a disruption to the evolutionary course of computing technology, driving numerous activities targeted at energy saving. The key principle which must be applied is the locality principle, which implies that data traffic must be minimized. Ideally, any data should be transported only once across a large distance such as a network or external memory interface. Consequently, parallel execution of many operations is called for, and local memory must be provided. Novel memory structures, addressed in various function-specific ways and application-specific modes, should be embedded within execution units of future machines, creating innovative memory-intensive systems. Software and hardware must be configured to minimize waiting states, and efficiently shut down all inactive sub-systems as soon as they run out of work, even for short periods. An example for power saving due to speedup of a common function, which reduces wasteful waiting in the system, can be seen in [11], where 20% efficiency gain is observed due to acceleration of networking. Careful balancing of system bottlenecks and resource contentions must be made in order to minimize the energy waste which typically occurs as a consequence of such events. Finally, there are many architectural opportunities to replace energy-inefficient charge-based electrical signaling over metal wires, by photonic communications, once the technological barriers of interfacing will be overcome.

References

1. http://www.nrdc.org/energy/data-center-efficiency-assessment.asp
2. R. Dennard, Design of ion-implanted MOSFETs with very small physical dimensions. IEEE J. Solid State Circuits **SC-9**(5), 256–268 (1974)
3. F.J. Pollack, New micro architecture challenges in the coming generations of CMOS process technologies (keynote address) (abstract only), in *Proceedings of the 32nd Annual ACM/IEEE International Symposium on Micro Architecture (MICRO 32)* (1999)
4. K. Moiseev, A. Kolodny, S. Wimer, *Multi-Net Interconnect Optimization* (Springer, 2015). ISBN 978-1-4614-0820-8
5. N. Magen, A. Kolodny, U. Weiser N. Shamir, Interconnect power dissipation in a microprocessor, in *International Workshop on System Level Interconnect Prediction* pp. 7–13 (2004)
6. M.T. Bohr, Interconnect scaling—the real limiter to high performance ULSI, in *IEDM* pp. 241–244 (1995)
7. D. Patterson, The trouble with multi-core. IEEE Spectr. **47**(7), 28–32 (2010). doi:10.1109/MSPEC.2010.5491011
8. Z. Guz, E. Bolotin, I. Keidar, A. Kolodny, A. Mendelson, U. Weiser, Many-core vs. many-thread machines: stay away from the valley. IEEE Comput. Archit. Lett. **8**(1) (2009)

9. S. Kvatinsky, E.G. Friedman, A. Kolodny, U.C. Weiser, Memory intensive computing, in *Proceeding of the Annual Non-Volatile Memories Workshop* March (2014)
10. I.L. Markov, Limits on fundamental limits to computation. Nature **512**, 147–154 (2014). doi:10.1038/nature13570
11. Reference to Eitan Zahavi's chapter in this volume (Green photonics 2015)

Chapter 9
Low-Cost Harvesting of Solar Energy: The Future of Global Photovoltaics

Eicke R. Weber, Stefan Glunz, H. -M. Henning, A. Palzer and R. Schindler

Abstract Photovoltaics is currently experiencing a rapid global expansion into the Terawatt age, exceeding even the most optimistic predictions of experts just a few years ago. This is driven by innovations, resulting in higher solar energy conversion efficiencies at ever decreasing costs, and the rapidly increasing global market size. This chapter presents a comprehensive discussion of PV technology developments and issues of integration ever increasing amounts of renewable energy into the electricity grid. Based on this we can look forward to establish a reliable, renewables-based energy supply for the world.

9.1 Introduction: The Needed Transformation of Our Energy System, Limited Fossil Fuels, the Climate Problem

Humankind is facing a problem of unheard dimensions during this century: after 200 years of rapid economic development in the first and second industrial revolution, characterized by harnessing first the energy of coal, utilizing the power of

E. R. Weber (✉) · S. Glunz · H. -M.Henning · A. Palzer · R. Schindler
Fraunhofer Institute for Solar Energy Sytems ISE, Freiburg, Germany
e-mail: Weber@berkeley.edu

S. Glunz
e-mail: Stefan.glunz@ise.fraunhofer.de

H. -M.Henning
e-mail: Hans-martin.henning@ise.fraunhofer.de

A. Palzer
e-mail: Andreas.palzer@ise.fraunhofer.de

R. Schindler
e-mail: roland.a.schindler@gmx.de

E. R. Weber
Berkeley Education Alliance for Research in Singapore BEARS, Singapore, Singapore

© Springer International Publishing AG 2017

G. Eisenstein and D. Bimberg (eds.), *Green Photonics and Electronics*,
NanoScience and Technology, DOI 10.1007/978-3-319-67002-7_9

the steam engine, and then of oil and gas, combined with the introduction of electricity and rapid communication, we are facing the challenge of limited resources: the world will have to change from the past resource-wasting economy to a sustainable economy. This is the only way we can form the basis of human life on earth—as we know it today—in 100, 500, 1000 years, and our challenge is to start this gigantic task during this century.

Many resources are still available to allow us to continue business-as-usual in our wasteful throw-away way of living for many decades, but energy is the first really urgent topic. We have no time to loose to start the transformation to a long-time sustainable energy system right now. This is not only caused by resource availability but more and more by the rapidly mounting climate problem: during more than a million years, the earth experienced an oscillating system of temperatures and CO_2-content of the atmosphere: 200 ppm of CO_2 during ice ages, and 280 ppm CO_2 during the intermediate warm periods. In the last 12,000 years we experienced especially favorable conditions, the stable climate of the holocene [1]. However, since about 1850 the CO_2 concentration started to rise, around 1900 it already exceeded 300 ppm, and today we are already well above 400 ppm. Just in the last 20 years we are starting to feel the very first clear consequences of what is ahead of us: record hot years, droughts, torrential rainfalls, storms with unheard-of wind speed exceeding 300 km/h… we are not yet aware, which of the consequences of the anthropogenic loss of climate stability—induced by our emissions—will be most damaging! We do not know the tipping points of the monsoon cycle, the gulf stream, the El Nino/El Nina pacific weather phenomena, but one fact is sure: once we exceed a critical tipping point, there will be no easy way back, it will be too late, and our children will accuse us: you knew the problem, you had the solution, even at reasonable cost, why in the world did you not act, why did you put your head in the sand?

In other words: among all the needed transformation processes towards sustainability the transformation of our energy system is the most urgent challenge. Our fossil resources would allow at least 50, may be 100 more years of business-as-usual, but the climate issue cannot wait so long, the fundamental changeover has to happen within the next decade!

In order to de-carbonize our energy system we should work on two fronts simultaneously: increasing energy efficiency reduces emissions immediately, as currently our energy is provided by dirty energy; rapid introduction of renewable energy paves the way towards the final de-carbonization of the energy sector. The transport sector is dependent on the way energy is provided and has to be de-carbonized as well, by switching to electric transport, provided by batteries and hydrogen fuel cells.

We should briefly note that energy is never produced, since the big bang about 13 bn years ago, physics is based on the principle of energy conservation. In the following, we will use the term *energy converters* for devices producing e.g. electricity, by converting energy into a form useful for human use.

First, we will discuss the issue of nuclear energy, which in principle would offer an alternative way to de-carbonize the energy sector without relying on renewable energy.

Here we have the choice between fission and fusion reactions: fission of heavy nuclei releases stored energy of inner-nuclear bonds, and fusion of light nuclei does the same, as elements near the center of the periodic table such as lead are most stable. In the fission reaction, large amounts of highly radioactive and toxic nuclei are produced, Plutonium being the prime example. As long as we can operate nuclear fission reactors under well-controlled conditions we might use them for electricity production, though the cost of final storage of long-lived radioactive waste, of decommissioning after the useful lifetime, and the cost of disaster insurance make the real, total cost of this electricity drastically non-competitive. In addition, there is the danger of catastrophic loss of the dangerous products of the fission reaction: we can well predict the behavior of a reactor under conditions we think of in advance, but in principle, human ingenuity is never sufficient to predict ahead of time all possible situations that might occur. Take the 2006 near-catastrophe in Forsmark, Sweden, where just the bold action of a technician against the rulebook prevented a Fukushima-type accident already then!

Fukushima itself did not happen because of damage to the reactor caused by the earthquake or tidal wave, this had been quite well taken care for (just the walls were a bit too low!)—no, it was the unforeseen, simultaneous failure of external electricity supply and both emergency generators—just like in Forsmark—that caused the disaster! So, nuclear fission cannot and should not be used to solve the world's energy problem. Unfortunately, the next big nuclear disaster will come, and with it the possible rapid shut-down of nuclear fission reactors worldwide. Reliance on nuclear fission power actually endangers security of energy supply.

Nuclear fusion is a much different process, that does not result in the cocktail of dangerous byproducts of the fission reaction. However, it is technologically very difficult to control the 100-million degree hot plasma. We can be optimistic, within the next 50 years we will probably succeed in controlling the fusion plasma. However, the question of how to extract useful power out of this super-hot plasma is far from a feasible solution: we do not yet have a suitable material for the inner walls of the central burning chamber, that could withstand facing this super-hot plasma for several years. Secondly, for harvesting the energy released by nuclear fusion we envisage at present heating within those walls tubing with water and steam, to drive 19th century-type steam engines, turbines, with the low efficiency of all thermal power plants! The result is that even if we master the physics problems of stabilizing the hot plasma, and the materials science problems of the wall material, we end up with a thermal power plant that produces electricity at estimated costs of 5 ct/kWh—in 50 years, when solar electricity will be available for 2 ct/kWh or less! Moreover, in 50 years solar and wind based electricity production will be dominant, and able to supply 100% of the electricity needs at many times. Flexible power plants, like (bio) gas-fired plants, will be a useful complement of solar and wind power, not nuclear plants that have to run around the clock at constant power!

Therefore, our future decarbonized energy system hast to rely on two main pillars, harvesting energy for human use from sun and wind, augmented by all other kinds of renewable energy, such as from hydropower, biomass, geothermal energy, tidal and wave energy, to mention just the most important ones. The global wind resources alone would be sufficient to provide all power we need, and the sun provides many thousand times the energy needed by humankind now and in all future!

The energy system of the future will be radically different from today's, and Jeremy Rifkin described it well in his vision of the Third Industrial Revolution [2]: Based on decentralized harvesting of renewable energy plus decentralized information and communication system. Decentralized harvesting of fluctuating renewable energy will pose new kind of challenges to the grid: it has to contain elements of storage and deal with surplus energy available when the sun shines and the wind blows more than needed. Bi-directional grids have to allow distribution of power from where it is locally produced to where it is needed. Local storage allows to achieve energy independence, reducing the needs from the grid. Central storage will augment the system, like dam-based hydro power plants that minimize outflow of water when surplus electricity is available, and maximize outflow when power is needed. If the percentage of fluctuating renewable energy climbs above 50%, large-scale hydrogen production out of superfluous electricity becomes interesting. Hydrogen can be stored, fed into the natural gas grid, or used for fuel cells, in cars or stationary. All of this will be described in more detail in Sect. 2 of this chapter.

Germany shows that it is indeed possible to feed ever increasing amounts of renewable energy into the grid while preserving and even improving grid stability: In the last 10 years, the fraction of renewable energy in the electricity grid more than doubled, today it is at 35%, and in the same time the System Average Interruption Duration Index SAIDI was cut in half, to only 12 min average grid interruption per year today! (Fig. 9.1).

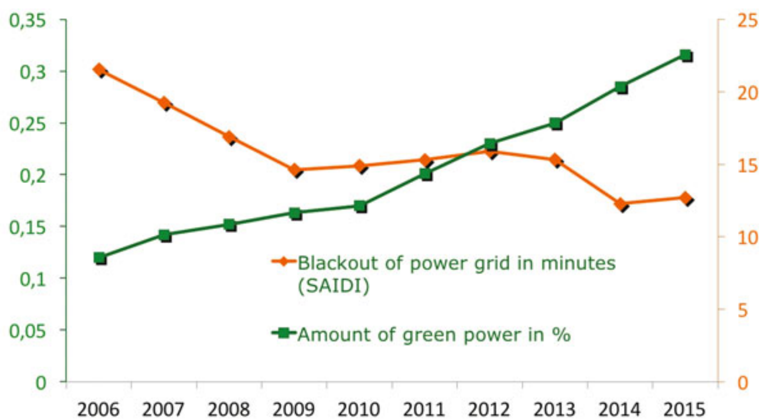


Fig. 9.1 System Average Interuption Duration Index SAIDI and fraction of renewable energy in the electricity grid 2006–2015 [3], graph after Fell [4]

In other words, the electricity grid in Germany has been *strengthened* by the introduction of suitable algorithms for nodal grid control and the maintenance of sufficient reserve power to not only maintain grid stability, but even considerably improve grid stability: a grid that can stabilize the sudden knock-out of a 10 MW PV field by a large cloud is as well much more resilient against all other interruptions, from storms to acts of terrorism!

In the following sections of this chapter, first a comprehensive simulation of a german energy system based primarily on renewable energy is presented as an example, followed by a discussion of the main technologies currently used to harvest solar energy through photovoltaics (PV) on a commercial scale, followed by a concluding section on the future global prospects of PV heading towards the Terawatt scale.

9.2 The Role of Photovoltaics in Our Future Energy System, Based on Simulation of the German Energy System for 80% and More of Renewable Energy

During the past decade, Germany has taken a leadership position in the market introduction of renewable energy technologies, that was instrumental in bringing down its cost dramatically. The driving force of the energy transformation in Germany is the political goal of drastically reduced greenhouse gas emissions, in order to limit the anthropogenic climate change and thus the danger of drastic influences on nature and the conditions of human life and economy. The German Government was one of the driving forces behind the COP-21 accord, that has been signed in November 2015 in Paris. The declared goal is to decrease the greenhouse gas (GHG) emissions of Germany by 2050 by at least 80% [5] and wherever possible by 95%, below 1990 levels [6, 7]. This objective is supported by a wide social consensus in Germany. The total amount of GHG emissions in the reference year 1990 amounted to 1,215 million tons of CO_2 equivalent (for this purpose, all greenhouse-relevant effects are converted into the climate-changing effect of CO_2 emissions). This value considered the CO_2 lowering in agriculture and forestry. For the years prior to 2050, reduction target values are defined as well: a reduction by 40% by 2020, by 55% by 2030, and by 70% by 2040.

The amount of the German GHG emissions in the past is presented in Fig. 9.2, together with the target values intended for the period up to 2050.

The largest share of GHG emissions are energy-related CO_2 emissions with close to 990 million tons in 1990 and 793 million tons in 2013 (see green bars in Fig. 9.2). Thus, energy-related CO_2 emission may be at maximum 198 million tons in 2050 in order to achieve the reduction goal of 80% compared to the reference year 1990. Here it is assumed that energy-related CO_2 emissions are reduced to the same extent as all other greenhouse gas emissions. A reduction by 95% would mean a target value of 49 million tons. The relative portion of energy-related CO_2

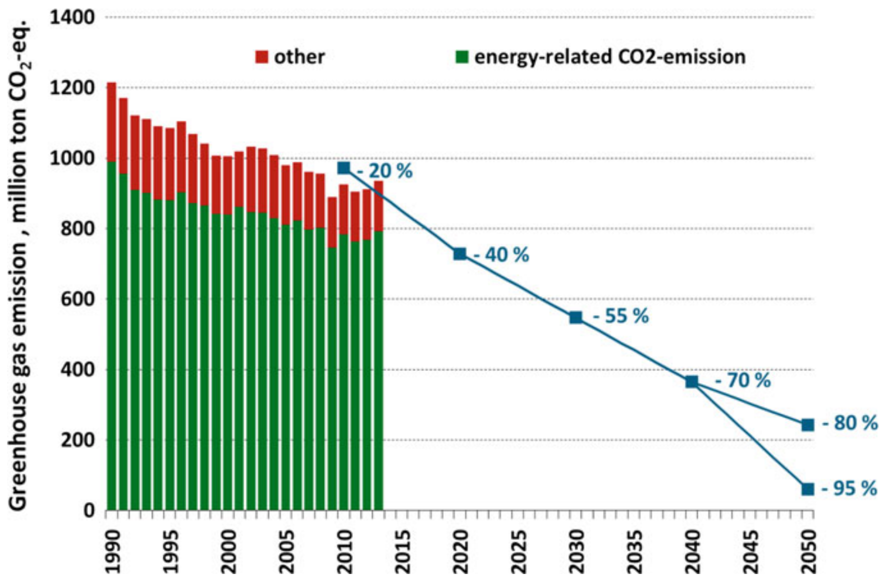


Fig. 9.2 Greenhouse gas (GHG) emissions in Germany from 1990 till 2013, and target values up to 2050 (*blue dots*). *Green bars* represent energy-related CO_2 emissions, *red bars* other GHG emissions, *cf.* data in [8]. Percentage reductions refer to the 1990 value

emissions in the total greenhouse emissions has increased from around 80–85% in the period from 1990 to 2013.

To investigate how these boundary conditions influence a future German energy system, Fraunhofer ISE developed a national energy system model called Regenerative Energy Model for Germany: REMod-D [9–11].

The basic functionality of the REMod-D model is founded on a cost-based optimisation of a German energy supply system whose energy-related CO_2 emissions do not exceed a specified target value and/or target pathway. The optimisation target is to dimension all generators, converters, and consumers at minimum cost, such that the energy balance of the overall system is met at every hour. Conventional power plants with lignite and hard coal as fuel, nuclear power plants, oil-fired power plants, gas turbines, CHP plants, and gas-fired and steam power plants are implemented as generators. Renewable energy can be supplied in the model using wind turbines onshore and offshore, photovoltaic systems and hydropower plants (Fig. 9.3).

Biomass can be used in different usage pathways either directly or after conversion into a different energy carrier. For example, wood can be used in boilers, in order to provide process heat for industrial applications and for the generation of low-temperature heat in the building sector. Biogas systems, gasification systems with subsequent synthesization into hydrogen, methane, or liquid fuels and biodiesel systems are implemented as systems for the conversion of biomass. Electrical energy storage systems in the form of stationary and mobile (in vehicles) batteries

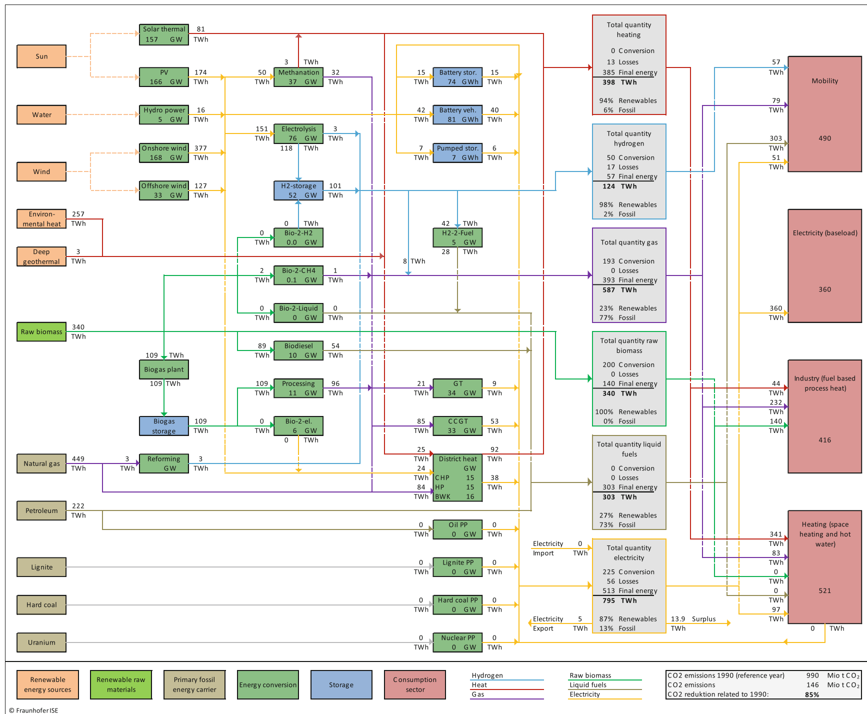


Fig. 9.3 Schematic of the energy system as presented in the REMod-D simulation model. The illustration shows all conversion pathways of fossil primary energy and/or renewable energies up to the respective consumption sector (Values are examples for a scenario with a reduction of energy-related CO_2 emissions by 85%, *cf.* [11])

or pumped-storage power plants are used as storage systems. Hydrogen storage systems and thermal hot water storage systems in different orders of magnitudes are considered as well. With respect to methane storage system, the simplified assumption is made that currently already existing storage capacities (including grid, ca. 210 TWh [12]) will also be available to the system in the future. They are not changed in the optimisation [11].

The energy demand side is divided into four groups, according to the different fields of use: Mobility, intrinsic electricity applications, heat for buildings (residential, non-residential, and industrial buildings), and process heat in the industry. The mobility sector is mapped in detail concerning passenger cars and trucks, with seven vehicle concepts each. The energy demand of aviation, shipping, and fuel-based railway traffic is considered in the balance, however, without temporal resolution. The basic electricity load is mapped using load profiles based on the data of European transmission grid operators that was reduced by the current load for heating systems. This load is calculated model-endogenously and is not included in the basic load [11].

The building sector is implemented with 18 possible heat supply technologies. Each of these heating technologies can be optionally supplemented with a heat storage device or solar thermal energy system. Figure 9.4 shows an example of the "electrical brine heat pump" system, hence, a brine-water heat pump with the ground as heat source. The possible energy flows between the individual system components are shown. Thermal storage devices can be charged through solar thermal energy, as well as through heat from excess power (directly or via the heat pump). The latter option allows the flexible use of power in the case of a negative residual load. Vice versa, the heat pump can be switched off and the heat storage discharged in the case of a positive residual load and simultaneous heat requirement.

The energy demand of the industry is derived from the statistical data of the German Federal Ministry of the Economy [13] and refers to the fuel-based energy supply for process heat. The electricity demand of the industry is included in the basic electricity load [11].

The REMod-D model is based on simple physical models of all components contained. The central component is the energy exchange across the electricity system. A load still to be covered after feed-in of renewable energy is compensated by the generation of electricity from systems of different sectors. Excess electricity, on the other hand, can be stored and/or converted into different electricity forms (chemical and thermal) and is thus accessible for all sectors. The operation of electricity-generating and electricity-using systems in the case of positive and/or negative residual load follows a defined management strategy. The component usage sequence in this management strategy follows the path of highest energy efficiency at simultaneously lowest CO_2 emissions. Figure 9.5 shows the different stages for the generation and usage of electricity in the case of a positive and/or negative residual load in the system.

To cover positive residual loads, CHP plants are operated first after the use of electrical storage systems and biogas CHP. The generated heat is then used to charge heat storage devices and/or to cover thermal loads if these are present at the same time. Any additional demand is covered by the operation of combined cycle gas turbine (CCGT) plants and CHP plants in "electricity only mode". The remaining load is covered by highly flexible gas and oil turbines and with the help

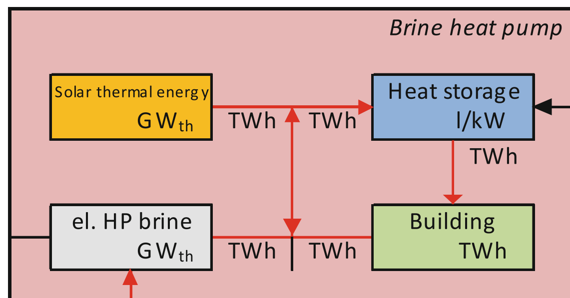


Fig. 9.4 Schematic design of heating systems based on the example of a ground-coupled, electrical heat pump (HP) (*red lines* = heat, *black line* = power) [10]

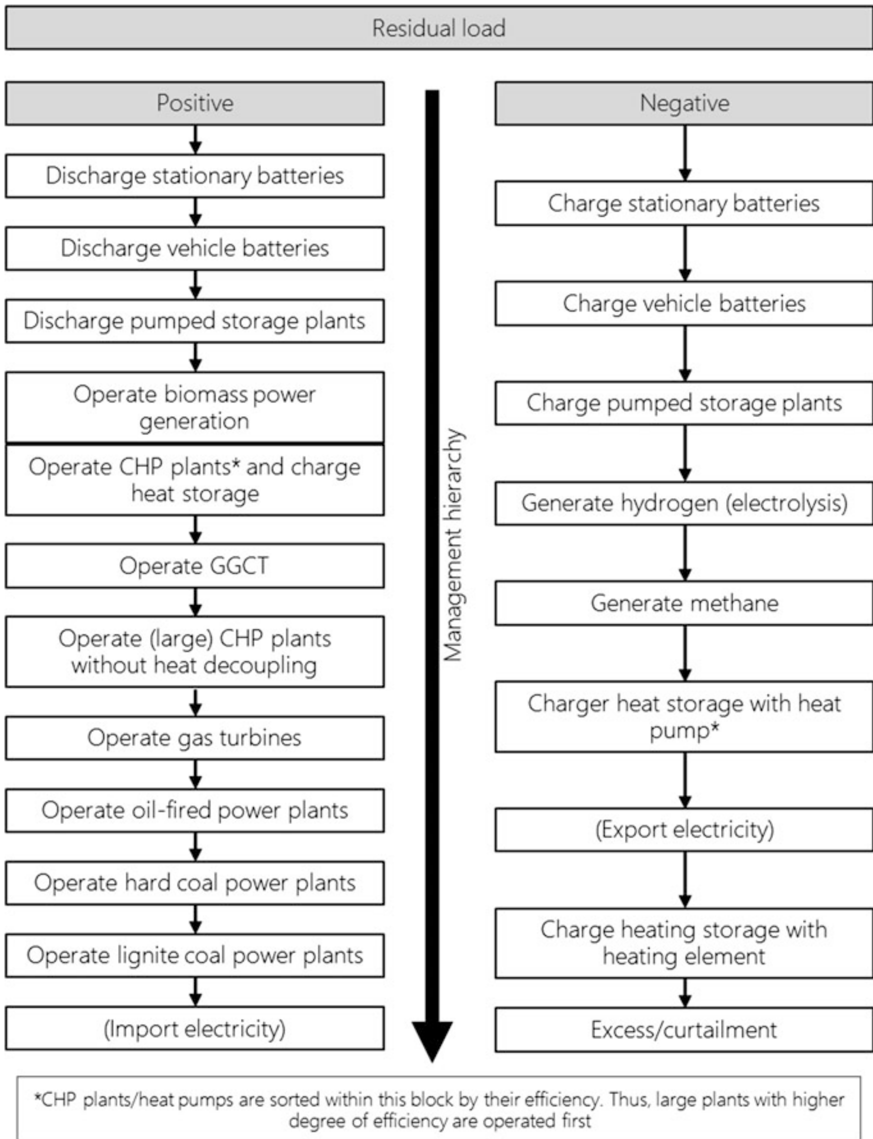


Fig. 9.5 Management sequence in the case of positive (*left*) and negative (*right*) residual load [11]

of the remaining, flexibly usable power of conventional lignite and hard coal power plants. In model calculations that also investigated the electricity import, it can also contribute to covering the electricity demand at the end of the usage cascade with a previously defined maximum power [11].

The weather influence is decisive for the different residual load states during the simulation. To map this influence adequately, three real data records of 2011–2013 were used within the scope of the calculation. The weather data used in the model for the calculation of feed-in and load profiles are based on publicly accessible data of the German Meteorological Service [14, 15]. Weather data from two different reference locations in Germany, Braunschweig for North Germany and Würzburg for South Germany, were processed in the model. Hourly outside temperature values and emission data are used from both locations. To consider the stochastic effects, the available weather years of 2011, 2012, and 2013 are randomly distributed to the period from 2014 to 2050 at the beginning of the calculation for the weather calculation in the investigated period from 2014 to 2050. Every iterative calculation of a transformation pathway within an optimisation uses this sequence always in the same form. This way, a consistent data record was generated for the entire period from 2014 to 2050, which is used as basis for the electricity generation from solar energy and wind, which is used for calculation of the heating load of the building sector, and which is used as the basis for the heat generation from solar thermal energy systems. The real profile of the power consumption data of these years is also used in the same sequence to ensure an adequate correlation between the profile of the basic electricity load and the profile of the power provision from renewable sources, which is defined by the profile of the meteorological variables [11].

Technology specific system costs are obtained from an exogenously specified cost function depending on the investigated year. In determining this cost function, the values of every technology for start year 2015 and target year 2050 were used as start and end value. Different data sources were used for this purpose, respectively. The expected curve profile of the costs of photovoltaic systems is presented in Figure 9.6 as an example. The *curve* profile is based on studies discussing the cost degression behaviour of the respective technologies. As result, a specific cost value in €$_{2013}$/kW is available to the model for every year.

In addition to the cost investigations for individual components, such as converters or storage devices, costs for energy saving measures in the building sector through energy renovations are considered as well in the model.

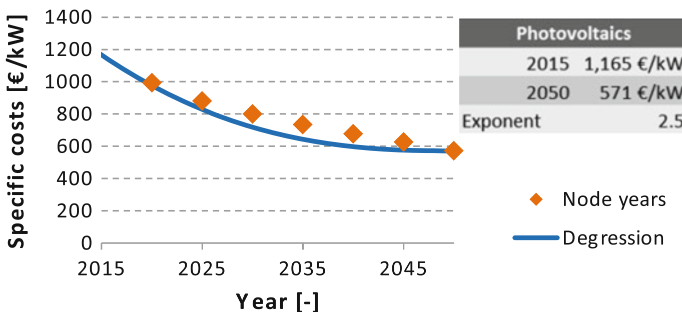


Fig. 9.6 Expected cost profile of photovoltaic systems up to 2050, *cf* [11], based on [16]

The influence of renewable energy converters like wind energy converters or photovoltaics is investigated in the third version of the energy system related study of Fraunhofer ISE [10]. It provides different future scenarios for the German energy system. The respective scenarios and their assumptions relative to the target value of energy-related CO_2 emissions, energy renovations, the vehicle sector composition used, and the exit from coal-fired electricity generation are summarised in Table 9.1. Every calculation contains a complete optimisation calculation for the given development of energy-related CO_2 emissions, respectively, to identify the transformation pathway leading to a transformation cost minimum under the given assumptions and definitions.

First, optimisation calculations were performed for the five different mobility scenarios under the same assumptions for all other boundary conditions described above. The target value of energy-related CO_2 emission was set to minus 80% relative to the 1990 reference value. Low energy renovations were assumed as well as the operation of coal-fired power plants until 2050. The "Mix" mobility scenario was then always defined for all other optimisation calculations. This should compensate for the uncertainty regarding the development of the vehicle sector. Furthermore, an ambitious energy renovation of the building sector and an accelerated exit from coal-fired electricity generation were assumed for calculations with the CO_2 reduction targets of 85% and 90%, as otherwise these stricter climate

Table 9.1 Overview of the investigated climate protection scenarios. As to coal-based power plants, accelerated phase-out till 2040 is compared with a non-accelerated scenario based on regular decommissioning [10]

Seq. no.	Target value CO_2	Energy renovation buildings	Vehicles	Coal plants	Abbreviation
# 1	−80%	Low	Classic	Not accelerated	80/low/class/ n.a.
#2			CH_4	Not accelerated	80/1ow/CH_4/ n.a.
#3			H_2	Not accelerated	80/1ow/H_2/ n.a.
#4			Electric	Not accelerated	80/low/ eletric/n.a.
#5			Mix	Not accelerated	80/low/mix/ n.a.
#6		Ambitious	Mix	Not accelerated	80/amb/mix/ n.a.
#7				Accelerated	80/amb/mix/ acc.
#8	−85%	Ambitious	Mix	Accelerated	85/amb/mix/ acc.
#9	−90%	Ambitious	Mix	Accelerated	90/amb/mix/ acc.

protection targets could not be achieved purely mathematically due to the permitted CO_2 quantities [10].

The results clearly indicate that renewable energy converters—to a large extent photovoltaics—are much needed to decrease the systems energy related CO_2-emissions. Actually, with at least 300 GW of combined energy converters from photovoltaics and wind, a multiple of the peak electricity demand (currently 80 GW in Germany) will have to be installed. The composition of the most important converters of renewable energies for the electricity generation, i.e., onshore wind energy converters, wind energy converters in the North and Baltic Sea (offshore), and photovoltaics for electricity generation for the nine investigated scenarios, is shown in Fig. 9.7. This figure shows the installed power expected in 2050, i.e., the target year investigated here.

When looking into the first five scenarios with different mobility concepts, while all other boundary conditions remain the same, it can be observed that the installed power of wind turbines and photovoltaic systems is significantly smaller in the scenario with a high portion of electric vehicles (scenario 80/low/electric/n.a.) than in the other four scenarios. The total installed power for wind turbines and photovoltaic systems is here slightly above 350 GW, while the values for the other four scenarios are slightly above 400 GW. This is explained by the higher efficiency during conversion of final energy (electricity) into useful energy (traction) of battery-electric motor drives compared to all other drive concepts.

The effect of an ambitious building energy renovation gets obvious in the comparison between otherwise similar scenarios with a CO_2 reduction of 80%, drive concept mix for mobility, and without accelerated exit from coal-fired

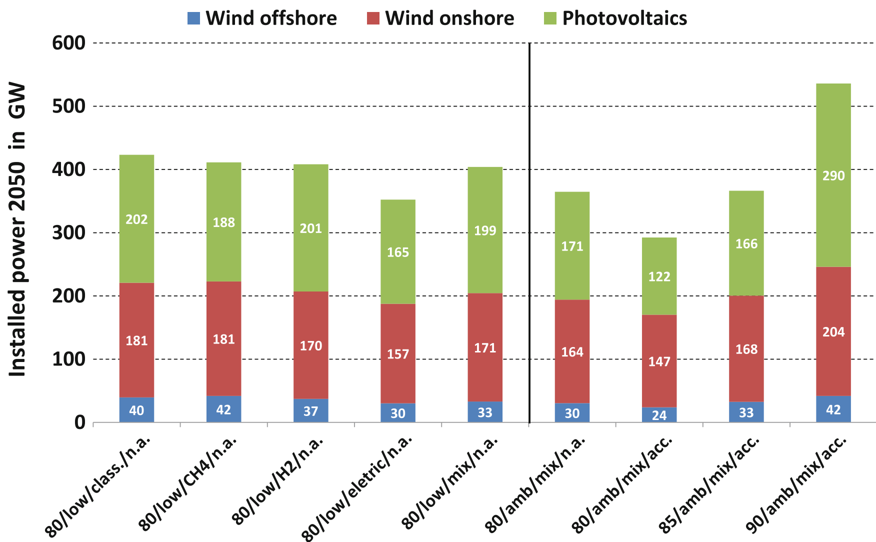


Fig. 9.7 Expected installed power of the most important converters of renewable energies (solar, wind) in 2050 for the investigated scenarios [10]

electricity generation (80/low/mix/n.a. and 80/amb/mix/n.a.). Significantly, the scenario with ambitious building energy renovation results in a significantly lower need for installed power of wind and solar converters. Here, the installed power is at ca. 365 GW in 2050, while the value for the scenario with lower building energy renovation is above 400 GW. An even more significant reduction is possible, if the exit from coal-fired electricity generation is accelerated and realised by 2040. In the case of an ambitious building energy renovation and a drive concept mix for motor vehicles, the required installed power of the power converters from solar and wind energy is at around 290 GW installed power in 2050 (80/amb/mix/acc. scenario). All scenarios discussed so far lead to a reduction in energy-related CO_2 emissions by 80%. A significant increase in the installed power of renewable energy converters is required if the CO_2 emissions should be reduced further, i.e., by 85%. In this case, the total installed power for solar and wind-based electricity generation amounts to 412 GW—or by 90%, with a total needed power of 536 GW.

The development of the expansion of wind turbines and photovoltaic plants is shown in Fig. 9.8 [10]. A mostly constant increase can be recognised over the entire period. However, it must be observed that we used upper limits for the capacity that reasonably can be net-added per year.

The amount installed per year is partly significantly above these limits, as replacement installations become necessary, as presented in Fig. 9.9 [10]. Here can be seen that a first significant amount of replacement installations for onshore wind turbines is required in the 2020s, and a second amount starting in the mid-2030s, and in the last 6 years from 2044 to 2050. In the case of photovoltaic plants, a phase

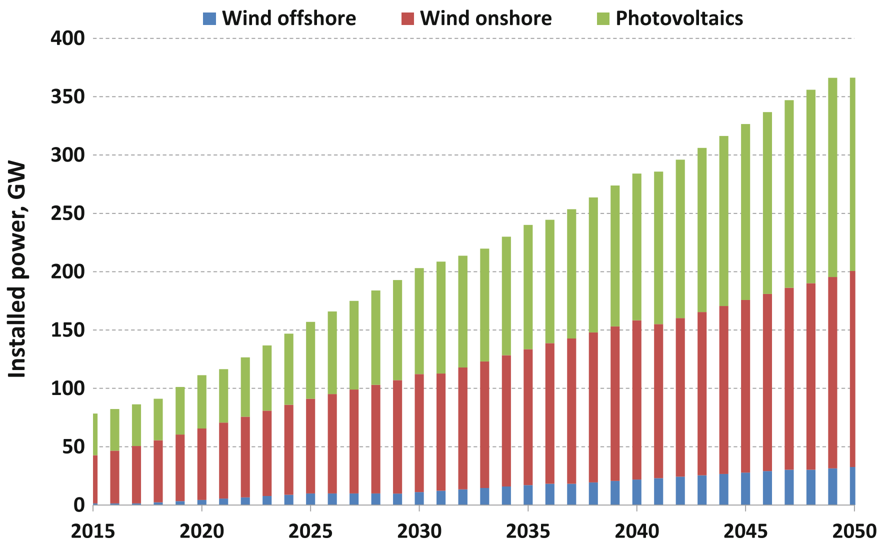


Fig. 9.8 Total installed power of wind turbines and PV plants in the 85–% scenario [10]

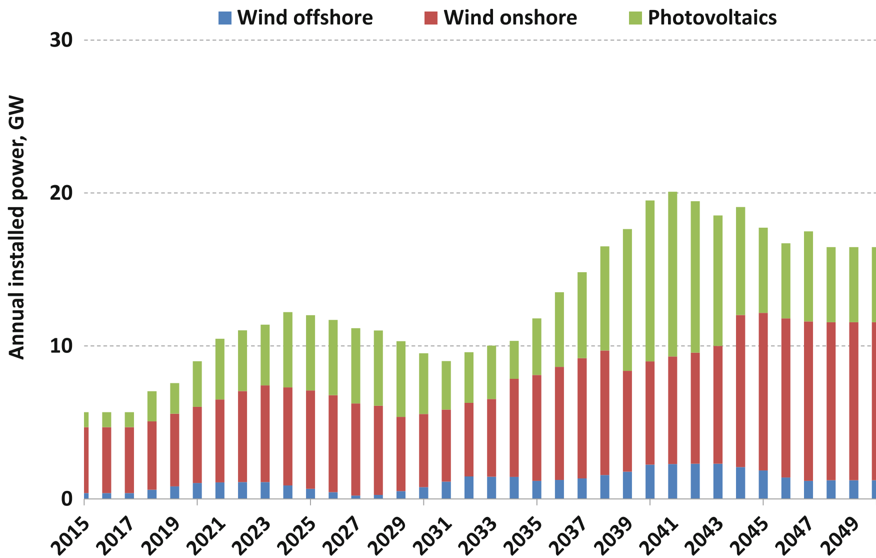


Fig. 9.9 Development of the yearly installed gross power (i.e., including replacement installations) of wind turbines and photovoltaic plants [10]

with high replacement installations occurs in the late 2030s and 2040s. Here, the many plants installed in 2010 to 2013 must be replaced.

These results demonstrate that a large amount of renewable energy converters, especially photovoltaic systems, are needed to fulfil the CO_2 emission reduction aims for a future German energy system. Depending on the assumed scenario boundary conditions, between 122 and 290 GW_{el} of PV will be needed in 2050 in Germany alone.

9.3 Crystalline Silicon Photovoltaics

Photovoltaic energy converters are dominated by crystalline Silicon technology. Fig. 9.10 shows that just 20 years ago thin film technologies enjoyed a 30% market share, but since 2010 the market share of all thin-film technologies dropped even below 10%.

Silicon has several advantages: It is non-toxic and abundantly available in the earth's crust. Crystalline silicon-based photovoltaic (PV) modules have proven their long-term stability over decades in the field and not only in accelerated module tests. The rapid technology development in this field has been recently reviewed in detail by Glunz [18].

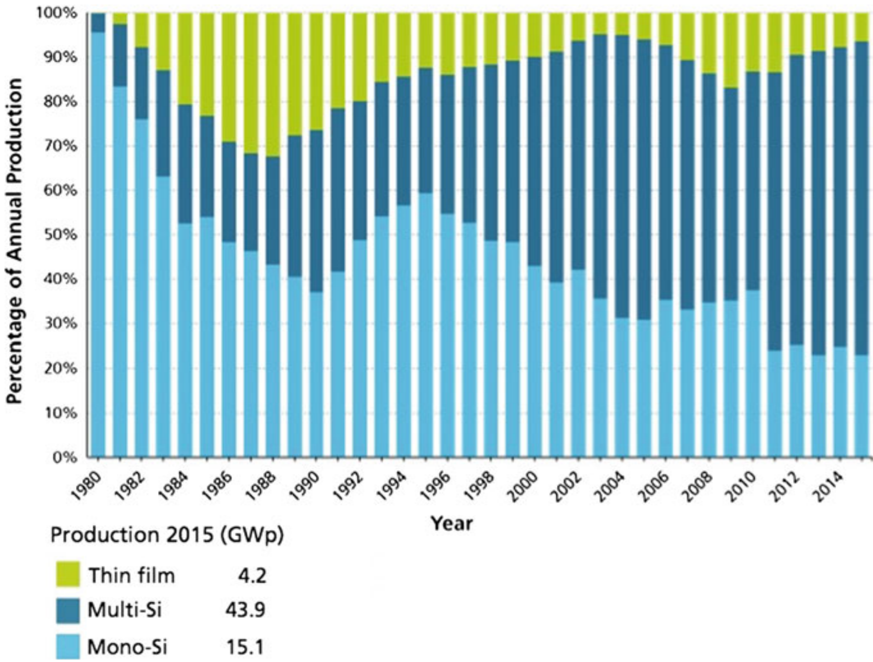


Fig. 9.10 Market share of PV technologies since 1980 [17]

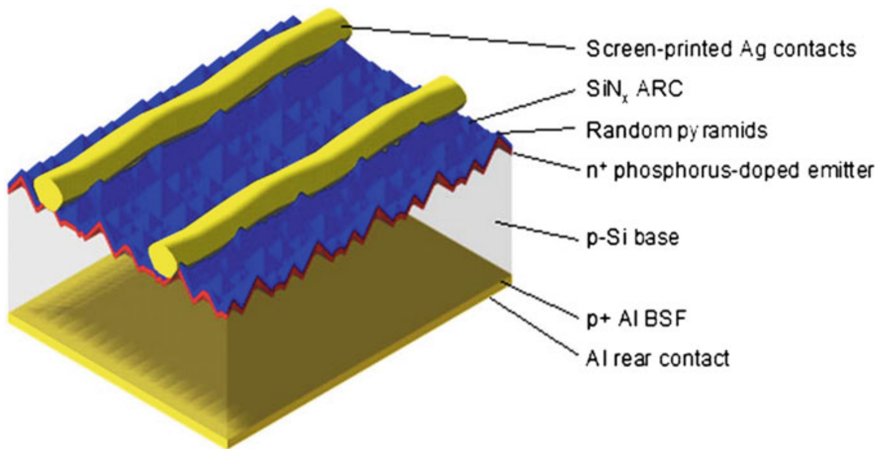


Fig. 9.11 Structure of a screen-printed Al-BSF solar cell based on a p-type Si wafer [18]

9.3.1 Al-Back Surface Field Technology

Concerning crystalline Si PV technologies, the market is clearly dominated by a rather simple approach: screen-printed aluminium back surface field (Al-BSF) cells on p-type silicon substrates, with in-diffused frontside n^+ emitter and the backside covered by an Al rear contact that gives rise to a back surface field, rejecting minority carriers from the highly recombinative full-area back Al contact, see Fig. 9.11.

In today's standard process sequence, the Al backside is created by a rather elegant alloying process of screen-printed Al-paste with the base silicon [19]. This alloying process takes place during a very short firing process in an inline belt furnace. During the cool-down phase the silicon that has been dissolved into the molten aluminium, recrystallizes and aluminium is incorporated in the silicon lattice, according to its solubility at the actual temperature.

In the last 20 years, most of the impressive increase in cell efficiency has been due to evolutionary improvements of this technology. Improved metal pastes and printing processes, increased emitter sheet resistance and better front surface passivation layers (SiN_x by plasma-enhanced vapour deposition, PECVD [20]) are a few examples of the changes which have been responsible for the increase in average monocrystalline silicon efficiency limits over the years, with values at around 14% in the 1990 s to more than 19% today.

One of the many strategies for such incremental improvements is known as selective emitter or two-step emitter, whereby the region under the front contact is highly doped, to facilitate good contact properties, and the region in-between the contacts is lowly doped ('shallow') to reduce light absorption and thus improve the blue response and reduce the dark saturation current. The highly doped emitter under the contacts ('deep' emitter) has a second advantage of 'shielding' the highly recombinative metal/silicon contact and additionally reduces recombination. By using such a selective emitter structure, the conversion efficiency of Al-BSF PV cells can reach a value of 19% [21].

9.3.2 Partial Rear Contact Technologies (PRC, PERC, PERL, PERT)

Even with this approach and improved rear side Al pastes, the performance of the Al-BSF structure is limited, as the rear side recombination still is high, due to the metallization layer covering the full backside. However, majority carriers in Si have a high mobility and therefore can be collected at individual contact points rather then metallizing the whole wafer backside. This allows to deposit a good dielectric rear side passivation, leaving local contact points or lines (partial rear contact, PRC). The PRC structure is the basis for a variety of successful cell architectures like the passivated emitter and rear cell (PERC) [22], the passivated emitter, rear

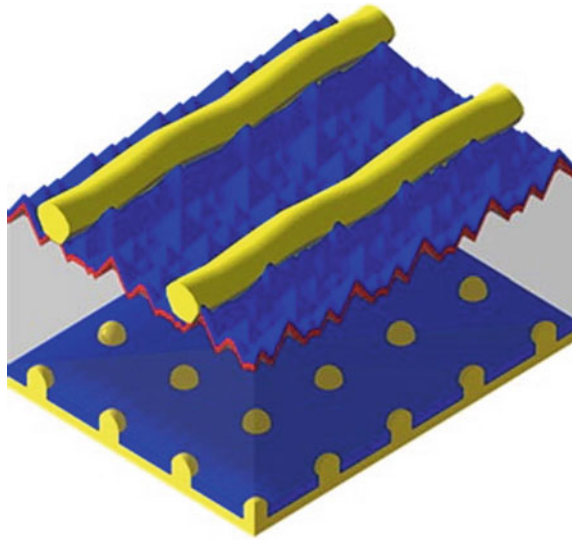


Fig. 9.12 Structure of a PRC silicon solar cell with surface passivation (*blue*) at front and rear surface. Only small point-like metal points (*yellow*) form the base contact at the rear [18]

locally-diffused (PERL) [23] or local back surface field structure (LBSF) [24] and the passivated, rear totally diffused cells (PERT) [25], *cf.* Fig. 9.12.

Although this cell structure achieved excellent efficiencies already as early as 1989 [22], it took 20 years until the transfer into industrial production was performed [26–29]. Average efficiencies of 20.6% were recently shown using industrial production equipment [30]. The PRC structure can be easily combined with the so-called metal-wrap through (MWT) cell architecture [31].

Obviously, the surface passivation layer plays a crucial role for this cell structure. Traditionally, the preferred layer type was thermally grown silicon dioxide (SiO_2). This approach is well known in MOS technology, and was successfully introduced to photovoltaics in the 1980s [32]. The reduction of surface recombination velocity in this case is mainly due to the low interface state density at SiO_2/Si interfaces.

Another route to reduce the surface recombination is to incorporate fixed charges in the passivating dielectric layer. This leads to a band bending in the silicon bulk, and a strong asymmetry of electron and hole density at the surface. Strong surface recombination is observed when similar amounts of electrons and holes are present. Therefore both, strong inversion (high population of minority carriers) or accumulation (high population of majority carriers) lead to a strong reduction of surface recombination. Although both inversion and accumulation lead to an excellent passivation quality, it was shown that in a finished solar cell accumulation is indeed the better option. Therefore, for solar cells on p-type silicon, layers with a high negative charge density are preferred. While SiN_x shows a positive charge, aluminum oxide (AlO_x) layers exhibit a strong negative charge [33] and are now the preferred passivation layers for PRC cells [34, 35]. They can be deposited by atomic layer deposition [34, 36, 37], by plasma-enhanced chemical vapour deposition [38] or by sputtering [39]. In many cases, this AlO_x layer gets an additional

SiN_x top layer to protect itself from damage by the aggressive printed Al rear metal paste [29, 35]. Further effective passivation layers are oxide–nitride stacks, which have led to very good efficiencies in industrial production [40].

The local contact points in PRC technology, which cover approximately 1% of the rear surface, are fabricated in the laboratory using photolithography. This process is too expensive for industrial PV production. One possibility is to open the contact points in the passivation layer by laser [35, 40, 41] or by printing etching pastes [42].

Then an Al paste is printed on the full rear surface and an alloying step is applied. This creates local Al-BSFs which reduce the recombination at the contact points. A second possibility to form local contacts is to leave the passivation layer untouched before the following metallization step. After the Al metal is either screen-printed or evaporated onto the rear, a laser is used to drive the metal through the passivation layer into the silicon. During this so-called laser-fired contact process [43], an effective local Al-BSF is formed as well [44, 45]. A new interesting route is to make use of commercial Al foil as the rear electrode, thereby avoiding complex equipment like screen printers or evaporation systems [46].

All these cells are normally fabricated on boron-doped Czochralski-grown (Cz). Unfortunately, this material contains high concentrations of oxygen and is known to suffer from a metastable defect [47–51] related to boron and oxygen, which is activated by illumination or carrier injection (light-induced degradation) [52]. Thus, the bulk minority carrier lifetime is limited to such an extent that the efficiency is reduced to 20% [26, 29, 53]. If the defect is deactivated [54] efficiencies of around 21% on large area [35] can be achieved.

There are several options available to reduce light-induced degradation: the use of thinner wafers to improve the ratio diffusion length/cell thickness [55], decreasing or avoiding the boron doping or oxygen contamination [49, 56], or the application of the regeneration process [54, 57]. A strong reduction of the oxygen concentration to values below 1 ppma results in a perfect suppression of light-induced degradation. Magnetic Czochralski silicon (MCz) contains low oxygen concentrations due to the suppression of melt convection, and has shown a very high efficiency potential [60, 93, 94]. Another material, PV- float-zone (FZ) silicon, also with a negligible oxygen concentration, was considered a few years ago to be introduced into the large scale PV production [58].

This material type has shown very high and stable carrier lifetimes, but unfortunately is presently not available from mass production. A third material type with low oxygen concentration is cast silicon, which results in multicrystalline silicon with a columnar crystal structure. For multicrystalline silicon, excellent minority carrier lifetimes have been measured [59]. This superior material quality is mainly due to reduced sensitivity to the most relevant impurities such as iron [60, 61].

In order to avoid boron doping, alternative acceptors such as gallium can be used. In fact, cells from gallium-doped Cz-silicon show no degradation [62]. The only issue occurring with this material might be the large variation of doping concentration over the ingot due to the low segregation coefficient of gallium. Nevertheless, adapted cell structures show excellent results over a wide doping range [62].

9.3.3 Solar Cells on n-type Si

Boron can also be avoided as a dopant if n-type silicon with phosphorus doping is used. n-type Czochralski-grown material shows no light-induced degradation even in the presence of a significant oxygen concentration [50, 63]. Due to these excellent material properties, solar cells on n-type silicon are currently at the center of interest of research and industry. Obviously, the widely used phosphorus diffusion [64] cannot be used to create the junction of such cells anymore. One of the following techniques is commonly used to create the p–n junction on n-type substrates:

- Boron diffusion [65, 66].
- Al alloying at the rear side of the cell [67–70].
- Heterostructure using p-doped amorphous silicon [71].

If the boron emitter is used on the front side, a simple cell p^+–nn^+ structure can be created by using phosphorus diffusion at the rear side. This is the analogue to the Al-BSF cell on p-type silicon with inverted polarities. An issue when working with boron-doped emitters is the fact that the well established SiN_x PECVD passivation layer, which is used for phosphorus-doped emitters, is not well suited anymore [72–74]. This is due to the fact that SiN_x features a positive built-in charge which results in a majority carrier depletion or inversion in the boron emitter [75]. On the other hand, negative charges have improved the performance of boron-doped emitters. Therefore, a negatively charged dielectric passivation layer would be beneficial. AlO_x was proven to be an ideal layer to improve the passivation and to reduce the dark saturation current of boron emitters [33]. Using such a layer for high-efficiency solar cells on n-type silicon resulted in very high efficiencies of up to 23.9% [53, 75].

However, process simplifications are mandatory in order to transfer this cell type into industrial production. An elegant process to create the local phosphorus back surface fields (BSF) is the so-called PassDop process [76]. After depositing the phosphorus-containing PassDop layer on the rear side using PECVD, a laser is used to open the contact points in the dielectric passivation layer. The phosphorus contained in the layer is simultaneously driven into the silicon and forms a very effective local BSF. Efficiencies of 22.4% have been achieved on small areas.

9.3.4 Heterojunction Solar Cells

Even for advanced cell structures with small contact areas and superior dielectric passivation, a large fraction of charge carrier recombination takes place at the surfaces. For the 23.9% n-type PERL cell described in the previous section, about 66% of the front recombination happens at the very small front contacts [77]. This recombination channel can be reduced by high-low junction as diffused or alloyed back-surface fields [19, 78].

The most effective way to suppress the recombination at the metal semiconductor interface and thus to fabricate a so called passivated contact is in the application of a heterojunction, i.e., through the deposition of a semiconductor material with wider bandgap on the crystalline silicon base [79]. The most common approach to realizing such a structure is the deposition of doped amorphous silicon [80–84]. An important feature of the best cell structure is the introduction of an intrinsic amorphous silicon layer between the crystalline bulk and the n- or p-doped a-Si layer [82]. Such cells are well-known under the name HIT and have been optimized by Sanyo/Panasonic to a superior degree [85].

Since the surfaces of this structure are perfectly passivated, it is advantageous to reduce the cell thickness. In fact, the best efficiencies were obtained on n-type wafers with a thickness of only 98 μm (Fig. 9.13).

Due to the high efficiency potential of this cell structure, strong research and development activities have been launched in various companies and research laboratories [87–93]. Typical efficiencies achieved by these groups are in the range between 21 and 22%. An excellent overview of this topic was recently given by de Wolf [86].

A major issue of this cell type is photon absorption in the amorphous silicon layer. Although these layers passivate the surface perfectly for carriers which are generated in the crystalline silicon bulk, the diffusion length in the material is too small so that carriers which are photogenerated within these layers recombine with a high probability. This leads to a rather low blue response of these cells [94]. The recombination losses are strongly increased due to such 'parasitic' absorption in the front region (p-doped and intrinsic a-Si) and also the optical losses are high because light is absorbed in the top TCO layers. Another way to reduce the optical absorption of cells with heterojunctions is to place all contacts at the rear side of the cell i.e., interdigitated back junction back-contact solar cells. Interdigitated back-junction back-contact cells were originally developed for concentrator applications [32], since the grid can be designed for high current densities without

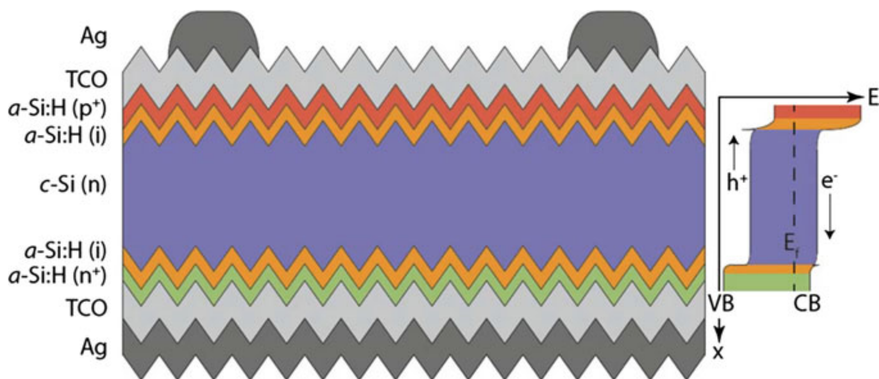


Fig. 9.13 Structure of a heterojunction solar cell and the related band diagram Taken from De Wolf et al. [86], courtesy of De Gruyter

increasing shadowing losses. They have been used mostly for special applications such as solar car races [95] up to the beginning of this millennium, when SunPower entered mass production [96]. The current world record of 25.6% has been established with heterojunction combined with an interdigitated back contact cell architecture [85].

In an alternative, new approach called TOPCon [97], the passivated contacts are actually realized by the growth of a ultrathin tunnel oxide layer and the PECVD deposition of a thin, highly doped silicon layer. Simulations show that the thickness of the oxide is rather critical, as these contacts use tunneling carriers for contact formation [98]. If the oxide is too thin, the passivation quality will be reduced, while thicker oxides hinder the majority carrier tunnel transport. In fact, the thickness of the tunnel oxide should be below 1.5 nm. Such oxides can be fabricated using wet-chemical [97] or UV/O_3 growth [99]. After the deposition of the silicon film, a high-temperature anneal and hydrogen passivation is performed to tune the morphology and the band gap of the silicon film. In this way it is attempted to combine the advantages of a heterojunction a-Si/c-Si structure [100], i.e. excellent carrier selectivity, with the ones of a classical poly-silicon-contact [101], i.e. higher temperature stability. For anneal temperatures between 700–900°C, a contact resistance of below 5 mΩ cm^2 and $J_{0,cont}$ below 10 fA/cm^2 were achieved. The lowest value achieved so far is 7 fA/cm^2 for an *n*-type contact. Record efficiencies of 25.1% for cells with contacts on both sides have been achieved which demonstrates the potential of this approach [102].

A remarkable fact resulting from the very effective suppression of the other recombination channels like surface and contact recombination is that now the intrinsic recombination mechanism, i.e., Auger-recombination, becomes important. This also shows that we are approaching with this new cell technology the physical limits of c-Si cell performance.

9.3.5 Crystalline Si PV Beyond the Shockley-Queisser Limit

The fundamental performance limit for a Si solar cell has been calculated already in 1961 by Shockley and Queisser to be about 29.4% [103]. Using the AM1.5 spectrum and assuming no light concentration, the limit for a single junction solar cell is 33%. Taking into account the unavoidable Auger recombination the maximum value for crystalline silicon solar cells was determined to be 29.4% [104]. With the current efficiency record of 25.6% for the Sanyo/Panasonic HIT cell [85] current PV technology is approaching this fundamental limit for a single-junction Si solar cell. How can we expect to exceed this limit?

As we will discuss in the next paragraph, using c-Si is perfect only for photon energies near the Si band-edge of 1.12 eV, which lies in the near-infrared. Photons of longer wavelength are transmitted, as they cannot produce electron-hole pairs. Photons of shorter wavelength are not utilized with best efficiency, as any excess

energy above the bandgap energy is lost in heat, in the process called thermalization to the band edges.

Therefore, in order to achieve higher efficiencies several approaches are being considered.

The first approach is called photon management through upconversion. With a suitable layer of nonlinear optical material that is well transparent for shorter wavelength and thus does not degrade the performance of the c-Si cell too much, two photons can be combined to one photon with an energy exceeding the Si bandgap.

The second approach is to combine two semiconductor materials with different bandgap, i.e., tandem solar cells, to utilize the solar spectrum more efficiently. The bottom cell might be silicon, in this case this technology may well be treated as an extension of c-Si PV technology, or it can be another semiconductor with an attractive bandgap for such structures, as will be discussed in the next section.

This discussion shows that also for a crystalline silicon solar technology, which has been the leader in the field for more than 50 years, there is still ample space for further technological improvements resulting in higher efficiency at lower production cost. Further reduction of production costs will be accompanied by a strong push towards high-efficiency cell structures based on the development of new process technologies. This combined effort will help to further reduce the already very low costs of photovoltaic electricity in the coming decades and allow photovoltaics to develop very soon into a significant pillar of our worldwide energy supply.

9.4 High-Concentration PV: CPV Technology

The highest efficiencies of any photovoltaic technology, so far, have been reached with solar cells made of combinations of different III–V compound semiconductor materials. These compound semiconductors offer a wide range of bandgaps and thus are particularly suitable for multi-junction solar cells, in which solar cells with different bandgaps are used to absorb distinct parts of the solar spectrum. This technology has been well described in recent reviews, such as the excellent recent summary by Phillips and Bett [105].

III–V monolithically stacked multi-junction solar cells have reached efficiencies of 38.8% under the global reference spectrum AM1.5 g, 35.8% under AM0 for space applications and 46.0% under the concentrated direct reference spectrum AM1.5d [106–108].

Solar cells with such highest efficiencies are now the standard cells in space applications as well as in terrestrial high-concentration systems. Due to the high cost needed to launch satellites into space the cost benchmark is based on power-per-weight, therefore cells with highest efficiency are preferred, almost independent of production costs. For terrestrial applications, the cost per kWh generated by the solar cell is the benchmark. The levelized cost of energy (LCOE),

defined as the total cost of a system divided by its lifetime energy production, is often used to rate the cost competiveness of different photovoltaic technologies. In general, high efficiencies are essential for lowering the levelized cost of energy (LCOE), as more energy can be produced from the same installation area. With respect to concentrating photovoltaics (CPV), it is necessary that the higher systems costs, due to the usage of a tracking system, do not outweigh the benefit of higher energy output. Moreover, III–V multi-junction solar cells are essentially more expensive than conventional, e.g. silicon-based solar cell technologies of the same cell area. The main cost driver is the substrate used: The price of typical GaAs or Ge wafers can be up to a hundredfold higher than for a Silicon wafer. Thus, even if their efficiency is about twice as high, substrate-based III–V multi-junction solar cells are currently too expensive for the use in flat-plate modules on Earth. Hence it was essential to develop high concentrating CPV systems with concentration factors above 300 in order to introduce III–V multi-junction solar cells to the terrestrial market [109].

Figure 9.14 shows the development of newly installed low- and high-concentration CPV systems per year. The obvious majority are high-concentration systems which use monolithic multi-junction solar cells with highest efficiencies. Today, the on-going development of III–V multi-junction solar cells is driven by both, the needs of space and terrestrial market, and both benefit from each other.

The currently most common technology for monolithic multi-junction solar cells are MOVPE (metal-organic vapour phase epitaxy) grown $Ga_{0.50}In_{0.50}P$/$Ga_{0.99}In_{0.01}As$/Ge triple-junction structures. They only contain lattice-matched layers. The internal architecture of this monolithically grown solar cell is sophisticated, since tunnel diodes, barrier and passivation layers as well as differently doped layers are needed. Figure 9.15 shows such a structure in order to demonstrate the complexity.

Anyhow, this device has achieved a record efficiency of 41.6% (AM1.5d, 364 suns) [110] and is commercially available with efficiencies around 42% under

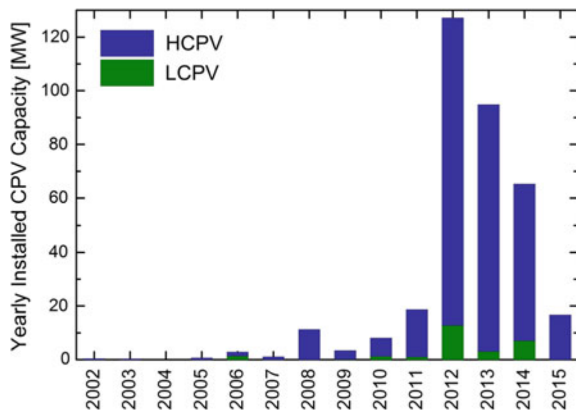


Fig. 9.14 Yearly installed CPV capacity with differentiation between low concentration (LCPV; concentration factor up to 30) and high concentration (HCPV; concentration factor beyond 300) systems

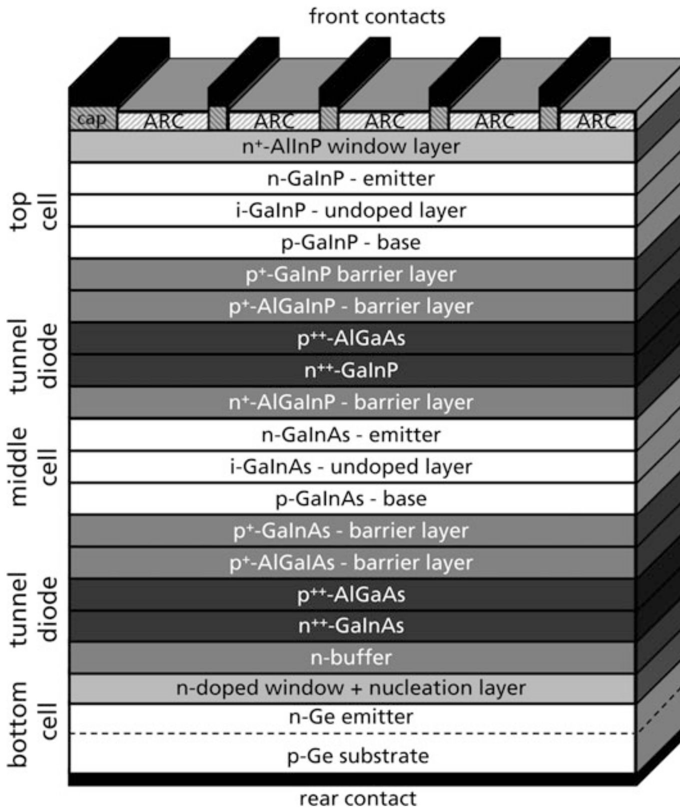


Fig. 9.15 Simplified sketch (not to scale) of a lattice-matched $Ga_{0.50}In_{0.50}P/Ga_{0.99}In_{0.01}As/Ge$ triple-junction solar cell [105]

concentrated sunlight. New concepts are being investigated to further increase the efficiencies. Today, a variety of materials as well as an extensive technological toolbox is available. Consequently, diverse solar cell architectures have been realised, which allow a more efficient use of the solar spectrum. By now, several approaches have already reached efficiencies higher than the lattice-matched triple-junction solar cell. They shall be discussed in some detail.

As to the substrates, several choices are available for the growth of III–V multi-junction solar cells. These differ in technical characteristics like bandgap, lattice constant, off-cuts and available doping of the substrate material, as well as in economic aspects such as price and availability. In many architectures the substrate not only functions as a seed layer for epitaxial growth, but also becomes the bottom cell of the subcell stack, e.g. through diffusion of doping material into the upper part of the substrate. On the other hand, the lattice constant of the substrate material should match that of the upper solar cell layers since this facilitates epitaxial growth

with high material quality. The famous graph of Fig. 9.16 shows various binary and ternary III–V semiconductors as well as common substrates as examples.

Currently, Germanium is the standard substrate for commercial III–V triple-junction solar cells, due to its relatively low bandgap, the lattice-match with suitable III–V semiconductors, and its commercial availability. Alternatives are still being investigated, as Ge is relatively expensive. One low cost alternative is Silicon, which found increasing attention in the last years. Silicon could enable high efficiencies in dual- and triple-junction solar cells. However, for multi-junctions with more than three pn-junctions its bandgap is too high. The main technical challenge lies in the epitaxial growth of suitable III–V semiconductors on Silicon, due to its smaller lattice constant. Various strategies such as metamorphic growth or wafer bonding are being investigated to overcome the lattice-mismatch. For a long time, GaAs has been used as a substrate material for single- and dual-junction solar cells. As a direct semiconductor, high absorption is facilitated and several common III–V compounds with similar lattice-constant are available to build solar cell stacks in good material quality. However, its high bandgap prohibits multi-junction architectures with more than two junctions with a GaAs bottom cell.

Several research groups are also investigating epitaxial growth of III–V multi-junction solar cells on InP substrates, as this allows lattice-matched growth of materials for high efficiency bandgap combinations [112–115]. Although scientific results are promising, high substrate costs currently pose a challenge to wide commercial adoption. Another substrate material is GaSb, which has mostly been investigated with a focus on thermo-photovoltaics (TPV), and mechanically stacked multi-junction devices [116–118]. GaSb has moved out of the research focus in recent years, but might be revisited as an infrared absorber in the future, due to its low bandgap. This substrate is currently even more expensive than InP. However, significant cost reductions are possible as Antimony is in principle available in large quantities.

Depending on the substrate used and the intended materials needed to make a specific solar cell architecture, different epitaxial growth concepts can be applied.

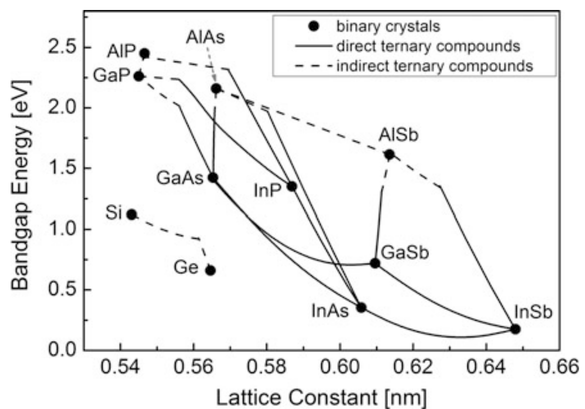


Fig. 9.16 Relation between bandgap energy and lattice constant for exemplary ternary semiconductors. *Lines* between binary crystals represent direct (*solid lines*) or indirect (*broken lines*) ternary compounds (based on Bett et al. [111])

The straight-forward approach is to grow III–V multi-junction solar cells only containing lattice-matched materials with respect to the substrate. However, this way only a restricted range of materials and thus bandgaps can be grown lattice-matched to the available substrates. Using the standard lattice-matched concept on Ge substrates leads to twice the current in the Ge subcell. In other words, the efficiency for this structure is not maximised.

Therefore, the concept of metamorphic growth has been developed. Here, materials with different lattice constant are grown on top of each other. This allows the adaption of the bandgaps of the top and middle cell on a Ge-substrate to maximise the efficiency. However, metamorphic growth is challenging since the change in the lattice constant introduces defects and strain. Therefore, specifically designed buffer structures have to be implemented [119–121]. They allow to gradually transition the lattice constant, and eventually provide a substrate surface for growth with a new lattice constant, see Fig. 9.17.

It is important in such structures that all dislocations and other crystal defects resulting from the difference in the lattice constants should be confined to within the buffer, so that the subsequent solar cell layers can be grown strain-relaxed with good material quality. In general, it is extremely challenging to realise buffer structures absorbing all defects. Some threading dislocations might propagate into the growing layers. Therefore, it is particularly challenging to implement a buffer

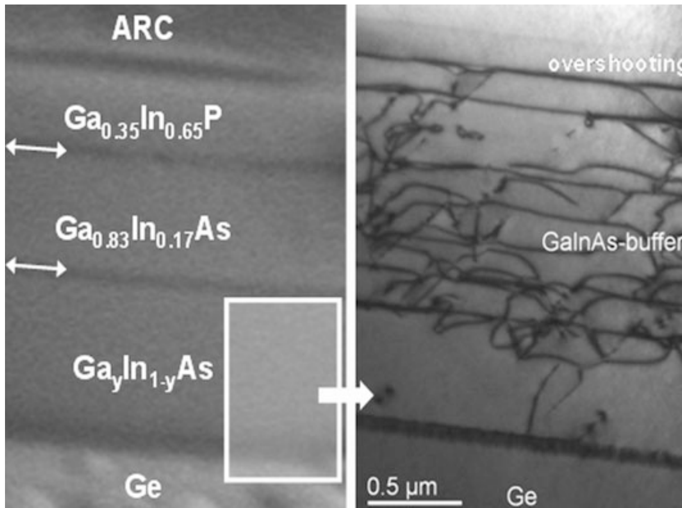


Fig. 9.17 REM image of the epitaxial structure of a metamorphic $Ga_{0.35}In_{0.65}P/Ga_{0.83}In_{0.17}As/Ge$ triple-junction solar cell (*left*). In addition to the sub cells the structure contains tunnel diodes (marked with *arrows*), an anti-reflective coating (ARC) and a $Ga_{1-y}In_yAs$ buffer. A 220 bright field TEM cross section shows that dislocations are confined in the stepgraded buffer (*right*). This allows epitaxial growth of the upper sub-cells in high material quality [120] (Permission Guter et al. Copyright 2009, AIP Publishing LLC)

structure early during the growth phase. In such a situation, inverted metamorphic growth is preferred [122–125].

In this approach, the multi-junction solar cell is grown inversely, i.e. with the top and middle cell being grown first on a lattice-matched substrate followed by another subcell, which is grown upon a metamorphic buffer. After growth lift-off techniques are used to remove the substrate from the solar cell and the grown structure is flipped around. From a technical point of view, the inverted metamorphic growth concept has mainly two advantages compared to upright growth. First, the growth of the buffer is postponed to the latest possible growth phase, while the major part of the multi-junction solar cell can be grown lattice matched to a suitable substrate. Thus, threading dislocations due to the transition of the lattice constant do not affect the upper cells. Second, the bandgap of the bottom cell can be chosen with fewer restrictions, as the cell structure is not grown on the Ge substrate. Economically, a cost benefit in production could arise if the same costly growth substrate is used for several epitaxial runs. Yet this might be counterbalanced by higher production costs and lower yield due to the complexity of cell processing. For space applications, the possible lower power to mass ratio and the possibility for flexible modules could be additional benefits.

The wide choice of III–V materials allows for various designs of III–V multi-junction solar cells, and is the key to high efficiencies. However, the pre-requisite for high efficiencies is excellent material quality. Realizing theoretically optimal multi-junction solar cell stacks with sufficient material quality might be the main challenge in the R&D of III–V solar cells. The central parameter to evaluate material quality is the charge carrier lifetime. It determines in particular the diffusion length of minority carriers. If the lifetime and thus the diffusion length in the emitter or base layer are too short, minority carriers will recombine before reaching the pn-junction and the cell current will remain low. The quality of the thin films is obviously affected by the purity of the epitaxial sources and by residues from previous growth in the epitaxial reactor. But even in a pure environment material quality is strongly influenced by the parameters of the epitaxy, e.g., pressure, temperature, and III/V ratio in MOVPE growth. Moreover, lattice-mismatch and thermal stress within the layer stack need to be dealt with adequately in order to limit the number of defects within the structure.

Particular challenges arise from the high number of (hetero)-interfaces within the layer stack. Defects at the interface can lead to high recombination velocities. Moreover, depending on the band alignment at the interface charge carriers might flow in unintended directions, e.g. away from the pn-junction. In order to reduce these effects emitter and base layer are enclosed by passivation layers. For these, front-surface-field (window) and back-surface-field layers III–V semiconductors are chosen, which shall allow for a sufficient interface quality and for a beneficial band alignment.

One example of a material system, which would be highly valuable for III–V multi-junction solar cells are dilute nitrides (GaIn) (NAs). It offers a wide range of bandgaps from 1 to 2 eV for different lattice-constants. However, in MOVPE-growth material quality has been limited leading to short minority carrier

diffusion lengths [126, 127]. Currently, dilute nitrides can only be grown with MBE in sufficiently high material quality [128].

A more complex class of materials are nanostructures. Research efforts are on-going to increase the efficiency of III–V solar cells through the integration of nanostructures (for an overview see [129, 130]. Specific nanostructures can be integrated into the absorbing layers or placed outside the solar cell. In some approaches the complete device is structured on the nanoscale.

Multiple Quantum Wells (MQW) or Quantum Dots (QD) were proposed to be incorporated into the intrinsic region of a subcell in a multi-junction solar cell. This gives a new technological option to achieve current matching conditions and thus to increase the efficiency of the standard $Ga_{0.50}In_{0.50}P/Ga_{0.99}In_{0.01}As/Ge$ triple-junction solar cell. Multiple Quantum Wells can be realised by thin alternating layers of semiconductors with higher and lower bandgap compared to the host material, leading to a confinement of charge carriers in one dimension and hence to discrete energy values. By placing such a stack into the intrinsic region of the p- and n-doped layers of a solar cell, the absorption can be extended to longer wavelengths. In the standard triple-junction solar cell this leads to higher current densities. An opposing trend is the decrease of open-circuit voltage, which might balance a possible efficiency increase. However, it was found that strained-balanced MQWs minimise the voltage loss [131].

Instead of MQWs, QDs can also be used to achieve the current matching. Also investigated was whether QD can be used to form an intermediate absorption band within the bandgap of a semiconductor. An overview about other applications of quantum dots in solar cells can be found in the book edited by Wu and Wang [132].

Corresponding intermediate band solar cells could increase the current, while preserving the voltage. Although experimental proofs of concept were provided, structures with high efficiencies have not yet been realised. The reason for the failure so far of all intermediate band solar cell approaches might be that creating an intermediate bandgap in any semiconductor to facilitate carrier excitation by photons of smaller energy than the main badgap not only results in carrier excitation, but as well carrier recombination. This has been well studied in low-temperature MBE-grown GaAs, where an intermediate bandgap can be introduced through a high concentration of As antisite defects forming a midgap level. At concentrations above 10^{20} cm^{-3} these degenerate into a midgap band that allows carrier absorption at energy above 0.8 eV. However, these layers show as well the shortest lifetimes observed in GaAs, in the fsec range [133].

Another recent approach is the growth of a forest of nanowires from III–V materials, which avoids dislocations in metamorphic material combinations and reduces overall material consumption [134–136]. Moreover, as growth on Si substrates is possible, the high costs of III–V substrates can be avoided. Although still in the early prototype phase, nanowire solar cells are promising. Recently an efficiency of 13.8% has been achieved with single-junction InP nanowires [134]. An overview about the experimental status and the theoretical background of III–V nanowire solar cells can be found in [137].

After epitaxial growth, the stack of semiconductor layers is processed into solar cell devices. Various post-growth technologies can be used. These include photolithography and etching for defining the structures, deposition of the anti-reflection coating as well as metallization techniques for top and rear contacts. As discussed above, monolithic growth of optimal multi-junction solar cell stacks can be challenging, in particular for devices with more than three junctions. Therefore, advanced post-growth technologies which allow realizing monolithic multi-junction solar cells without the requirement of monolithical growth of the complete structure are interesting.

One post-growth technology of special interest is wafer bonding, combining independently grown (multi-junction) solar cells. It can help to overcome the limitations in monolithic growth arising from different lattice constants or thermal expansion coefficients. One of the two stacks is usually grown inversely. During the bonding process, the cell stacks are in principle just pressed together. However, as the bonding interface lies between subcells within the solar cell stack, it needs to fulfil several requirements. In particular high optical transparency is important in order to avoid parasitic absorption. This is why direct wafer bonds are favourable in contrast to metal-bonding. Moreover, the wafer bond needs to have high mechanical strength and a low electrical resistance.

One option is to bond two wafers at room temperature after deoxidizing the surfaces with wet-chemical etching and use thermal annealing at high temperatures to ensure a good bond. However, the induced thermal stress might result in cracks. Another option is to activate the bond surfaces with highly energetic argon atoms followed by a bond at room temperature (Fig. 9.18). Through rapid thermal annealing as well as high doping levels at the interface, low resistances have already been demonstrated with fast atom beam activated direct wafer bonding [138].

After the bonding step, the substrate of the upper stack needs to be removed, which can for example be achieved with laser, stress induced, ion implantation or wet chemical lift-off processes. In the future it is intended to re-use the substrate

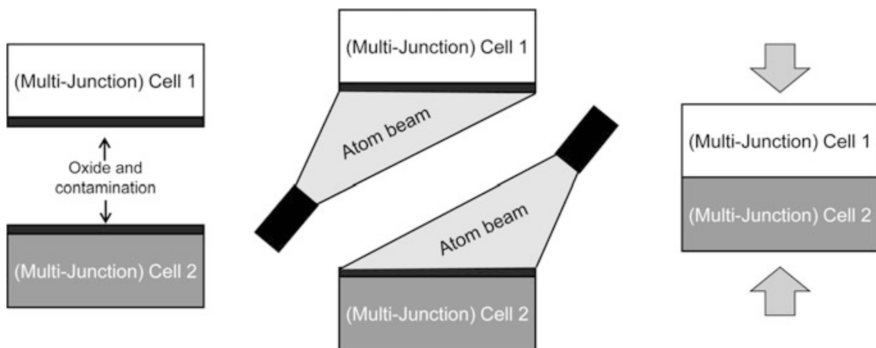


Fig. 9.18 Schematic illustration of fast atom beam activated wafer bonding. Before the actual bonding of the two (multi-junction) solar cells their surfaces are cleaned with atom beams in order to allow for a bond interface with low resistance and high mechanical stability

several times, in order to reduce costs. Although this process is technologically challenging, promising results have already been achieved.

An obvious option for combining independently grown solar cells was mainly used before MOVPE-growth was well established. This option uses simple mechanical stacking. Solar cells are arranged on top of each other without direct contact and without removing the substrate(s). However, due to the precise positioning required, the use of several substrates and the necessity of duplicative processing costs, this does no longer appear to be competitive with monolithically grown structures or even wafer bonded devices. Moreover, optical losses in the substrate and at the interfaces decrease the efficiency of mechanical stacks.

Another approach is based on spectrum-splitting architectures, which split the solar spectrum and direct the light toward single- or dual-junction solar cells with adequate absorption ranges. Thereby the challenge of lattice-match is circumvented. Several devices in prototype stage have already been realised. However, as for mechanical stacks competitive costs will be difficult to achieve due to greater assembly effort, more extensive processing and the use of several substrates.

The wide variety of III–V semiconductors available in conjunction with the extensive technical toolbox has resulted in the heterogeneity of different III–V multi-junction solar cells. Figure 9.19 presents several cell architectures which are under current consideration. For more extensive reviews, see for example references [139–145].

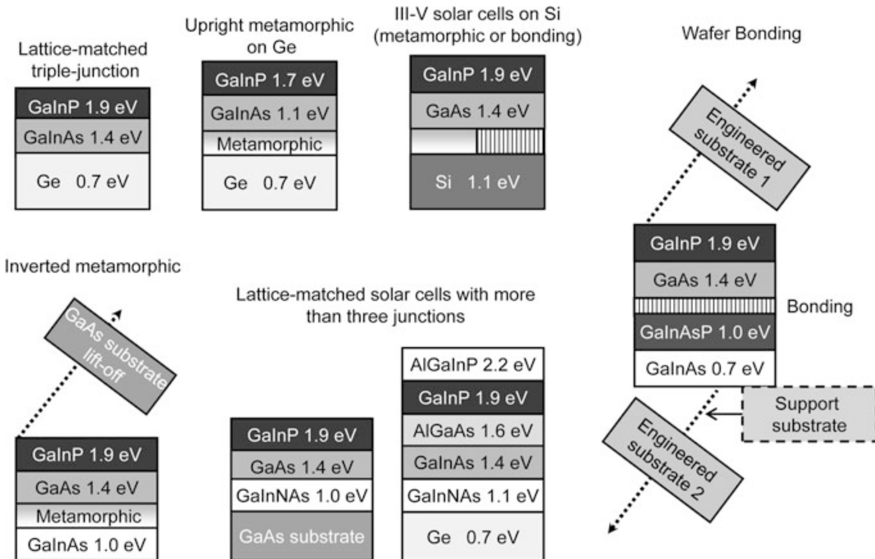


Fig. 9.19 Schematic illustration of some currently investigated III–V multi-junction solar cell concepts. Some concepts have also been realized with different materials and bandgap combinations [105]

One class of approaches employing wafer-bonding shall be briefly discussed. It has been proposed already in 2002 to create multi-junction solar cells with more than three junctions using wafer-bonding [146]. Recently, a five-junction solar cell with 35.8% under AM0 and 38.8% under AM1.5g was presented [106, 107]. Three upper subcells of AlGaInP (2.2 eV), AlGa(In)As (1.7 eV) and Ga(In)As (1.4 eV) were grown inverted on a GaAs or Ge substrate and then combined via wafer-bonding with subcells of GaInPAs (1.1 eV) and GaIn(P)As (0.7 eV). The latter were grown upright on an InP substrate. All subcells are grown lattice-matched to their respective substrate. Another semiconductor-bonded four-junction solar cell with different bandgaps (1.9/1.4/1.0/0.73 eV) achieved an efficiency around 33.5% under AM0 [147]. Recently, a four-junction solar cell consisting of a GaInP/GaAs dual-junction wafer bonded to a GaInAsP/GaInAs dual-junction for terrestrial concentrator applications was developed. The challenge here is to obtain a very low ohmic resistance at the wafer bonding interface. This device achieved a world-record efficiency of 46.0% (AM1.5d, 508 suns) [106, 108].

However, this very exciting multi-junction solar cell technology applied in high concentration systems still has not yet found the attention in the market place it deserves. At present, this technology still is more expensive in the cost per installed Watt as well as the levelized cost of each kWh produced. However, with the growth of this industry, the cost learning curve can be well expected to extend even below the cost learning curve of c-Si PV modules.

9.5 Thin Film PV Technologies

Photovoltaic energy converters today are dominated in by crystalline Silicon technology as shown in Fig. 9.10. As the c-Si solar cells for terrestrial applications have matured, the cost for these solar cells have been dominated by material costs for silicon wafers, glass cover plates and encapsulants.

The costs of solar cells may be reduced by replacing Silicon wafers as absorber materials by suitable other semiconductor materials. Compared to bulk Si Cells with thickness of around 200 µm the total thickness of the semiconductor absorbers in thin film solar cells is less than a few µm. This is possible due to the different absorption coefficients for c-Si and thin film materials. The mainstream materials are amorphous Silicon (a-Si), cadmium telluride (CdTe), and copper indium diselenide (CIS or CIGS), see Fig. 9.20 [148, 149].

Besides possible lower material costs thin film solar cell may be cheaper to manufacture as energy costs, handling costs and capital costs are reduced. This is especially valid for roll to roll printed processes.

Thin film solar cells function the same way as bulk solar cells: p-type and n-type semiconductor layers form a junction, where the carriers generated due to the light absorption are separated. The p- and n-type semiconductor layers are deposited by suitable techniques. In contrast to wafer based technologies thin film solar cells are interconnected directly during the processing of the cells by structuring the layers

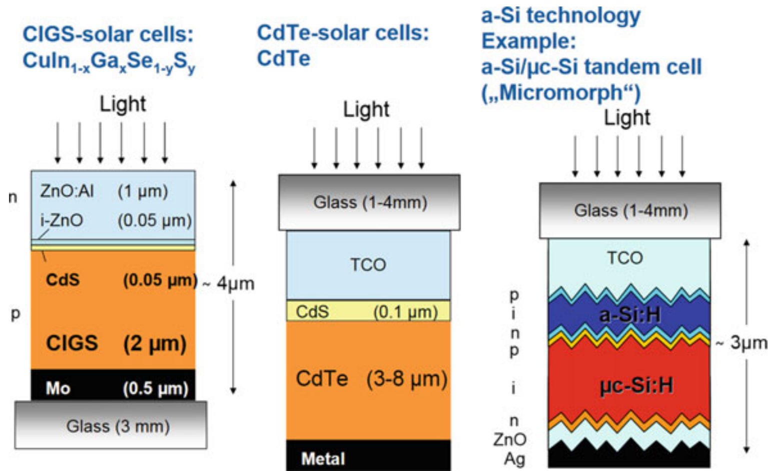


Fig. 9.20 Thin film solar cell structures, TCO is the Transparent Conducting Oxide [148, 149]

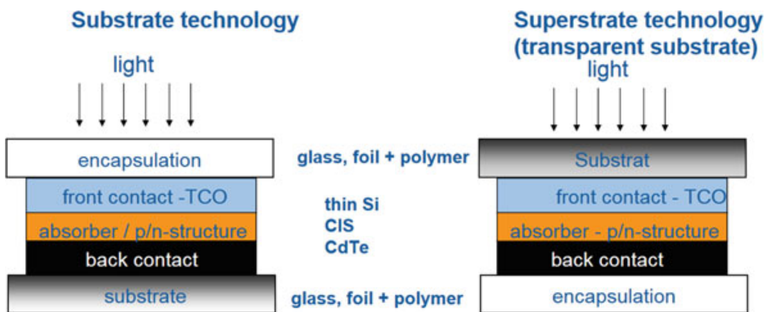


Fig. 9.21 Sub- and superstrate cell configurations

deposited. The solar cells can be formed either on a substrate, with the incident light solar cell impinging the Semiconductor layers first, or in a substrate configuration, where the light need to pass the superstrate before entering the semiconductor layers. This can be seen in Fig. 9.21.

Due to the thickness of the thin films solar cells can be flexible and lower in weight than c-Si cells, thus allowing a wide variety of applications.

In 2009, 14% of the thin film PV production was hydrogenated amorphous silicon (a-Si:H) based, 7% CIGS based, and 79% CdTe based [148].

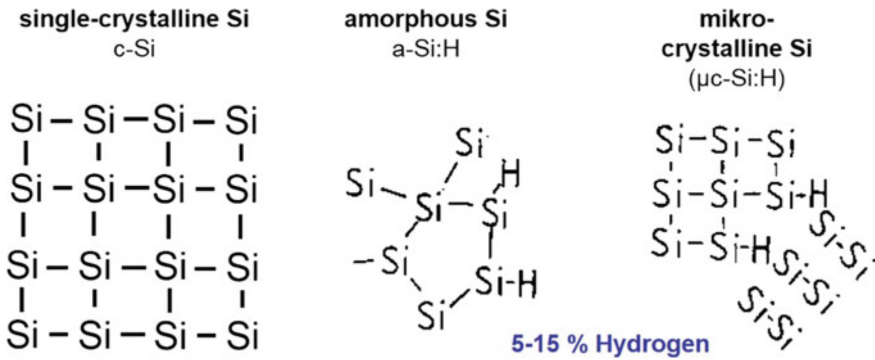


Fig. 9.22 Schematic of crystalline, amorphous and microcrystalline Silicon

9.5.1 Thin Film Silicon Solar Cells

Thin film silicon solar cells in contrast to bulk silicon are highly disordered materials. While in crystalline silicon the atoms are in a regular three dimensional order, amorphous silicon has no clear order. In microcrystalline silicon we find only very small grains. This is depicted in Fig. 9.22.

9.5.1.1 Amorphous Silicon

Most of the commercial thin film silicon solar cells today are -a-Si:H based solar cells. The processing of these solar cell occurs in a low temperature regime below 250 °C. The silicon based layers are deposited mainly by plasma enhanced chemical vapor deposition (PECVD) techniques. Most of the thin film Si solar cells are single junction devices. Figure 9.21 exhibits the basic solar cell structure. Light enters the intrinsic a-Si:H layer, which serves as the absorber layer through the p-layer. This facilitates the collection of holes from the absorber layer, which have a lower mobility than electrons in a-Si:H.

In case of a superstrate configuration, the p-i-n deposition sequence starts on a transparent substrate carrier. Usually a glass plate coated with a transparent conducted oxide is used as a carrier. The back contact is a highly reflecting metal layer, usually deposited onto a TCO interlayer, which is used to improve the reflection from the back contact.

In case of the substrate configuration, the substrate carrier forms the backside of the cell. This allows opaque substrates, e.g., stainless steel or opaque polymer foils. A highly reflecting back contact with textured surface consisting of silver or aluminum and a TCO layer is deposited onto the carrier. After deposition of the n-i-p a-Si:H based layers a TCO top layer with a metal grid is formed as the top contact. The frontside of the substrate cell is finished by a transparent encapsulant layer with an additional glass plate.

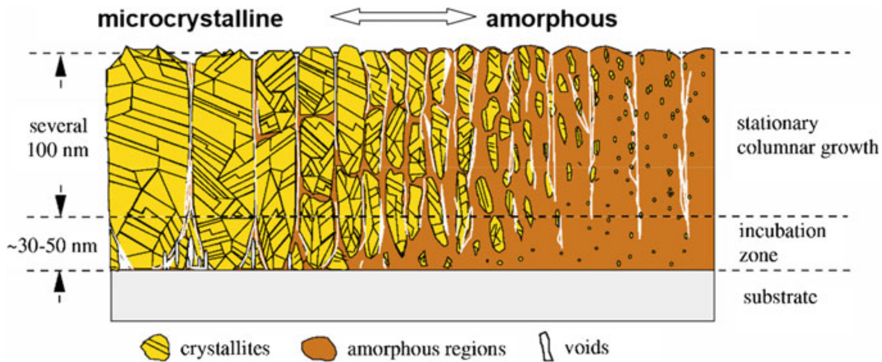


Fig. 9.23 Schematics of transition amorphous—microcrystalline silicon [152, 153]

Inherent to a-Si:H is the creation of metastable defects when material is exposed to light [150]. The performance degradation a a-Si:H based solar cell due to illumination can be partly avoided by using thinner intrinsic layers. However this is associated with lower absorption. The solution for having cells with better stability is the use of a stacked or multijunction solar cell structures or using material, which is more stable against light exposure [151].

9.5.1.2 Microcrystalline Silicon

Microcrystalline silicon is a mixed phase material (μc-Si:H), see Fig. 9.23. It contains -a-Si and micro crystallites of silicon [152, 153]. The performance of solar cells from transition type microcrystalline Silicon is based on the fact that a considerable amorphous silicon content passivates the grain boundary defects, whereas the crystalline fraction is high enough to bring about sufficient absorption in the near infrared part of the spectrum. The composition of the film is process sensitive.

9.5.1.3 Multijunction Solar Cells

For power generation multijunction-solar cells are used, as single junction cells suffer from low conversion efficiencies. In multijunction solar cells two or more solar cells are stacked on top of each other, either mechanically or by monolithic integration [154, 155]. The absorber layer of each cell can be tailored to specific part of the solar spectrum. Top cells absorb the short wavelength part of the spectrum whereas bottom cells absorb the remaining long wavelength part of the spectrum (Fig. 9.24).

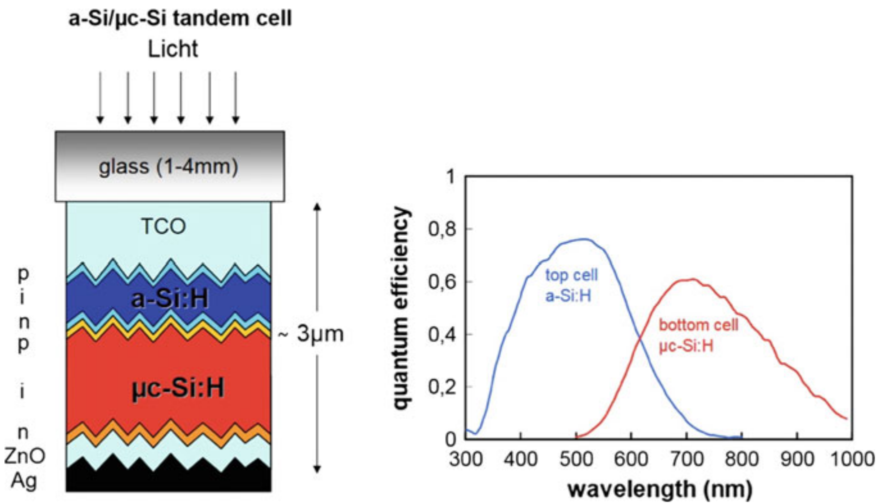


Fig. 9.24 Tandem cell structure and quantum efficiency for the tandem cell

Amorphous silicon PV modules (including tandem and tripel junction modules) have a market share of about 0.6 GWp in 2015, which is 1% of the total global PV production.

9.5.2 Cupper Indium Diselenide (CIS) Solar Cells

Chalcopyrite materials are considered to be the most promising thin film materials [156]. Their band gaps are well matched to the solar spectrum and their favourable electronic properties. Table 9.2 depicts the elements for which possible combinations with photovoltaic effects are possible.

The absorber layer is mostly co-evaporated or co-sputtered copper, gallium, and indium onto a substrate at room temperature, the resulting film is then annealed in a selenide vapor. An alternative process is to co-evaporate copper, gallium, indium and selenium onto a heated substrate [157].

A non-vacuum-based alternative process deposits nanoparticles of the precursor materials on the substrate and then sinters them in situ. Electroplating is another low cost alternative to apply the CIGS layer [158, 159].

Common to all these deposition processes is a subsequent selenization process. The Se supply and selenization environment is important in determining the

Table 9.2 Elements suitable for chalcopyrite-type solar cells

Cu	Al	S
Ag	Ga	Se
Au	In	Te

properties and quality of the film. When Se is supplied in the gas phase (for example as H_2Se or elemental Se) at high temperatures, the Se becomes incorporated into the film by absorption and subsequent diffusion. During this step, called chalcogenization, complex interactions occur to form a chalcogenide. These interactions include formation of Cu-In-Ga intermetallic alloys, formation of intermediate metal-selenide binary compounds and phase separation of various stoichiometric CIGS compounds. Because of the variety and complexity of the reactions, the properties of the CIGS film are difficult to control. Use of H_2Se provides the best compositional uniformity and the largest grain sizes. However, H_2Se is highly toxic and is classified as an environmental hazard.

CIS or CIGS modules have a market share of about 1.1 GWp in 2015, which is 2% of the total global PV production.

9.5.3 Cadmium Telluride Solar Cells

Thin film photovoltaic devices based on CdTe absorbers today are dominating the thin film PV market [160]. The development to this strong position happened within the past few years surpassing thin film silicon technology.

The CdTe technology at present has the lowest specific production costs of any PV technology [161]. The p/n junction is formed with CdS layer as the n-type partner, CdTe as the p type absorber. From theory, the CdS/CdTe junction should be a hetero p-junction. But in most practical cases the electrical junction forms between the CdTe absorber layer and an intermixed CdSTe layer which forms during CdTe deposition. Figure 9.25 gives a schematic diagram of the junction location.

CdTe can be deposited at very high rates (> 20 μm), which is much higher as for other thin film absorber materials. This relates in in-line processing shorter CdTe deposition zones (and lower hardware costs associated).

CdTe modules have a market share of about 2,5 GWp in 2015, which is 6% of the total global PV production.

9.6 The Future of PV: Further Market Development, PV Going into the Terawatt Range

Within the last 30 years, the price reduction of silicon PV modules can be well described by a price experience curve with a learning factor of 20%, see Fig. 9.26 [163]: for each doubling of the globally installed capacity on the average a 20% reduction in price could be observed!

Due to a strong global overproduction capacity since 2010, this price reduction was even stronger than this extrapolation in the last years, resulting in module

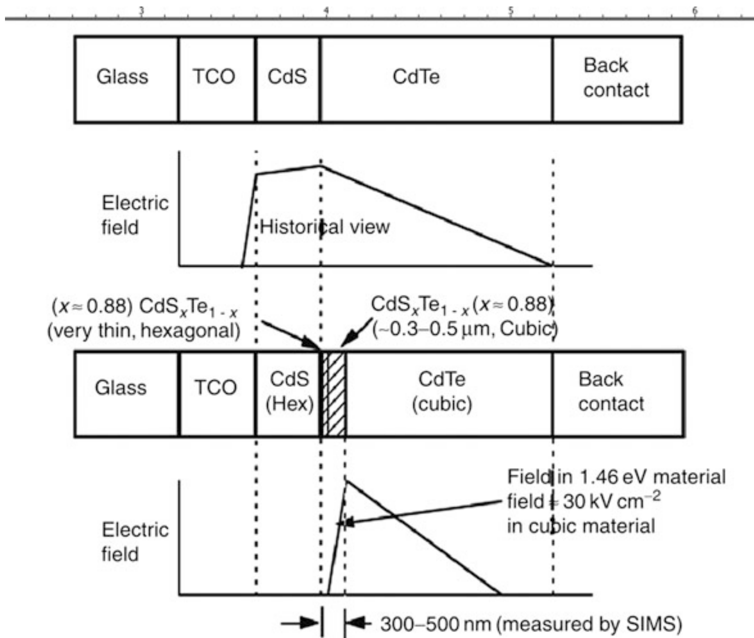


Fig. 9.25 Diagramm illustrating the location of the junction in a CdS/CdTe device [162]

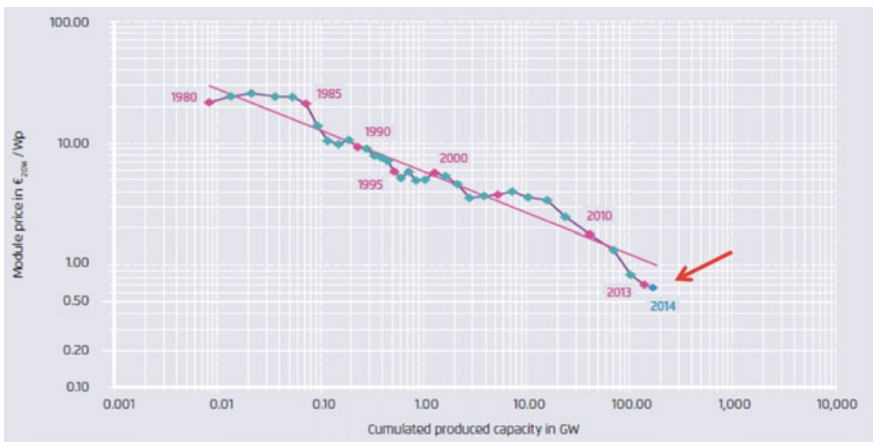


Fig. 9.26 Price experience curve of crystalline Si PV modules [163]

prices below the extrapolated price-experience curve, and well below $1/W$_p$. This is an excellent situation for customers and PV installers, but rather challenging for producers of silicon solar cells and modules, as well as equipment manufactures.

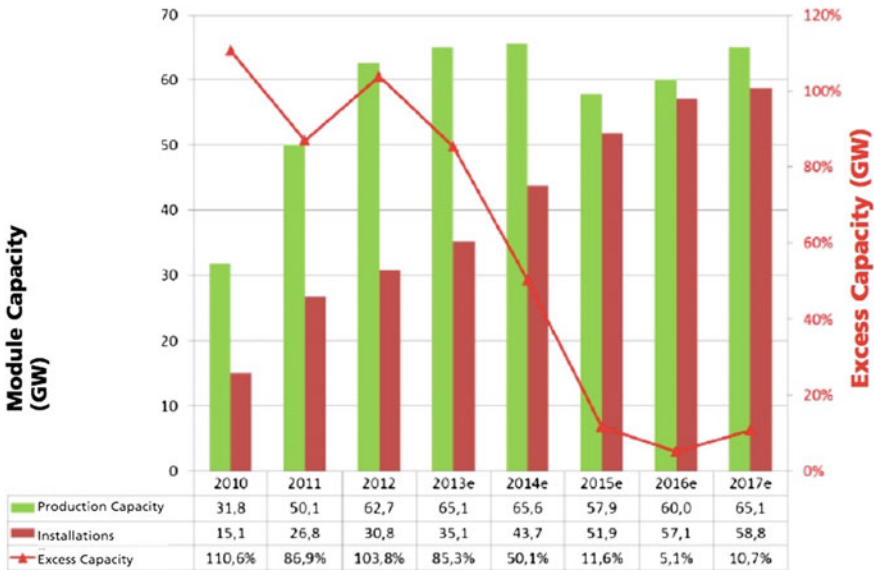


Fig. 9.27 Global annual PV production capacity and PV installations [164]

However, the time of strong global over-production capacity is presently ending, see Fig. 9.27: In 2016 it is expected that the global market size of ca. 60 GW_p is basically matched by the global production capacity in the same range, around 70 GW.

Further strong growth of the global PV market is expected, see Fig. 9.28, so that in the next years we can expect a doubling of the global PV production capacity, from the current level of about 60 GW_p/year to 100–120 GW_p/year in 2020.

In March of 2016, a high-level workshop convened in Freiburg, Germany, organized by the Global Alliance of Solar Research Institutes GA-SERI. The topic was PV heading into the Terawatt level. As a result of this workshop a detailed report discussing the expected further growth of the global PV-market was published [166] (Fig. 9.29).

It was concluded that starting from a 57 GW_p-market reported by Bloemberg in 2015 different market growth scenarios resulted in the following total deployment numbers for 2030: no-growth, i.e., constant 57 GW_p/year above 1 TW_p, with a 15% growth rate it would reach 3 TW_p deployment by 2030, and with a 25% groth rate it could result in 8 TW_p of PV deployment!

A scenario published in 2014 by the International Energy Agency IEA predicts only about 1.5 TW_p of PV deployment in 2030, see Fig. 9.30, and close to 5 TW_p in 2050. Nevertheless, even the IEA forecast depicts drastically the dramatic market increase we are facing in the coming decade and beyond. From the scenarios discussed in the March workshop, we can predict much more than 5 TW_p of PV

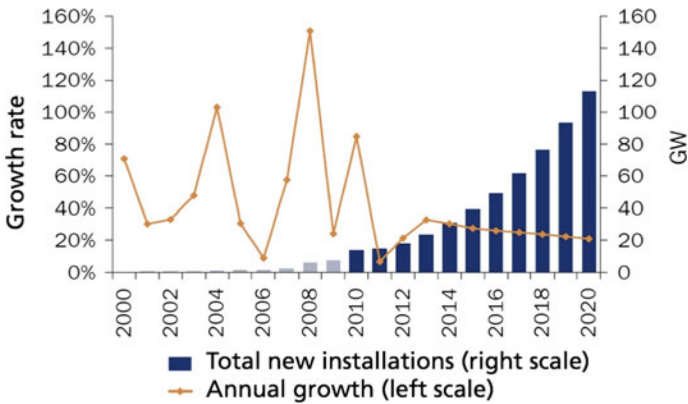


Fig. 9.28 Expected annual PV installations [165]. It is noteworthy that since 2010, the real annual installations (2016: ca. 60 GW) well exceeded all Sarazin predictions!

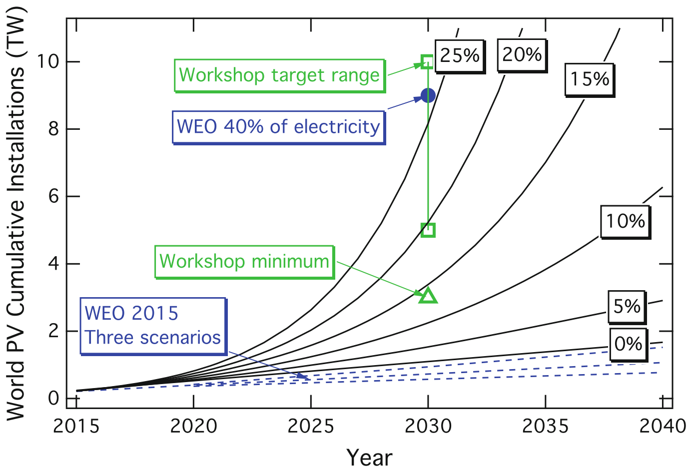


Fig. 9.29 Expected cumulative PV installations, as proposed by the March workshop on PV heading into the Terawatt range [166]

deployment by 2050, which would correspond only to less than 10% of the world's electricity supply in 2050.

These scenarios are of course driven by the continuing reduction of the costs of PV-generated electricity. In 2016, PV electricity in Germany has been auctioned for about 7 €ct/kWh, reported from auctions of large-scale PV electricity in Germany, and in Abu Dhabi, a recent auction resulted in a serious offer for only 3 $ct/kWh! Given the different insolation conditions at the two sites, these two numbers actually are quite similar!

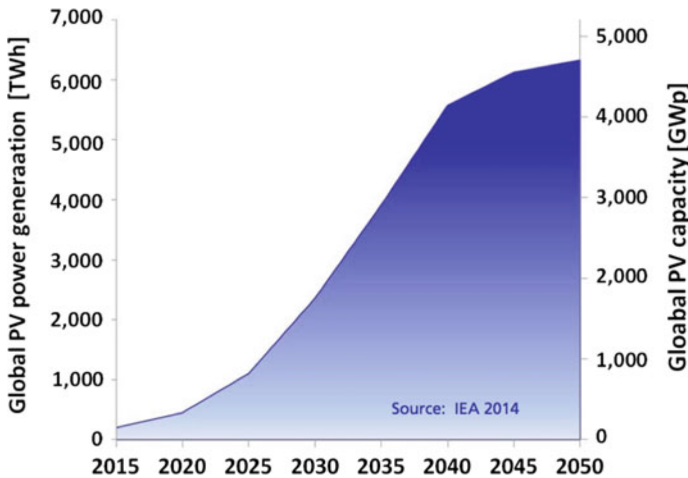


Fig. 9.30 Expected cumulative PV installations in one of the scenarios discussed by the IEA in 2014. Currently, just 300 GW_p are installed, showing that we are just at the beginning of an impressive growth curve of the global PV industry in the coming 5-10 years!

It can be well expected that further growth of the PV industry will bring prices of PV installations further down. This will be aided by the transition from GW-type of factories to production volumes of several GW_p/year for an individual company, the adoption of the advanced cell concepts discussed in this chapter, in conjunction with better process control and cost savings on all levels of the production chain, from the Si raw material to cell and module technology. The combined effect of these developments will bring down the cost of PV electricity to 2 ct/kWh in many locations of the world, probably well before 2050.

In addition, and not discussed in this chapter, energy storage is as well developing very rapidly. The costs especially for batteries are decreasing similar dramatically as the cost for PV modules, driven today especially by the rapidly growing needs of batteries for electro-mobility. Thus, decentralized PV systems with batteries for the nighttime needs become rapidly a cost-effective option.

Based on these expectations, even the prospect of a world-connecting electricity supergrid based on high-voltage DC power lines becomes a realistic vision. This would finally solve the problem of the required 24-h availability of electricity, as the sun shines on about half of our planet at any given time.

In summary, in the coming years the outlook for the PV industry is very promising. The low costs of PV generated electricity will result in a further doubling of annual PV installations by 2020, combined with the need for doubling the global PV cell and module production capacity. Beyond 2020, the annual PV market might climb from about 100 GW_p/year to 200–300 GW_p/year 2025, heading for deployments of several TW_p of PV in 2030 and beyond.

This deployment of greenhouse gas-neutral PV electricity will be accompanied by a similar development of wind-based electricity generators, and it will be driven

mainly by economic and technologic factors. If we can avoid that stakeholders of the old, fossil- and nuclear based energy system of the Second Industrial Revolution hamper this desirable development through political action and negative campaigns, that we already experienced in Germany, we might indeed contribute a key factor to avoid the gravest dangers of anthropogenic changes of the earth's atmosphere, so that the goals of the international climate accord achieved at COP-21 in Paris might come into reach!

References

1. A. Gore, *Earth in the Balance: Ecology and the Human Spirit* (Houghton Mifflin, Boston, 1992)
2. J. Rifkin, *The Third Industrial Revolution* (Palgrave Macmillan, New York, 2011)
3. Versorgungsqualität – SAIDI-Werte 2006–2014. Bundesnetzagentur (Bonn, 2015), http://www.bundesnetzagentur.de/DE/Sachgebiete/ElektrizitaetundGas/Unternehmen_Institutionen/Versorgungssicherheit/Stromnetze/Versorgungsqualitaet/Versorgungsqualitaet-node.html. Accessed 7 Oct 2016
4. H.J. Fell 2016, unpublished
5. Die Energiewende in Deutschland. Mit sicherer, bezahlbarer und umweltschonender Energie ins Jahr 2050. Bundesministerium für Wirtschaft und Technologie (Berlin, 2012), https://www.bmwi.de/Dateien/BMWi/__Archiv/energiewende-in-deutschland,property=pdf,bereich=bmwi2012,sprache=de,rwb=true.pdf. Accessed 7 Oct 2016
6. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A Roadmap for Moving to a Competitive Low Carbon Economy in 2050. European Commission (Brussels 2011), http://eur-lex.europa.eu/legal-content/SK/TXT/?uri=CELEX:52011DC0112. Accessed 7 Oct. 2016
7. The 2020 Climate and Energy Package. European Commission. (Brussels, 2011), http://ec.europa.eu/clima/policies/strategies/2020/index_en.htm. Accessed 7 Oct 2016
8. Emissionen von Luftschadstoffen. Umweltbundesamt (Dessau-Roßlau, 2015), http://www.umweltbundesamt.de/themen/luft/emissionen-von-luftschadstoffen. Accessed 7 Oct. 2016
9. A. Palzer, H.M. Henning, Ener. Techn. **2**, 13 (2009)
10. H.M. Henning, A. Palzer: What will the energy transformation cost? in: *Pathways for Transforming the German Energy System by 2050. Fraunhofer ISE* (Freiburg, 2015), https://www.ise.fraunhofer.de/en/publications/studies/what-will-the-energy-transformation-cost. Accessed 18 Oct 2016
11. A. Palzer: *Sektorübergreifende Modellierung und Optimierung eines zukünftigen deutschen Energiesystems unter Berücksichtigung von Energieeffizienzmaßnahmen im Gebäudesektor*, Dissertation, KIT, Karlsruhe, (2016)
12. T. Trost, S. Horn, M. Jentsch, M. Sterner, Z. Energiewirtsch. **36**, 173 (2012)
13. Zahlen und Fakten. Energiedaten. Nationale und internationale Entwicklung. Bundesministerium für Wirtschaft und Technologie (Berlin, 2015), http://bmwi.de/DE/Themen/Energie/Energiedaten-und-analysen/Energiedaten/gesamtausgabe,did=476134.html. Accessed 7 Oct 2016
14. CDC (Climate Data Center). Deutscher Wetterdienst (Offenbach, 2014), http://www.dwd.de/cdc. Accessed 7 Oct 2016
15. Veröffentlichungen zum COSMO-Modell (COSMO-EU, COSMO-DE, COSMO-DE EPS). Deutscher Wetterdienst (Offenbach, 2014). www.dwd.de. Accessed 7 Oct 2016

16. Current and future cost of photovoltaics, in *Long-term Scenarios for Market Development, System Prices and LCOE of Utility-Scale PV Systems*. Agora Energiewende, Berlin (2015), https://www.agora-energiewende.de/fileadmin/Projekte/2014/Kosten-Photovoltaik-2050/AgoraEnergiewende_Current_and_Future_Cost_of_PV_Feb2015_web.pdf. Accessed 7 Oct 2016
17. PSE AG Photovoltaic Report 2016, Data: from 2000 to 2010: Navigant; from 2011: IHS (Mono-/Multi- proportion from Cell Production)
18. S.W. Glunz, Advanced Concepts in Photovoltaics, eds. by A.J. Nozik, G. Conibeer, M.C. Beard. *Royal Society of Chemistry*, (Cambridge, 2014), p. 1
19. E.L. Ralph, *Proceedings of 11th IEEE PVSC,* (Scottsdale, Arizona, USA, 1975), p. 315
20. A. Aberle, T. Lauinge, R. Hezel, *Proceedings of 14th EU PVSEC,* (Barcelona, Spain 1997), p. 684
21. G. Hahn, *Proceedings of 25th EU PVSEC,* (Valencia, Spain, 2010), 1091
22. A.W. Blakers, A. Wang, A.M. Milner, J. Zhao, M.A. Green, Appl. Phys. Lett. **55**, 1363 (1989)
23. J. Zhao, A. Wang, M.A. Green, *Proceedings of 21st IEEE PVSC,* (Kissimmee, Florida, USA, 1990), p. 333
24. J. Knobloch, A. Aberle, W. Warta, B. Voss, *Proceedings of 5th International PVSEC* (Kyoto, Japan, 1990), p. 955
25. J. Zhao, A. Wang, M.A. Green, Prog. Photovolt. **7**, 471 (1999)
26. A. Mohr, P. Engelhart, C. Klenke, S. Wanka, A.A. Stekolnikov, M. Scherff, R. Seguin, S. Tardon, T. Rudolph, M. Hofmann, F. Stenzel, J.Y. Lee, S. Diez, J. Wendt, W. Brendle, S. Schmidt, J.W. Müller, P. Wawer, M. Hofmann, P. Saint-Cast, J. Nekarda, D. Erath, J. Rentsch, R. Preu, *Proceedings of 26th EU PVSEC* (Hamburg, Germany, 2011), p. 2150
27. A. Mohr, S. Wanka, A. Stekolnikov, M. Scherff, R. Seguin, P. Engelhart, C. Klenke, J.Y. Lee, S. Tardon, S. Diez, J. Wendt, B. Hintze, R. Hoyer, S. Schmidt, J.W. Müller, P. Wawer, Energy Procedia **8**, 390 (2011)
28. P. Engelhart, D. Manger, B. Klöter, S. Hermann, A.A. Stekolnikov, S. Peters, H.-C. Ploigt, A. Eifler, C. Klenke, A. Mohr, G. Zimmermann, B. Barkenfelt, K. Suva, J. Wendt, T. Kaden, S. Rupp, D. Rychtarik, M. Fischer, J.W. Müller, *Proceeding of 26th EU PVSEC* (Hamburg, Germany, 2011), p. 821
29. B. Sun, J. Sheng, S. Yuan, C. Zhang, Z. Feng, Q. Huang, *Proceedings of. 38th IEEE PVSC* (Austin, Texas, USA, 2012), p. 1125
30. S. Zhang, W.W. Deng, X.J. Pan, H.J. Jiao, D.M. Chen, H.W. Huang, Y.F. Cui, J.M. Xu, J. Feng, M. Zhong, Y.F. Chen, P.P. Altermatt, Z.Q. Feng and P.J. Verlinden, *Proc. 42nd IEEE PVSC* (New Orleans, Louisiana, USA, 2015), p. 1
31. E. van Kerschaver, S. de Wolf, J. Szlufcik, *Proceedings of 28th IEEE PVSC* (Anchorage, Alaska, USA, 2000), p. 209
32. R.A. Sinton, Y. Kwark, J.Y. Gan, R.M. Swanson, IEEE Electron Dev. Lett. **7**, 567 (1986)
33. B. Hoex, J. Schmidt, R. Bock, P.P. Altermatt, M.C.M. van de Sanden, W.M.M. Kessels, Appl. Phys. Lett. **91**, 112107/1 (2007)
34. J. Schmidt, A. Merkle, R. Brendel, B. Hoex, M.C.M. van de Sanden, W.M.M. Kessels, Prog. Photovolt. **16**, 461 (2008)
35. A. Metz, D. Adler, S. Bagus, H. Blanke, M. Bothar, E. Brouwer, S. Dauwe, K. Dressler, R. Droessler, T. Droste, M. Fiedler, Y. Gassenbauer, T. Grahl, N. Hermert, W. Kuzminski, A. Lachowicz, T. Lauinger, N. Lenck, M. Manole, M. Martini, R. Messmer, C. Meyer, J. Moschner, K. Ramspeck, P. Roth, R. Schönfelder, B. Schum, J. Sticksel, K. Vaas, M. Volk, K. Wangemann, Sol. Energ. Mat. Sol. Cells **120**, 417 (2013)
36. P. Saint-Cast, J. Benick, D. Kania, L. Weiss, M. Hofmann, J. Rentsch, R. Preu, S.W. Glunz, IEEE Electron Dev. Lett. **31**, 695 (2010)
37. A. Richter, F.M. Souren, D. Schuldis, R.M. Görtzen, J. Benick, M. Hermle, S.W. Glunz, *Proceedings of 27th EU PVSEC* (Frankfurt, Germany, 2012), p. 1133
38. P. Saint-Cast, D. Kania, M. Hofmann, J. Benick, J. Rentsch, R. Preu, A. Phys, Lett. **95**, 151502 (2009)

39. T.-T. Li, A. Cuevas, Phys. Status Solid. RRL **3**, 160 (2009)
40. K.A. Münzer, J. Schöne, A. Teppe, M. Hein, R.E. Schlosser, M. Hanke, J. Maier, A. Yodyungyong, J. Isenberg, T. Friess, C. Ehling, K. Varner, S. Keller, P. Fath, *Proceedings of 26th EU PVSEC* (Hamburg, Germany, 2011), p. 843
41. S.W. Glunz, R. Preu, S. Schaefer, E. Schneiderlöchner, W. Pfleging, R. Lüdemann, G. Willeke, *Proceedings of 28th IEEE PVSC* (Anchorage, Alaska, USA, 2000), p. 168
42. M. Bähr, G. Heinrich, O. Doll, I. Köhler, C. Maier and A. Lawerenz, *Proceedings of 26th EU PVSEC* (Hamburg, Germany, 2011), p. 1206
43. E. Schneiderlöchner, R. Preu, R. Lüdemann, S.W. Glunz, Prog. Photovolt. **10**, 29 (2002)
44. S.W. Glunz, E. Schneiderlöchner, D. Kray, A. Grohe, H. Kampwerth, R. Preu, G. Willeke, *Proceedings of 19th EU PVSEC* (Paris, France, 2004), p. 408
45. D. Kray, S. Glunz, Prog. Photovolt. **14**, 195 (2006)
46. J.-F. Nekarda, F. Lottspeich, A. Wolf, R. Preu, *Proceedings of 25th EU PVSEC* (Valencia, Spain, 2010), p. 2211
47. H. Fischer, W. Pschunder, *Proceedings of 10th IEEE PVSC* (Palo Alto, California, USA, 1973), p. 404
48. S.W. Glunz, S. Rein, W. Warta, J. Knobloch, W. Wettling, *Proceedings of 2nd World Conference on Photovoltaic Energy Conversion* (Vienna, Austria, 1998), p. 1343
49. S.W. Glunz, S. Rein, J.Y. Lee, W. Warta, J. Appl. Phys. **90**, 2397 (2001)
50. J. Schmidt, A.G. Aberle, R. Hezel, *Proceedings of 26th IEEE PVSC* (Anaheim, California, USA, 1997), p. 13
51. J. Schmidt, K. Bothe, Phys. Rev. B **69**, 0241071 (2004)
52. S.W. Glunz, E. Schäffer, S. Rein, K. Bothe, J. Schmidt, *Proceedings of 3rd World Conference on Photovoltaic Energy Conversion* (Osaka, Japan, 2003), p. 919
53. S.W. Glunz, J. Benick, D. Biro, M. Bivour, M. Hermle, D. Pysch, M. Rauer, C. Reichel, A. Richter, M. Rüdiger, C. Schmiga, D. Suwito, A. Wolf, R. Preu, *Proceedings of 35th IEEE PVSC Honolulu* (Hawaii, USA, 2010)
54. A. Herguth, G. Schubert, M. Kaes and G. Hahn, *Proceedings of 4th World Conference on Photovoltaic Energy Conversion, Waikoloa, Hawaii, USA*, 2006, 940
55. K.A. Münzer, K.T. Holdermann, R.E. Schlosser, S. Sterk, *Proceedings of 2nd World Conference on Photovoltaic Energy Conversion,* (Vienna, Austria, 1998), p. 1214
56. S. Rein, W. Warta, S.W. Glunz, *Proceedings of 28th IEEE PVSC* (Anchorage, Alaska, USA, 2000), p. 57
57. B. Lim, V.V. Voronkov, R. Falster, K. Bothe, J. Schmidt, Appl. Phys. Lett. **98**, 1 (2011)
58. J. Vedde, T. Clausen, L. Jensen, *Proceedings of 3rd World Conference on Photovoltaic Energy Conversion* (Osaka, Japan, 2003), p. 943
59. A. Cuevas, M.J. Kerr, C. Samundsett, F. Ferrazza, G. Coletti, Appl. Phys. Lett. **81**, 4952 (2002)
60. D. Macdonald, L.J. Geerligs, Appl. Phys. Lett. **85**, 4061 (2004)
61. B. Michl, J. Benick, A. Richter, M. Bivour, J. Yong, R. Steeman, M.C. Schubert, S.W. Glunz, Energy Procedia **33**, 41 (2012)
62. S.W. Glunz, S. Rein, J. Knobloch, W. Wettling, T. Abe, Prog. Photovolt. **7**, 463 (1999)
63. T. Yoshida and Y. Kitagawara, *Proceedings of 4th International Symposium on High Purity Silicon IV* (San Antonio, Texas, USA, 1996), p. 450
64. J. Mandelkorn, C. McAfee, J. Kesperis, L. Schwart, W. Pharo, J. Electrochem. Soc. **109**, 313 (1962)
65. R.R. King, R.M. Swanson, IEEE T. Electron Dev. **38**, 1399 (1991)
66. J. Benick, B. Hoex, O. Schultz, S.W. Glunz, *Proceedings of 33rd IEEE PVSC* (San Diego, USA, 2008)
67. D.L. Meier, H.P. Davis, R.A. Garcia, J. Salami, A. Rohatgi, A. Ebong, P. Doshi, Sol. Energ. Mat. Sol. Cells **65**, 621 (2001)
68. C. Schmiga, M. Hörteis, M. Rauer, K. Meyer, J. Lossen, H.-J. Krokoszinski, M. Hermle, S. W. Glunz, *Proceedings of 24th EU PVSEC* (Hamburg, Germany, 2009), p. 1167

69. C. Schmiga, M. Rauer, M. Rüdiger, K. Meyer, J. Lossen, H.-J. Krokoszinski, M. Hermle, S. W. Glunz, *Proceedings of 25th EU PVSEC* (Valencia, Spain, 2010), p. 1163
70. R. Bock, S. Mau, J. Schmidt, R. Brendel, Appl. Phys. Lett. **96**, 1 (2010)
71. M. Tanaka, M. Taguchi, T. Takahama, T. Sawada, S. Kuroda, T. Matsuyama, S. Tsuda, A. Takeoka, S. Nakano, H. Hanafusa, Y. Kuwano, Prog. Photovolt. **1**, 85 (1993)
72. P.P. Altermatt, H. Plagwitz, R. Bock, J. Schmidt, R. Brendel, M.J. Kerr, A. Cuevas, *Proceedings of 21st EU PVSEC* (Dresden, Germany, 2006), p. 647
73. M.J. Kerr, *Surface, emitter and bulk recombination in silicon and development of silicon nitride passivated solar cells*, PhD Dissertation (Australian National University, Canberra, 2002)
74. J. Libal, R. Petres, T. Buck, R. Kopecek, G. Hahn, R. Ferre, M. Vetter, I. Martín, K. Wambach, I. Roever, P. Fath, *Proceedings of 20th EU PVSEC* (Barcelona, Spain, 2005), p. 793
75. J. Benick, B. Hoex, M.C.M. van de Sanden, W.M.M. Kessels, O. Schultz, S.W. Glunz, Appl. Phys. Lett. **92**, 253504/1 (2008)
76. D. Suwito, U. Jäger, J. Benick, S. Janz, M. Hermle, S.W. Glunz, IEEE T. Electron Dev. **57**, 2032 (2010)
77. J. Benick, B. Hoex, M.C.M. van de Sanden, W.M.M. Kessels, O. Schultz, S.W. Glunz, Appl. Phys. Lett. **92**, 253504 (2008)
78. M.P. Godlewski, C.R. Baraona, H.W. Brandhorst, Jr., *Proceedings of 10th IEEE PVSC* (Palo Alto, California, USA, 1973), p. 40
79. E. Yablonovitch, T. Gmitter, R.M. Swanson, Y.H. Kwark, Appl. Phys. Lett. **47**, 1211 (1985)
80. W. Fuhs, K. Niemann, J. Stuke, Bull. Am. Phys. Soc. **19**, 394 (1974)
81. K. Okuda, H. Okamoto, Y. Hamakawa, Jap. J. Appl. Phys. **22**, L605 (1983)
82. M. Taguchi, M. Tanaka, T. Matsuyama, T. Matsuoka, S. Tsuda, S. Nakano, Y. Kishi, Y. Kuwano, *Proceedings of 5th International PVSEC* (Kyoto, Japan, 1990), p. 689
83. M. Taguchi, E. Maruyama, M. Tanaka, Jap. J. Appl. Phys. **47**, 814 (2008)
84. M. Taguchi, Y. Tsunomura, H. Inoue, S. Taira, T. Nakashima, T. Baba, H. Sakata, E. Maruyama, *Proceedings of 24th EU PVSEC* (Hamburg, Germany, 2009), p. 1690
85. K. Masuko, M. Shigematsu, T. Hashiguchi, D. Fujishima, M. Kai, N. Yoshimura, T. Yamaguchi, Y. Ichihashi, T. Mishima, N. Matsubara, T. Yamanishi, T. Takahama, M. Taguchi, E. Maruyama, S. Okamoto, IEEE J. Photovolt. **4**, 1433 (2014)
86. S. de Wolf, A. Descoeudres, Z.C. Holman, C. Ballif, Green **2**, 7 (2012)
87. J.L. Hernandez, K. Yoshikawa, A. Feltrin, N. Menou, N. Valckx, E. van Assche, D. Schroos, K. Vandersmissen, H. Philipsen, J. Poortmans, D. Adachi, M. Yoshimi, T. Uto, H. Uzu, T. Kuchiyama, C. Allebe, N. Nakanishi, T. Terashita, T. Fujimoto, G. Koizumi, K. Yamamoto, Jap. J. Appl. Phys. **51**(10N), A04 (2012)
88. D.L. Bätzner, Y. Andrault, L. Andreetta, A. Büchel, W. Frammelsberger, C. Guerin, N. Holm, D. Lachenal, J. Meixenberger, P. Papet, B. Rau, B. Strahm, G. Wahli, F. Wünsch, Energy Procedia **8**, 153 (2011)
89. A. Descoeudres, Z.C. Holman, L. Barraud, S. Morel, S.de Wolf, C. Ballif, IEEE J. Photovolt. **3**, 83 (2013)
90. D. Muñoz, T. Desrues, A.S. Ozanne, N. Nguyen, S. de Vecchi, F. Souche, S. Martin de Nicholas, C. Denis, P.J. Ribeyron, *Proceedings of 26th EU PVSEC* (Hamburg, Germany, 2011), p. 861
91. J.-H. Choi, S.-K. Kim, J.-C. Lee, H. Park, W.-J. Lee, E.-C. Cho, *Proceedings of 26th EU PVSEC* (Hamburg, Germany, 2011), p. 3302
92. L. Korte, E. Conrad, H. Angermann, R. Stangl, M. Schmidt, Sol. Energ. Mat. Sol. Cells **93**, 905 (2009)
93. M. Bivour, C. Reichel, M. Hermle, S.W. Glunz, Sol. Energ. Mat. Sol. Cells **106**, 11 (2012)
94. Z.C. Holman, A. Descoeudres, L. Barraud, F.Z. Fernandez, J.P. Seif, S. de Wolf, C. Ballif, IEEE J. Photovolt. **2**, 7 (2012)
95. P.J. Verlinden, R.M. Swanson, R.A. Crane, Prog. Photovolt. **2**, 143 (1994)

96. K.R. McIntosh, M.J. Cudzinovic, D.D. Smith, W.P. Mulligan, R.M. Swanson, *Proceedings of 3rd World Conference Photovoltaic Energy Conversion* (Osaka, Japan, 2003), p. 971
97. F. Feldmann, M. Bivour, C. Reichel, M. Hermle, S.W. Glunz, Sol. Energ. Mat. Sol. Cells **120**, 270 (2014)
98. H. Steinkemper, F. Feldmann, M. Bivour, M. Hermle, IEEE J. Photovolt. **5**, 1348 (2015)
99. A. Moldovan, F. Feldmann, M. Zimmer, J. Rentsch, J. Benick, M. Hermle, Sol. Energ. Mat. Sol. Cells **142**, 123 (2015)
100. M. Taguchi, K. Kawamoto, S. Tsuge, T. Baba, H. Sakata, M. Morizane, K. Uchihashi, N. Nakamura, S. Kiyama, O. Oota, Prog. Photovolt. **8**, 503 (2000)
101. I.R.C. Post, P. Ashburn, G.R. Wolstenholme, IEEE T. Electron Dev. **39**, 1717 (1992)
102. S.W. Glunz, F. Feldmann, A. Richter, M. Bivour, C. Reichel, H. Steinkemper, J. Benick, M. Hermle, *Proceedings of 31st EU PVSEC* (Hamburg, Germany, 2015), p. 259
103. W. Shockley, H.J. Queisser, J. Appl. Phys. **32**, 510 (1961)
104. A. Richter, M. Hermle, S.W. Glunz, IEEE J. Photovolt. **3**, 1184 (2013)
105. S.A. Philips, A. Bett in *Advanced Concepts in Photovoltaics*, A.J. Nozik, G. Conibeer, M.C. Beard (eds.), (RSC, Cambridge, 2014), p. 87
106. M.A. Green, K. Emery, Y. Hishikawa, W. Warta, E.D. Dunlop, Prog. Photovolt. **24**, 905 (2016)
107. P.T. Chiu, D.C. Law, R.L. Woo, S.B. Singer, D. Bhusari, W.D. Hong, A. Zakaria, J. Boisvert, S. Mesropian, R.R. King, N.H. Karam, *Proceedings of 40th IEEE PVSC* (Denver, USA, 2014), p. 0011
108. F. Dimroth, T.N.D. Tibbits, M. Niemeyer, F. Predan, P. Beutel, C. Karcher, E. Oliva, G. Siefer, D. Lackner, P. Fuß-Kailuweit, A.W. Bett, R. Krause, C. Drazek, E. Guiot, J. Wasselin, A. Tauzin, T. Signamarcheix, IEEE J. Photovolt. **6**, 343 (2016)
109. M. Wiesenfarth, H. Helmers, S.P. Philipps, M. Steiner A.W. Bett, *Proceedings of 27th EU PVSEC* (Frankfurt, Germany, 2012), p. 11
110. R. King, A. Boca, W. Hong, D. Larrabee, K.M. Edmondson, D.C. Law, C. Fetzer, S. Mesropian, N.H. Karam, *Proceedings of 24th EU PVSEC* (Hamburg, Germany, 2009), p. 55
111. A.W. Bett, F. Dimroth, G. Stollwerck, O.V. Sulima, Appl. Phys. **69**, 119 (1999)
112. M.W. Wanlass, T.J. Coutts, J.S. Ward, K.A. Emery, T.A. Gessert, C.R. Osterwald, *Proceedings of 22nd IEEE PVSC* (Las Vegas, Nevada, USA, 1991), p. 38
113. M.P. Lumb, M.K. Yakes, M. Gonzalez, R. Hoheisel, C.G. Baile, W. Yoon, R.J. Walters, *Proceedings of 37th IEEE PVSC* (Seattle, Washington, USA, 2011), p. 00094
114. C.T. Lin, W.E. McMahon, J.S. Ward, J.F. Geisz, M.W. Wanlass, J.J. Carapella, W. Olavarria, M. Young, M.A. Steiner, R.M. Frances, A.E. Kibbler, A. Duda, J.M. Olson, E.E. Perl, D.J. Friedman, J.E. Bowers, *Proceedings of 38th IEEE PVSC* (Austin, Texas, USA, 2012), p. 000944
115. M.S. Leite, R.L. Woo, J.N. Munday, W.D. Hong, S. Mesropian, D.C. Law, H.A. Atwater, Appl. Phys. Lett. **102**, 033901 (2013)
116. A.W. Bett, O.V. Sulima, Semicond. Sci. Technol. **18**, 184 (2003)
117. L.M. Fraas, G.R. Girard, J.E. Avery, B.A. Arau, V.S. Sundaram, A.G. Thompson, J.M. Gee, J. Appl. Phys. **66**, 3866 (1989)
118. A.W. Bett, S. Keser, G. Stollwerck, O.V. Sulima, W. Wettling, *Proceedings of 26th IEEE PVSC* (Anaheim, California, USA, 1997), p. 931
119. A.W. Bett, C. Baur, F. Dimroth, J. Schöne, Mater. Res. Soc. Symp. Proc. **836**, 223 (2004)
120. J. Schöne, E. Spiecker, F. Dimroth, A.W. Bett, W. Jäger, Appl. Phys. Lett. **92**, 081905 (2008)
121. W. Guter, J. Schöne, S.P. Philipps, M. Steiner, G. Siefer, A. Wekkeli, E. Welser, E. Oliva, A.W. Bett, F. Dimroth, Appl. Phys. Lett. **94**, 223504 (2009)
122. D. Aiken, E. Dons, S.-S. Je, N. Miller, F. Newman, P. Patel, J. Spann, IEEE J. Photovolt. **3**, 542 (2013)
123. A.B. Cornfeld, M. Stan, T. Varghese, J. Diaz, A.V. Ley, B. Cho, A. Korostyshevsky, D. J. Aiken, P.R. Sharps, *Proceedings of 33rd IEEE PVSC* (San Diego, California, USA, 2008, p. 1

124. J.F. Geisz, D.J. Friedman, J.S. Ward, A. Duda, W.J. Olavarria, T.E. Moriarty, J.T. Kiehl, M. J. Romero, A.G. Norman, K.M. Jones, Appl. Phys. Lett. **93**, 123505 (2008)
125. R.M. France, J.F. Geisz, M.A. Steiner, D.J. Friedman, J.S. Ward, J.M. Olson, W. Olavarria, M. Young, A. Duda, IEEE J. Photovolt. **3**, 893 (2013)
126. M. Kondow, K. Uomi, A. Niwa, T. Kitatani, S. Watahiki, Y. Yazawa, Jpn. J. Appl. Phys. **35**, 1273 (1996)
127. J.F. Geisz, D.J. Friedman, Semicond. Sci. Technol. **17**, 769 (2002)
128. T.J. Garrod, J. Kirch, P. Dudley, S. Kim, L.J. Mawst, T.F. Kuech, J. Cryst. Growth **315**, 68 (2011)
129. K.W.J. Barnham, I.M. Ballard, B.C. Bowne, D.B. Bushnell, J.P. Connolly, N. J. Ekins-Daukes, M.Fuhrer, R. Ginige, G. Hill, A. Ioannides, D.C. Johnson, M.C. Lynch, M. Mazzer, J.S. Roberts, C. Rohr, T.N.D. Tibbits in *Nanotechnology for Photovoltaics*, ed. by L. Tsakalakos, CRC Press Taylor & Francis Group, (Boca Raton, 2010), p. 187
130. S.P. Philipps, W. Guter, E.Welser, J. Schöne, M. Steiner, F. Dimroth, A.W. Bett Next Generation of Photovoltaics, in *New Concepts,* ed. by A.B. Cristóbal López, A. Martí Vega, A. Luque López, (Springer, Berlin, Heidelberg, 2012), p. 1
131. J.G.J. Adams, B.C. Browne, I.M. Ballard, J.P. Connolly, N.L.A. Chan, A. Ioannides, W. Elder, P.N. Stavrinou, K.W.J. Barnham, N.J. Ekins-Daukes, Prog. Photovolt. **19**, 865 (2011)
132. J. Wu, Z.M. Wang (eds.), *Quantum Dot Solar Cells* (Springer, New York, 2013)
133. K.P. Korona, M. Kaminska, J.M. Baranowski, E.R. Weber, Mater. Sci. Eng. **B 22**, 41 (1993)
134. J. Wallentin, N. Anttu, D. Asoli, M. Huffman, I. Åberg, M.H. Magnusson, G. Siefer, P. Fuss-Kailuweit, F. Dimroth, B. Witzigmann, H.Q. Xu, L. Samuelson, K. Deppert, M.T. Borgström, Science **2013**, 339 (1057)
135. H. Goto, K. Nosaki, K. Tomioka, S. Hara, K. Hiruma, J. Motohisa, T. Fukui, Appl. Phys. Express **2**, 035004 (2009)
136. P. Kailuweit, M. Peters, J. Leene, K. Mergenthaler, F. Dimroth, A.W. Bett, Prog. Photovolt. **20**, 945 (2011)
137. R.R. LaPierre, A.C.E. Chia, S.J. Gibson, C.M. Haapamaki, J. Boulanger, R. Yee, P. Kuyanov, J. Zhang, N. Tajik, N. Jewell, K.M.A. Rahman, Phys. Status Solidi RRL **7**, 815 (2013)
138. S. Essig, F. Dimroth, J. Solid State Sci. Technol. **2**, Q178 (2013)
139. D.C. Law, R.R. King, H. Yoon, M.J. Archer, A. Boca, C.M. Fetzer, S. Mesropian, T. Isshiki, M. Haddad, K.M. Edmondson, D. Bhusari, J. Yen, R.A. Sherif, H.A. Atwater, N.H. Karam, Sol. Energy Mater. Sol. Cells **94**, 1314 (2008)
140. D.J. Friedman, Curr. Opin. Solid State Mater. Sci. **14**, 131 (2010)
141. A. Luque, J. Appl. Phys. **110**, 031301 (2011)
142. S.P. Philipps, W. Guter, E. Welser, J. Schöne, M. Steiner, F. Dimroth, A.W. Bett Next Generation of Photovoltaics, in *New Concepts,* eds. by A.B. Cristóbal López, A. Martí Vega, A. Luque López, (Springer, Berlin, Heidelberg, 2012), p. 1
143. S.P. Philipps, F. Dimroth, A.W. Bett Practical Handbook of Photovoltaics, in *Fundamentals and Applications (Second Edition)*, eds. by A. McEvoy, T. Markvart, L. Castaner, (Academic Press, Boston, 2012), p. 417
144. R.R. King, D. Bhusari, D. Larrabee, X.Q. Liu, E. Rehder, K. Edmondson, H. Cotal, R.K. Jones, J.H. Ermer, C.M. Fetzer, D.C. Law, N.H. Karam, Prog. Photovoltaics **20**, 801 (2012)
145. A.W. Bett, S.P. Philipps, S. Essig, S. Heckelmann, R. Kellenbenz, V. Klinger, M. Niemeyer, D. Lackner, F. Dimroth, *Proceedings of 28th EU PVSEC* (Paris, France, 2013), p. 1
146. J. M. Zahler, A. Fontcuberta i Morral, C.G. Ahn, H.A. Atwater, M. W. Wanlass, C. Chu and P.A. Iles, *Proceedings of 29th IEEE PVSC* (New Orleans, Louisiana, USA, 2002), p. 1039
147. . D. Bhusari, D. Law, R. Woo, J. Boisvert, S. Mesropian, D. Larrabee, H. F. Hong, N.H. Karam, *Proceedings of 37th IEEE PVSC* (Seattle, Washington, USA, 2011), p. 001937
148. M. Zeeman and R.E.I. Schropp in Comprehensive Renewable Energy, in *Photovoltaic Solar Energy,* vol. 1, ed. by A. Sayigh, (Elsevier, Amsterdam, 2012), p. 389

149. B. Rech, Solarzellen aus dünnen Siliziumschichten – Stand der Technik und Herausforderungen für die Zukunft. Helmholtz Zentrum (Berlin, 2008), https://www.wias-berlin.de/events/insk/herbst08/rech.pdf. Accessed 18 Oct 2016
150. D.L. Staebler, C.R. Wronski, Appl. Phys. Lett. **31**, 292 (1977)
151. J. Meier, R. Flückiger, H. Keppner, A. Shah, Appl. Phys. Lett. **65**, 860 (1994)
152. L. Houben: *Plasmaabscheidung von mikrokristallinem Silizium. Merkmale und Mikrostruktur und deren Deutung im Sinne von Wachstumsvorgängen*, Dissertation, FZJ, Universität Düsseldorf, (1999)
153. O. Vetterl, F. Finger, R. Carius, P. Hapke, L. Houben, O. Kluth, A. Lambertz, A. Mück, B. Rech, H. Wagner, Sol. Energ. Mat. Sol. Cells **62**, 97 (2000)
154. D. Fischer, S. Dubail, J.A.A. Selvan, N.P. Vaucher, R. Platz, Ch. Hof, U. Kroll, J. Meier, P. Torres, H. Keppner, N. Wyrsch, M. Goetz, A. Shah, K-D. Ufert, *Proceedings of 25th IEEE PVSC* (Washington DC, USA, 1996), p. 1053
155. S. Guha, J. Yang, *Proceedings of 15th International PVSEC* (Shanghai, China, 2005), p. 35
156. T. Unold, C.A. Kaufmann Comprehensive Renewable Energy in *Photovoltaic Solar Energy*, vol. 1, ed. by A. Sayigh, (Elsevier, Amsterdam, 2012), p. 399
157. M. Powalla, B. Dimmler, Thin Solid Films **361–362**, 540 (2000)
158. Solar Facts and Advice. Alchemie Limited Inc., Delaware (2013), http://www.solar-facts-and-advice.com/CIGS-solar-cell.html. Accessed 18 Oct 2016
159. B. Basol, Commercialization of High Efficiency Low Cost CIGS Technology Based on Electroplating (NREL, Colorado, 2010), www1.eere.energy.gov/solar/pdfs/48590.pdf. Accessed 18 Oct 2016
160. T.A. Gessert, Comprehensive Renewable Energy, in *Photovoltaic Solar Energy,* vol. 1, ed. by A. Sayigh, (Elsevier, Amsterdam, 2012), p. 423
161. C. Martin: Cheaper Than China Again, First Solar Vindicates U.S. Investment. Bloomberg L.P. (New York, 2016), http://www.bloomberg.com/news/articles/2016–04-14/first-solar-making-panels-more-cheaply-than-china-s-top-supplier. Accessed 18 Oct 2016
162. R.G. Dhere, Y. Zhang, M.J. Romero, S.E. Asher, M. Young, B. To, R. Noufi, T.A. Gessert, *Proceedings of 33rd IEEE PVSC* (San Diego, California, USA, 2008), p. 1
163. Navigant Consulting; EUPD PV Module Prices (Since 2006), Graph: ISE (2014)
164. Lux Research Inc., Grafik: PSE AG (2014)
165. Sarazin Bank, (2010)
166. N.M. Haegel, R. Margolis, T. Buonassisi, D. Feldman, A. Froitzheim, R. Garabedian, M. Green, S. Glunz, H.M. Henning, B. Holder, I. Kaizuka, B. Kroposki, K. Matsubara,S. Niki, K. Sakurai, R.A. Schindler, W. Tumas, E.R. Weber, G. Wilson, M. Woodhouse and S. Kurtz, Science **356**, 141–143 (2017)

Chapter 10
Novel Thin-Film Photovoltaics—Status and Perspectives

Benjamin Oesen, Sascha Ullbrich, Johannes Widmer and Karl Leo

Abstract In this chapter, we discuss recent advances in novel thin-film photovoltaic devices which allows novel and low-cost applications of photovoltaics. In particular, we discuss organic and perovskite photovoltaics. At the end, we compare the outdoor harvesting efficiency of these novel systems.

10.1 Introduction: Status of Photovoltaics in General

Due to the limited fossil resources and the climate change problem, it is obvious that humankind has to switch to renewable energies in the mid-term and long term future. From the various sources of alternative energies available, solar energy has by far the largest potential: Already one hour of sunlight on the whole globe provides as much energy as humankind approximately needs per year. Therefore, technologies to harvest sunlight have been intensively researched in the recent decades.

The most promising and elegant technology to harvest sunlight is photovoltaics: Here, sunlight impinging on a semiconductor diode directly generates electrical energy. The fact that this energy is created without any moving parts and without any pollution or noise allows to use photovoltaics technology very efficiently. Therefore, photovoltaics has developed after the first realization of efficient solar cells at Bell laboratories in the 1950s [1] into one of the most important renewable energies.

B. Oesen · S. Ullbrich · J. Widmer · K. Leo (✉)
Dresden Integrated Center for Applied Physics and Photonic Materials (IAPP) and Institute for Applied Physics, TU Dresden, 01062 Dresden, Germany
e-mail: karl.leo@iapp.de

B. Oesen
e-mail: Benjamin.oesen@iapp.de

S. Ullbrich
e-mail: Sascha.ullbrich@iapp.de

J. Widmer
e-mail: johannes.widmer@iapp.de

© Springer International Publishing AG 2017

G. Eisenstein and D. Bimberg (eds.), *Green Photonics and Electronics*,
NanoScience and Technology, DOI 10.1007/978-3-319-67002-7_10

For a long time, the cost of energy generation using photovoltaics was so high that only special applications like energy generation for satellites could be successfully addressed by photovoltaics. In the past few years, however, research, technology, and production have made so much progress that the cost of for photovoltaics has dropped into a region where for many applications, photovoltaics is competitive with conventional ways to generate energy.

The main driver in this development was the further reduction of cost for the main stream technology which is based on crystalline silicon: Within a little more than a decade, the price per Watt peak of silicon dropped from around 10 US$ to well below one US$. With such low cost of silicon solar cells, the energy generation cost in sunny regions of the world are already today below 10 US cents per kWh.

Stimulated by this cost reduction, installations of photovoltaics have grown by orders of magnitude. Today, about 200 GW of photovoltaic modules have been deployed worldwide. It is expected that this number will further increase by orders of magnitude and that in the long term (around 2050), photovoltaics will already produce a major share of the world's electrical energy.

Despite the fact that photovoltaic solar cells based on crystalline silicon solar cells have many advantages like comparatively high efficiency (16–22% in the module), high stability (>20 years), and low cost, there are many reasons to search for further photovoltaic technologies: first, there are applications where crystalline silicon simply cannot be used because the modules need to be flexible, transparent or have very low weight. For instance, many applications require conformable modules, for instance on surfaces of vehicles or buildings. Second, photovoltaics integrated in buildings often ideally require modules with a controllable transparency which allows to use the technology simultaneously as sunshade. Finally, for many applications very low weight is required because e.g. roofs cannot carry high loads.

Despite decades of research on alternative photovoltaic systems, so far no breakthrough has been achieved for alternatives to silicon photovoltaics. Although the market share of crystalline silicon was somewhat fluctuating over time, it still covers the large majority of the current PV market (see Fig. 10.1). The only thin film technologies which have reached a significant market share are CdTe and CIGS (Copper-Indium-Gallium-Diselenide). Other thin film systems, such as amorphous silicon, have never achieved a real breakthrough.

The main challenge for any new thin film technology is the efficiency. As argued above, silicon solar cells have reached quite high efficiencies: The thermodynamic limit for the efficiency of silicon solar cells is around 30%, and the best laboratory cells have reached more than 25%, meaning that the efficiency potential of silicon has been very well exploited. Furthermore, the gap between the best laboratory results and the best module results is continuously diminishing. One key reason for this development is the fact that the silicon technology is in many ways similar to the processing technology used in microelectronics, and the huge experience base of silicon microelectronics could be partly applied to silicon photovoltaics.

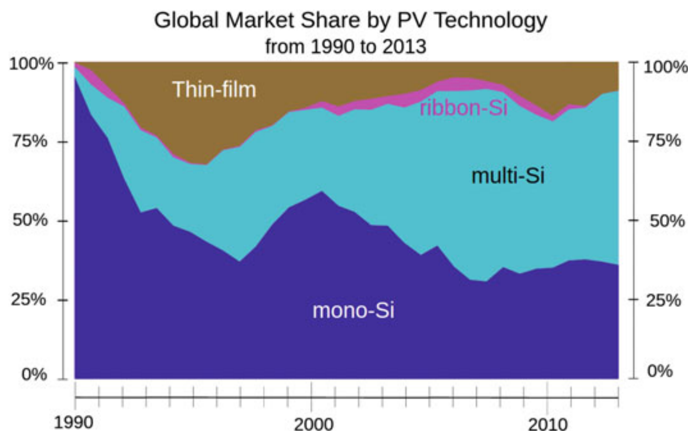


Fig. 10.1 Market share of different photovoltaic technologies versus time

Nevertheless, the learning curve of silicon is already well exploited and the progress in cost reduction is ultimately limited.

The main reason why thin film technologies have not yet been able to compete with silicon is that the efficiencies were always significantly lagging behind the efficiency of silicon. While at first glance, the efficiency might not be that important because sunlight has no cost and one might use large areas, in practice for several reasons high efficiency is crucial: First, the cell or module costs are only a part of the total cost and the system costs will grow out of bounds if the efficiency is too low. Second, in most applications the area is limited and expensive space restricts the use of low cost technologies.

The best progress so far has been achieved by the inorganic compounds CdTe and CIGS: In the past few years, for both technologies the efficiency could be significantly raised and laboratory records of more than 20% have been reached [2]. However, the manufacturing technologies of inorganic thin film photovoltaics have turned out to be more difficult than expected. In contrast to silicon, the stoichiometry is critical and needs to be carefully controlled. For that reason, the homogeneity of the cells is often a serious problem. Nevertheless, CdTe could achieve a larger market share and the market share of CIGS has been recently growing. Organic photovoltaics have been intensively investigated because they are based on carbon, which is an abundant element in nature. Furthermore, organic solar cells are processed at low temperature which will allow easy deposition even by technologies such as printing. This will allow to deposit organic solar cells in roll-to-roll coaters, which might reduce the cost to very low levels. Recent cost predictions have calculated that organic solar cell modules with prices as low as about 8 US$ per m^2 should be possible [3].

In this contribution, we discuss novel organic photovoltaic systems which have many advantages, such as flexibility, abundant materials, transparency, etc., allowing new applications at possibly much lower cost compared to conventional

inorganic semiconductor photovoltaics. Organic photovoltaics can be deposited on plastic substrates and can therefore be flexible. Combined with the very light weight and the fact that they can be made transparent, organic photovoltaics offer many new application scenarios.

10.2 Organic Photovoltaics

10.2.1 Basics of Organic Photovoltaics

In general, semiconductors are materials, which have electrical properties somewhere between metals and insulators. The term organic stems from organic chemistry which generally deals with carbon based molecules. Due to a high variety of potential bond configurations, it is possible to synthesize customized materials for many applications. In the field of photovoltaics, this allows to tailor the energy levels and the optical properties either to maximize the absorption and efficiency or to adjust the visual appearance of the final product.

For the fabrication of organic solar cells, basically two technologies are used: vacuum deposition by thermal evaporation and deposition from solution. In vacuum processing, the organic molecules are heated in material sources in a vacuum chamber and consequently deposited onto the substrate, which is situated opposite to the source (see Fig. 10.2 left). This technique requires rather lightweight molecules or a rigidly bound molecular structure to avoid a decomposition of the

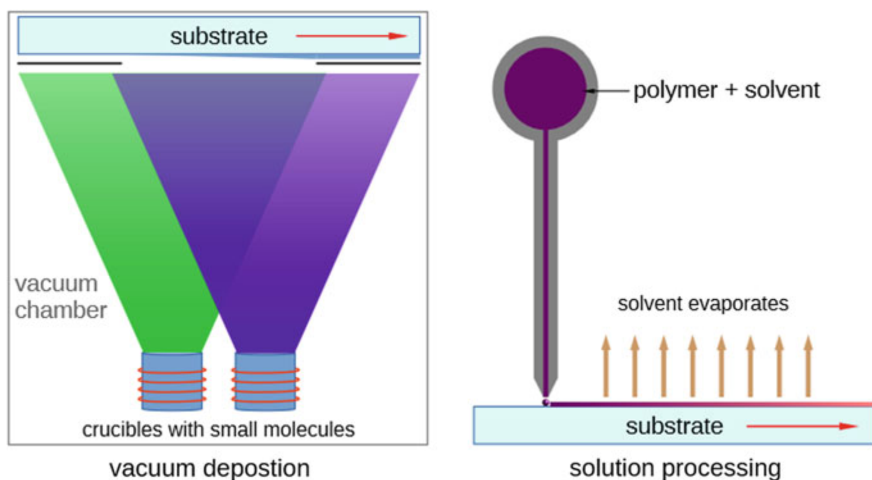


Fig. 10.2 Schematic illustration of the two dominant technologies for the deposition of organic solar cells: vacuum deposition by thermal evaporation (*left*) and solution processing by printing or comparable techniques (*right*)

molecules during the processing. By using oscillating quartz monitors, it is possible to control the film thickness with an accuracy in the range of 1 nm. Co-evaporation, i.e. evaporating two materials at the same time in a specific ratio, easily allows to process blend layers and doped transport layers. To control the morphology, it is possible to heat or cool the substrate during the deposition to achieve smooth layers avoid short cuts in the device, which will be shown on an example in the next section.

Solution processing, on the other hand, does not require a vacuum and is closely related to a classical printing technology. The organic materials are dissolved in a suitable solvent and brought to the substrate. While the solvent evaporates during the process or a subsequent heating process, the organic material remains on the substrate (see Fig. 10.2 right). In contrast to vacuum deposition, there is no limitation for the molecular size, allowing to use polymers as well. A good solubility is a major requirement for solution processing. The morphology of the layers can be influenced by varying the drying speed, using different solvents, or introducing additives into the solution.

While inorganic materials can form well-ordered crystals, organic materials are mostly amorphous or nanocrystalline. As a consequence of this disordered arrangement and the comparably weak van-der-Waals binding between neighboring molecules, the energy landscape is not band-like and delocalized, but charge carriers are rather localized on a single molecule. However, delocalization of energetic states exists on a smaller scale, the molecule orbitals. Therefore, in organic semiconductors, the analog to conduction and valance band are the lowest unoccupied molecular orbital (LUMO) and the highest occupied molecular orbital (HOMO), respectively. These molecular orbitals define the energy gap of an organic material. Hence, an incoming photon can be absorbed if the photon energy is larger than or equal to the difference of the LUMO and HOMO energies. The absorption spectra of a few exemplary organic small molecules used for OPV together with the AM1.5G spectrum[1] are shown in Fig. 10.3.

Furthermore, organic dyes can have very high absorption coefficients [4, 5]. Consequently, even very thin layers in the range of a few tens of nanometers can absorb enough light to reach a competitive power conversion efficiency (PCE).

Unlike inorganic semiconductors, organic materials show relatively low dielectric constants. Hence, a photoexcited electron will be promoted from the HOMO to the LUMO or a higher energy level, but it will remain Coulombically bound to its counterpart, the hole. This electron-hole pair is referred to as exciton and can be treated as a quasi-particle. An exciton is mobile within the material but it has a limited lifetime before it recombines, resulting in an average diffusion length, which an exciton can typically travel before recombining. The binding energy is in the order of 0.1–0.5 eV and the thermal energy at room temperature ($kT = 25$ meV) is insufficient to efficiently split the exciton into free charge carriers.

[1]Solar spectrum for standardized testing and characterization of solar cells.

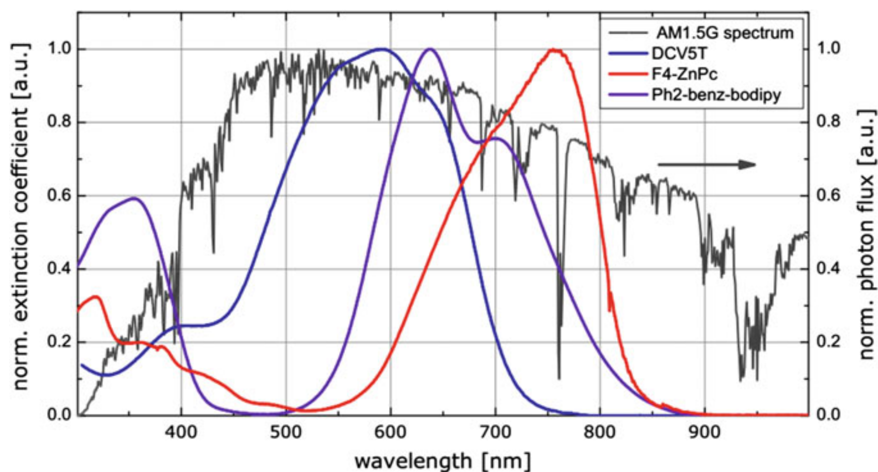


Fig. 10.3 The graph shows the normalized extinction coefficient of a few exemplary donor materials used in organic solar cells (DCV5T(2,2'-((3'',4''-dimethyl-[2,2':5',2'':5'',2''':5''',2''''-quinquethiophene]-5,5''''-diyl)bis(methanylylidene))dimalononitrile [23, 47]), F4-ZnPc(Tetrafluoro-zinc-phthalocyanine), and Ph2-Benz-bodipy(12,12-difluoro-10,14-diphenyl-12H-11λ^4,12λ^4- [4, 5, 41, 47] triazaborinino[4,3-a:6,1-a']diisoindole [4, 5])). Additionally, the dark grey line represents the AM1.5G sun spectrum

In order to drive a current and generate power, the exciton needs to be dissociated. By using a suitable combination of two materials in the active layer it is possible to split the exciton at the interface. Here, a parallel offset between the two HOMO and LUMO energies is required to overcome the exciton binding energy [6]. The material with the lower lying energy levels is referred to as acceptor, the other is denoted as donor. In case an exciton reaches the donor-acceptor interface within its lifetime, it is energetically preferable for one charge carrier to hop onto the adjacent molecule, while the other is blocked due to an energy barrier, resulting in the electron situated on the acceptor and the hole on the donor. Figure 10.4 shows exemplarily the dissociation of an exction that was generated on the acceptor material.

There are several ways to build heterojunction devices: One concept is the planar heterojunction (PHJ). Here, acceptor and donor material are successively processed, leading to a planar interface between both materials. As long as the film thicknesses are below the exciton diffusion length of typically several nanometers, most excitons will reach the interface and will split into free charges. PHJ devices usually profit from good charge transport and low recombination losses, but suffer from limited carrier generation due to the small interfacial area. Despite the strong absorption of organic dyes, the mentioned thickness limit is not sufficient to absorb enough light. To overcome this limit another device architecture, the bulk heterojunction (BHJ), was introduced [7]. Donor and acceptor material are mixed with a certain ratio and can create an interwoven network of percolation paths of donor

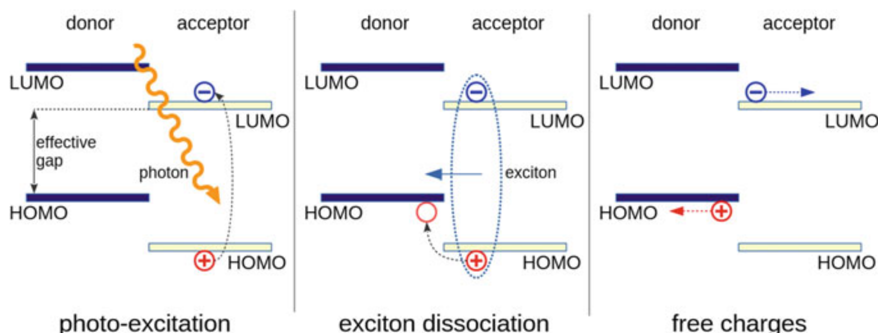


Fig. 10.4 Illustration of the energy levels in a heterojunction and the steps from photo-excitation to free charge carriers

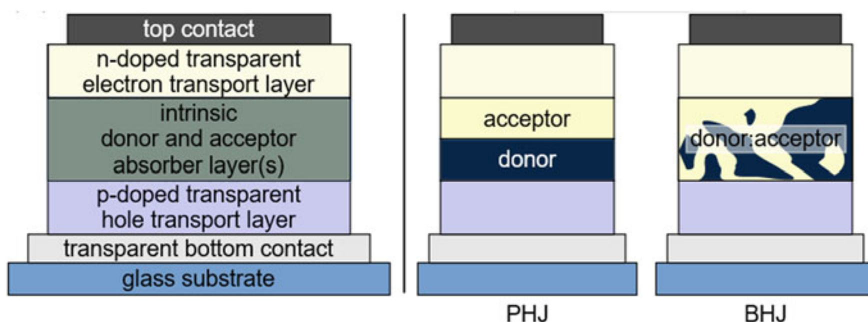


Fig. 10.5 Depiction of the p-i-n concept and two architecture types—planar and bulk heterojunction—used in typical organic solar cells

and acceptor domains. Since the interfaces in a BHJ are folded in the whole volume, the distance for an exciton to reach an interface is generally smaller, resulting in higher dissociation rates and accordingly more free charge carriers. Nowadays, the most efficient organic solar cells rely on multiple bulk heterojunctions [8, 9], however, new approaches in the form of cascade[2] solar cells [10] are very promising, too (Fig. 10.5).

According to Würfel [11], an ideal configuration of a solar cell is made of an absorber and semipermeable membranes which allow electrons and holes to reach only their respective electrodes. Today, high performance organic solar cells are built in a p-i-n structure which basically fulfills this concept. The letters p, i, and n indicate the sequence of layers and their purpose. The intrinsic absorber layer “i” is responsible for charge generation and is situated between the n-type layer—the

[2]The active layer consists of a sequence of at least three absorber materials, whereby adjacent materials form either a planar or bulk heterjunction.

electron transport layer (ETL)—and the p-type layer—the hole transport layer (HTL). Due to an energetic offset between the LUMO of the intrinsic and p-type layer, electrons are blocked while holes can pass. The same applies vice versa to the n-type layer, whose HOMO has an energetic step with respect to the HOMO of the absorber layer. It is rather unfavorable to have an absorbing ETL or HTL, since exciton diffusion lengths are generally small and a photoexcitation within the absorption layer is preferred. Therefore, wide gap materials are commonly used as transport layers in organic solar cells as they prevent parasitic absorption. The performance of an organic solar cell can be improved furthermore, by using p-and n-doped transport layers [12]. Doping is the targeted addition of specific atoms or molecules into a material to change its electrical behavior [13, 14]. A dopant, whose LUMO has the same energy level as or lower energy level than the HOMO of the matrix material, will create additional holes and is referred to as p-dopant. Analogously, an n-dopant will yield more electrons in the LUMO of the matrix material. Dopants lead to higher conductivities in the transport layers as well as to Ohmic contacts between electrodes and transport layers [15]. Organic solar cells typically have thicknesses in the range of a few hundreds of nanometers, i.e. within the coherence length of sunlight. An interference pattern is formed, which results in minima and maxima in the optical field distribution. This effect can be exploited by optimizing the thicknesses of the transparent transport layers and consequentially shifting the intensity maxima towards the active layer, as it will be described in the next section.

Transport layers are also beneficial to increase the maximal voltage for a given material system. When a solar cell is operated at open circuit conditions, the photo-generation of charge carriers leads to positive charging of the donor and a negative charging of the acceptor material, i.e. increasing the hole and electron densities, respectively. On the donor material, a higher hole density will shift the hole quasi Fermi level closer to the donor HOMO level. Accordingly, for higher electron densities on the acceptor, the electron quasi Fermi level will shift towards the LUMO level. In the ideal case, the difference between these quasi-Fermi levels defines the open-circuit voltage of the device. However, a direct contact of the heterojunction with the electrodes reduces the open-circuit voltage to the energetic difference of the contacts' work functions. Using transport layers as spacer layers decouples the heterojunction from the electrodes. This way the difference of the donor's HOMO energy and the acceptor's LUMO energy—the effective gap—will constitute the upper limit for the open-circuit voltage.

In summary it can be stated that a good energy level alignment, to ensure a good charge transport and low voltage losses, and a high complementary absorption of the donor and acceptor material to enable high photocurrents are the key ingredients for a high performance of an organic solar cell. However, in order to tap the full potential, there are several ways, e.g. different processing conditions, mixing ratios, etc. to further improve the power conversion efficiency, which will be presented in the following sections.

10.2.2 Thin Film Optics and Interference

The primary aim of every solar cell design is the absorption of as much light as possible as first step of the energy conversion. High absorption is primarily achieved by using highly absorbing materials like color-intensive dyes, in a thick enough layer to achieve possibly complete light absorption. However, the maximum feasible absorber thickness is limited by the conductivity of the material [16]: Too thick absorber layers will represent electric resistances for current extraction, leading to a reduction of the device efficiency. In an actual device, the optimum absorber thickness is a trade-off between light absorption and efficient charge extraction.

The limitation of the absorber thickness and the fact that every absorber only absorbs a limited wavelength range (see Fig. 10.3) imply that light is only partially absorbed upon one pass through the absorption layer [17]. The transmitted light is reflected at the metal back contact of the solar cell, allowing for a second pass through the absorber on its way back towards the incident surface.[3] Interestingly, this optical path length of the incident sunlight within an organic thin-film solar cell, which is typically in the range of tens or few hundreds of nanometers, is within the coherence length of sunlight: Due to the limited spectral width of the solar spectrum, a coherence length of approximately 600 nm is obtained [18]. In consequence, partial interference between the incident and reflected light can be observed and the resulting light intensity distribution exhibits clear maxima and minima of a standing wave pattern dominating over the propagating contribution. The effect of this interference on the performance of a solar cell is demonstrated in Fig. 10.6: By placing the absorber materials at various distances from the reflecting back contact via variation of the thickness of the electron transport material (ETM) of a p-i-n stack, they can be shifted into the maximum of the intensity distribution (symbolized by the solid arcs within the stack), or out of it [19]. The short-circuit current density j_{SC}, which equals the photo-current density j_{Ph}, can be used as a direct measure for the light absorption in the cell, because it is with a good accuracy proportional to the total absorbed photon flux. The proportionality is described by the external quantum efficiency (EQE) describing which share of photons leads to a successful extraction of electron-hole pairs from the solar cell, which is in most devices approximately independent of the local and global light intensity. Measuring finally j_{SC} from this series of solar cells with varying thickness of the ETM, the spatial distribution of the optical field strength can be directly mapped. The maximum in j_{SC} at an ETM thickness of 40 nm (C_{60} data) corresponds to the situation where the absorber layers are placed in the first interference maximum of the distribution. The minimum at 120 nm corresponds to the first interference minimum, where only the propagating fraction of the light contributes to

[3]As a secondary effect, another partial reflection at the transparent surface will furthermore reflect part of the remaining intensity back into the device, creating an effective cavity between front surface and back contact [15].

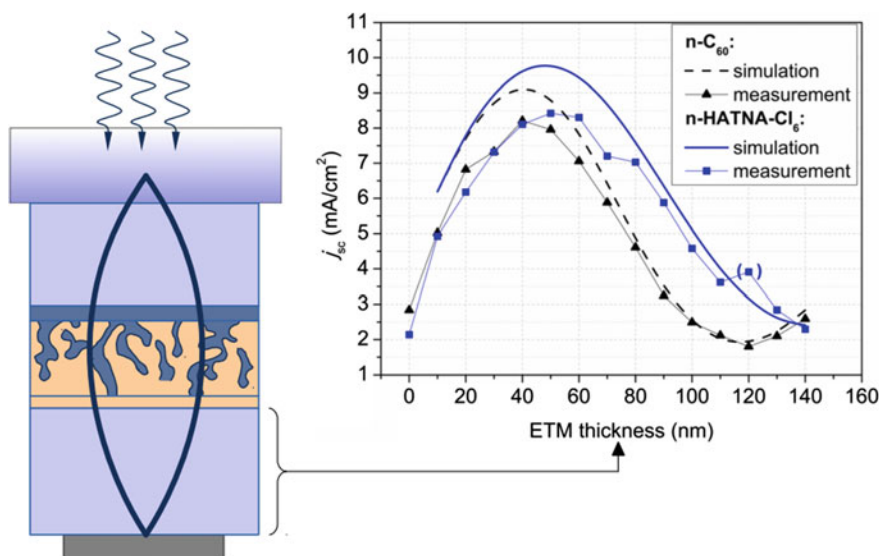


Fig. 10.6 Thin-film optical interference of incident light with light reflected at the back electrode. The thickness of the electron transport material (ETM) is varied, placing the absorber layers into and out of the maximum of the light intensity. The short-circuit current density j_{SC} is a measure for the absorbed light, which is highest if the absorber layers are in the maximum of the intensity distribution. Two different electron transport materials are investigated, n-doped C_{60} and n-doped HATNA-Cl_6. The measurement data (points) are compared to optical simulation data (*lines*). Picture (*left*) adapted from Christiane Falkenberg. Graph (*right*) reprinted from [19]

photo-current generation, and in the following two data points above 120 nm, the onset towards the second interference maximum can be observed.

This investigation illustrates the peculiarity of the extreme thin-film approach of organic photovoltaics, enabling the exploitation of thin-film optics for device tuning and optimization.

10.2.3 Morphology of the Blend Layer

The heart of photo-current generation in an organic solar cell is the heterojunction of the absorber materials donor and acceptor introduced above. The most successful concept for photon harvesting is the combination of both materials in a blend layer, the so-called bulk heterojunction (BHJ), or the subsequent deposition of the absorber materials with the first of them being rough enough to create a situation which is effectively similar to a bulk heterojunction. Both approaches lead to a finely folded three-dimensional interface between the donor phase and the acceptor phase of the heterojunction. The characteristic size of the geometric features constituting this heterojunction defines the area of this interface: the smaller the feature

size the larger the effective interface area. The feature size has an immediate impact on the resulting solar cell performance. If the features are too large, much larger than the exciton diffusion length within the material, excitons which are created in the middle of a feature fare away from any interface will probably recombine and be lost for electricity generation before reaching the separating interface. On the other hand, the donor-acceptor interface not only enables separation of charges, but also the recombination of electrons on the acceptor with holes on the donor, which represents a major loss mechanism for photo-current generation, rendering too fine structuring counterproductive for photo-current generation [20, 21]. Furthermore, in the case of a donor-acceptor blend layer, fine intermixing bares the risk of creating disconnected three-dimensional islands of either phase within the other phase, which obviously create a contribution to recombination, but not to current generation. Consequently, a too small feature size is equally detrimental for the solar cell performance. From the preceding discussion, the optimum feature size is expected in the range of twice the effective exciton diffusion length in the absorber materials.

For controlling the feature size of a blend layer respectively the roughness of a neat layer, the mobility of molecules during layer deposition is exploited. Here, vacuum and solution processed layers need to be distinguished technologically, although the resulting impact on the device physics is analogous. In vacuum deposition, molecules are continuously deposited on top of the already existing parts of the solar cell, i.e. they are initially at the surface towards vacuum. The dynamics of the binding to this surface can be described in a simplified way as two subsequent processes. First the molecule is adsorbed at the surface but still mobile to move across the surface, experiencing an energetic landscape with local minima and maxima due to the morphologic configuration of the surface. Secondly when coming across a local minimum deeper than its thermal energy, or when being restricted by other molecules deposited in the vicinity, the molecule is spatially confined to its final place in the film. In solution processing, these steps are analogous, with an initially partially mobile state of the molecules as long as the solvent is still present in the film, and a spatial confinement of molecules after evaporation of the solvent when the film has dried. In neat layers of materials which tend to crystallize, it is instructive that a higher or longer mobility of molecules enables the formation of larger crystallites. These crystallites can pile the molecules up into the third dimension and thus create a rough surface. In blend layers of two materials being deposited at the same time crystallization can equally take place, and it is accompanied by a second often dominating effect, the de-mixing of material phases. As the surface energy of two different materials is generally different, given high enough spatial mobility, they tend to de-mix over time similar to a mixture of e.g. a hydrophilic and a hydrophobic liquid. Initially, the molecules are deposited in a random distribution onto the sample surface, and with time they de-mix and form separate phases for each material. In summary, the coarseness of the intermixing is directly correlated to the spatial mobility of the molecules during film formation. In solution processing, this parameter is typically controlled by the selection of the solvent and the drying conditions, in vacuum processing by the substrate temperature during deposition. An impressive example showing the morphology change of

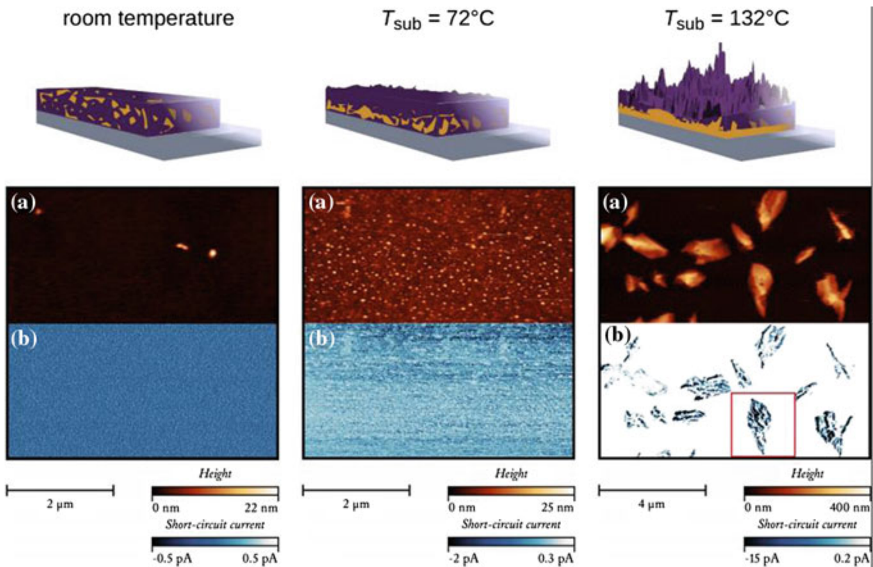


Fig. 10.7 Morphology of blend layers deposited onto different substrate temperatures T_{sub}. (All measurements are consequently performed at room temperature.) *Top* artistic view. *Below* Microscopic images by atomic force microscopy: **a** height profile **b** photo-current distribution. The intermediate substrate temperature of 72 °C yields in this series the optimum morphology for a solar cell Pictures by Chris Elschner (*top*) and Tobias Mönch (*bottom*) [22]

vacuum deposited blend layers upon variation of the substrate temperature is shown in Fig. 10.7 for the material system is DCV5T (Fig. 10.3) blended with C_{60} [22].

10.2.4 Optimized p-i-n Cells

Putting things together, an efficient solar cell can be designed with the following ingredients: (a) highly absorbing materials in a (b) p-i-n stack with matching transparent transport materials and (c) optically well-adjusted layer thicknesses, all processed in a way to achieve (d) optimum layer morphology.

This approach was followed by Meerheim, Körner, et al. using the previously mentioned absorber system DCV5T:C_{60} to achieve an impressive single junction power conversion efficiency (PCE) of 8.3% [8, 23]. Looking into the details of this device, a high EQE above 70% is observed at a wavelength of 580 nm (see Fig. 10.8), which corresponds to the maximum in absorption of the donor material DCV5T. It proves the high absorption strength of the donor material and the high internal quantum efficiency, i.e. the majority of the photons absorbed by the donor-acceptor system are successfully converted into electron-hole pairs and extracted as electric current, leading finally to a remarkable short-circuit current

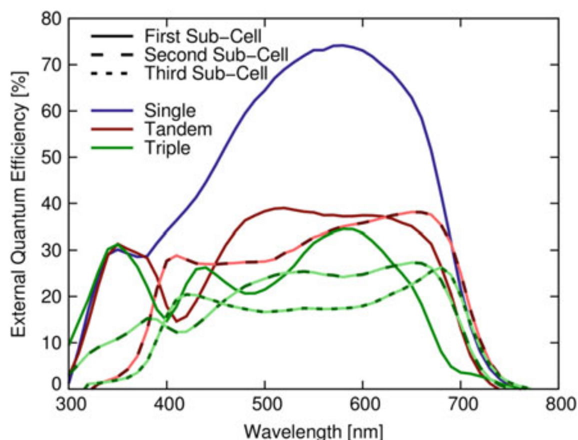


Fig. 10.8 Spectrally resolved external quantum efficiency (EQE) of solar cells based on the absorber material system DCV5T:C_{60}, realized as a single-junction cell (*blue*), a tandem cell (*red*) and a triple-junction cell (*green*). For the multi-junction cells, the EQE data of the sub-cells is shown separately. Data previously published in [8]

density of 13.2 mA/cm^2. The achieved open-circuit voltage of 0.97 V is a typical value expected for an organic material system with the given effective gap [21, 24], and the high fill factor of 65.8% proves the well optimized transport properties of the material system.

Looking at the factors limiting the efficiency of this device, a further discussion of the EQE is instructive: Outside the maximum position, it quickly drops to lower values and light absorption is far from complete in the wavelength ranges outside the EQE maximum, which has a width of approximately 300 nm. In consequence, much of the light leaves the cell unused, and increasing the amount of absorber material in the device or even the combination with other dyes absorbing in different wavelength regions would be desirable. These considerations lead to the concept of multi-junction devices as discussed in the following.

10.2.5 Tandem and Multi-junction Cells—Maximizing the Power Output

Boosting the output power of solar cells, a straight-forward approach is to stack several cells on top of each other. The reason is that every solar cell—independent of the used technology—absorbs the incident light only partially, and even uses the energy of the incident photons only partially [11]. These deficiencies are described by the limited usable wavelength range, e.g. the absorption onset of the semiconductor’s energy gap and by thermalization of electron-hole pairs towards the energy gap as well as incomplete absorption of the usable wavelength range, as discussed above. Thermalization losses can on the one hand be reduced by using small-band absorbers, on the other hand, this leads to an even increased share of the light intensity passing through the cell without being absorbed. Placing a second solar cell, which is optimized for the illumination spectrum transmitted by the first cell,

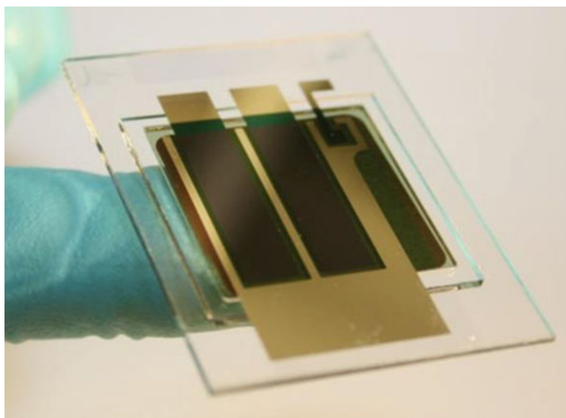


Fig. 10.9 Highly efficient organic solar cell by Heliatek

behind the first solar cell allows for harvesting this remaining intensity (Fig. 10.9).[4]

The primary requirement for the practical realization of this concept is that the light, which is not used by the first cell, is not absorbed by the back contact but rather transmitted to the second cell. This can best be assured by integrating both cells into one single device with only one pair of external contacts, a so-called tandem solar cell consisting of two sub-cells respectively junctions. In organic thin-film solar cells, especially if deposited by vacuum evaporation, this concept can be realized in a straight-forward manner by depositing two cells on top of each other [15]. The electrical contact between them needs to be transparent and can be realized e.g. by a recombination contact consisting of a highly p-doped and a highly n-doped organic semiconductor layer as shown in Fig. 10.10 [25]. These doped layers are part of a p-i-n stack anyway, and integration can be done with minor adaptions to the stack.

Such direct series connection of two solar cells brings along major benefits for the energy harvesting, but also new challenges for device design. Beneficially, two devices in series add their output power in terms of voltage, while the current density is not higher than in a single cell. Ohmic losses—they are a relevant factor in organic semiconductors and in transparent contacts with limited conductivity, especially upon up-scaling for commercial use—are proportional to the current density, but independent of the cell voltage, i.e. their detrimental effect is minimized by the tandem cell design leading generally to a preservation of the fill factor. The series connection of two cells in a tandem junction furthermore implies that the same current flows through both devices at any time [26]. This latter property constitutes the main rationale for device design of efficient tandem devices. The

[4]This concept can be extended beyond two cells, however, the further discussion is limited to two sub-cells where it covers all relevant issues.

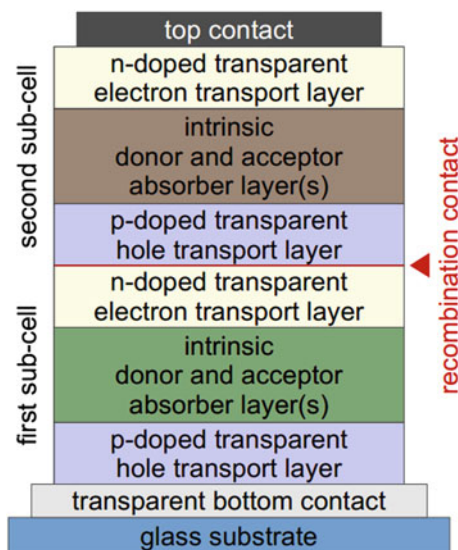


Fig. 10.10 Schematic of a tandem solar cell stack (*illumination from bottom*) consisting of two p-i-n type sub-cells which are connected in series by a recombination contact (*highlighted in red*) formed by the adjacent n- and p-doped transport layers or the respective sub-cells

current density at the maximum power point (MPP[5]) of all sub-cells must be equal to ensure that all sub-cells actually operate at the MPP simultaneously. Referring to the discussion above that the second sub-cell can only harvest the light which is transmitted through the first sub-cell,[6] it is instructive that the thickness of the absorber layers in the first sub-cell can be reduced, thereby reducing the current density from this sub-cell and simultaneously increasing its optical transmission and consequently the current density in the second sub-cell. At the optimum combination of thicknesses, current matching is achieved and the output current density at MPP is equal for both sub-cells and the maximum possible electric energy is generated from the absorbed light.

This approach of multi-junction cells can be used with the same absorber system for all sub-cells to compensate the limited absorption of a single cell, designing co-called homo-multijunction devices. This way, the efficiency of a DCV5T:C_{60} device can be increased from the previously mentioned 8.3% in a single junction cell to 9.6% in a tandem device and 9.7% in a triple junction device [8]. Even more beneficial is the combination of different absorber material systems with complementary absorption in a hetero-multijunction device. Meerheim et al. demonstrated a triple junction cell with a hybrid homo-hetero-triple-junction approach based on the previously discussed device, by introducing the long-wavelength absorber

[5]The MPP denotes the operation voltage, at which the output power, i.e. the product of the output voltage and the output current, is highest. It is found at a voltage slightly below the open-circuit voltage, with its actual position depending on the specific device and its operating conditions.

[6]In this simplified description, reflection from the back contact is omitted. In a more detailed view, again, the full interference pattern needs to be taken into account, as discussed above for a single junction cell.

Ph2-benz-bodipy (Fig. 10.3) in the third sub-cell, boosting the power conversion efficiency to 10.4% [9]. In a similar approach with different materials, Che et al. even reported an efficiency of 11.1%, which is today one of the highest efficiencies reached with a disclosed device structure [27]. The current world record for the efficiency of organic solar cells was set by the company Heliatek in a commercially developed vacuum deposited multi-junction device with an efficiency of 13.2% [28].

10.3 Perovskite Photovoltaics

Recently, a novel class of thin-film photovoltaics has generated large interest: organic-inorganic perovskites [29] have taken an enormous development in efficiency and have grown from a few percent efficiency to more than 20% efficiency in just a few years. This rate of efficiency growth is by far the highest which has ever been achieved by a photovoltaic system and raises expectations that the technology might become the most efficient thin film photovoltaics technology available. The development of perovskite photovoltaics started with first reports in 2009 that these materials can generate efficiencies of about 3% [30]. Then the development was taken up by many groups [31, 32] and within a few years, highly efficient cells [33] and a certified record value of 20.1% was reached. The efficiency of these perovskites is thus almost as large as for the inorganic thin film systems CdTe and CIGS. One should keep in mind though that some record values reported for perovskites have to be taken with some caution because the materials tend to show some hysteresis in their IV-curves. Therefore, the shape of the IV-curve is strongly dependent whether it is taken in forward or backward direction and at which speed, potentially causing errors in efficiency determination. A recent study has shown that this hysteresis is probably called by charge storage in the device. Therefore, some of the literature results cannot really be trusted.

Nevertheless, the certified efficiency of 20.1% shows the very large potential of the material. The main reason why such a high efficiency was reached is that the material shows rather large diffusion length and high mobilities for a thin film system. Furthermore, it seems to have unusually low recombination both at surfaces and grain boundaries which allows to collect the current very efficiently and to achieve high voltages. Figure 10.11 shows the image of a typical high efficiency perovskite structure. Most of the perovskite solar cells realized so far are an extension of the dye sensitized solar cell concept, containing a titanium dioxide sponge which is infiltrated with the perovskite. However, it has been shown in various reports that perovskite solar cells also perform very well as flat heterojunction cells so that the dye sensitized principle is not necessary for high efficiency in that system.

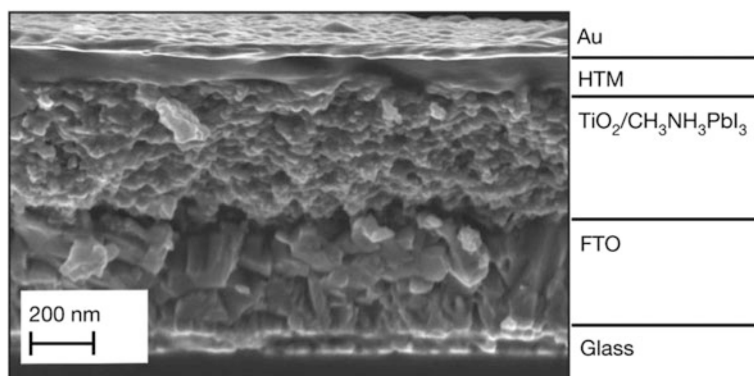


Fig. 10.11 Structure of a perovskite solar cell [32]

Despite these this excellent progress, there are a number of open questions:

- The so far best material contains lead which might not be tolerable due to its toxicity.
- The stability of the perovskite solar cells has not been investigated in much detail. Recent results indicate that the material has potential for larger stability, however, results are not sufficient to prove that a widespread use is possible.
- The manufacturing processes for perovskite solar cells still are mostly at a laboratory stage and the techniques for mass production have to be developed.

Despite these shortcomings, research on organic-inorganic perovskites is pursued by many groups worldwide and it can be expected that at least some of the problems mentioned above can be solved soon. We here do not further discuss the perovskites, since we concentrate on organic photovoltaics. It should be mentioned, however, that the perovskite absorber material can be combined very well with organic semiconductors. Usually organic semiconductors are used as hole transport materials in perovskite cells. The hole transport material needs to be adjusted very closely with the active material to achieve high efficiency, but techniques known from organic solar cells such as doping can be employed as well [34].

10.4 Application of Different Solar Cell Technologies

10.4.1 Application Scenarios

An idealized case of solar power generation for solar cells is described by the standard test conditions (STC) [35]. In detail, the environmental factors, the positioning, and the spectral quantity and quality of the incident light source are specified. The temperature of the solar cell T_c is defined as $T_c = 25$ °C. The

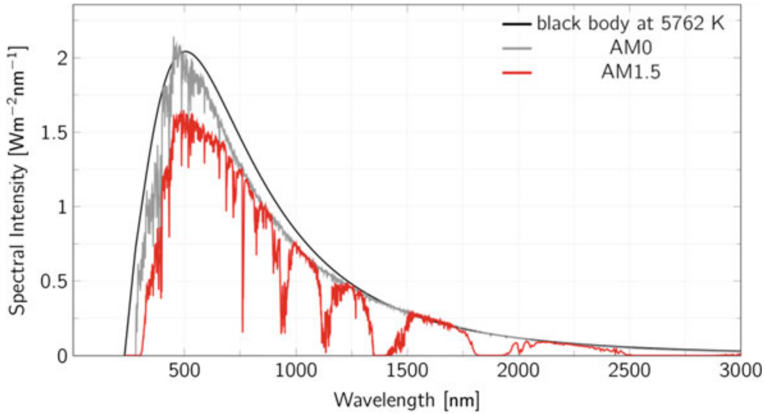


Fig. 10.12 Comparison of the black-body radiation at a temperature of T = 5762 K, an extraterrestrial sun spectrum (AM0) [48], and the AM1.5 spectrum [49]

intensity of incident light perpendicular to the solar cell surface is defined as $E_0 = 1000\ W/m^2$. Typically, the shape of the illumination spectrum is set to an "Air Mass 1.5" (AM1.5) spectrum. An air mass of 1.5 describes the sun spectrum with the absorption from the atmosphere subtracted, assuming that the light passes through the atmosphere at an inclination rendering its path 1.5 times as long as in the case of perpendicular irradiation. The extraterrestrial shape of the sun spectrum is called AM0. A comparison of these spectra is shown in Fig. 10.12.

These standardized conditions are used for the measurement of the solar cell's fingerprints (open-circuit voltage V_{oc}, short-circuit current density j_{sc}, maximum power point MPP, and fill factor FF) and the power conversion efficiency (PCE). The PCE at STC is the baseline for comparing different solar cells and technologies.

However, these idealized conditions are an exception, since they are only fulfilled for a position at a latitude of approximately 48° on days with clear sky and a low ambient temperature for optimally tilted solar cells at noon. Additionally, these conditions apply only for flat roofs without shadowing.

In actual photovoltaic installations, at most times conditions are strongly different from the STC. Thus, the question arises, how to characterize solar cells for realistic conditions and not only for the standardized conditions. Therefore, the harvesting factor H of a solar cell at the location-specific weather and climate conditions is required in addition to the PCE under STC conditions. The harvesting factor can be expressed as the PCE under realistic conditions PCE^{real} in relation to the PCE under STC PCE^{STC}.

$$H(t) = \frac{PCE^{real}(t)}{PCE^{STC}} \qquad (10.1)$$

The harvesting factor of a solar cell accounts for the realistic conditions and redefines the point of view from "What are the optimal conditions for a solar cell

technique?” to “Which solar cell technique is the best to be used in this specific situation?”. For example, one of the most challenging situations would be a solar cell mounted vertically at a house façade oriented to the north. In this situation the largest amount of illumination is diffuse sun light. The average sun intensity is much lower than considered for STC. Here, a solar cell that has a low efficiency under STC but a good low light performance might be better in comparison to a cell with a high PCE^{STC} but a deficient low light performance.

The harvesting factor summarizes the effects of the environmental conditions on the solar cell efficiency and gives a realistic estimation for the expected generated power.

10.4.2 Energy Harvesting Under Real Application Conditions

The harvesting factor is a tool to decide which solar cell technology suits best for a specific application scenario. At the “Institut für Angewandte Photophysik” (IAPP) in Dresden (51.0°N, 13.7°E), Germany, a long-term outdoor measurement of a mono-crystalline solar module (c-Si, Hanwha Q Cells Q. Peak-G3 265, 1.67 m^2), an amorphous silicon solar module (a-Si, Signet Solar Si-S4, 1.37 m^2) and a copper indium gallium diselenide solar module (CIGS, Avancis FB 110, 1.07 m^2) is performed. Additionally, an organic solar cell (DCV5T as a donor material in a blend with C_{60}[7]) with an aperture area of 6.44 mm^2 are compared to the inorganic types.[8] The surface of the modules and cells faces in southward direction and is tilted by an angle of 33° with respect to the earth surface. The STC efficiencies of all five devices are shown in Fig. 10.16. The illumination intensity (I) is measured by a pyranometer[9] for the OPV and an irradiance reference sensor[10] for the inorganic techniques. The reference sensor is calibrated to the pyranometer (cf. Fig. 10.13). The ambient temperature is measured by a weather station and the solar cell temperature (T_c) is calculated.[11]

---

[7]The stack sequence is as follows: indium tin oxide/ N,N-Bis(fluoren-2-yl)-naphthalenetetracarboxylic diimide [36] n-doped with 7wt% (weight-percent) tetrakis(1,3,4,6,7,8-hexahydro-2H-pyrimido[1,2-a]pyrimidinato)ditungsten (II) [37], [38] (thickness 5 nm)/ C60 (15 nm)/ DCV5T: C60 (mixing ratio 2:1 by volume, 40 nm)/ 9,9-bis[4-(*N*,*N*-bis-biphenyl-4-yl-amino)phenyl]-9*H*-fluorene) (BPAPF [39]) (5 nm)/ BPAPF p-doped with 10wt% NDP9 (commercial p-dopant by Novaled) (30 nm)/ N,N'-((Diphenyl-N,N'-bis)9,9,-dimethyl-fluoren-2-yl)-benzidine [40] p-doped with 10 wt% NDP9 (10 nm)/ NDP9 (1 nm)/ Al (100 nm).

[8]The OPV cell is measured with a Keithley 2400 source-measure unit, whereas the large modules are measured with a PVPM 2540C source-measure unit.

[9]Kipp and Zonen CMP 11.

[10]SOZ-03.

[11]$T_c = T_{ambient} + (T_{NOCT} - 20) * (I/ 800\ W/m^2)$ [41].

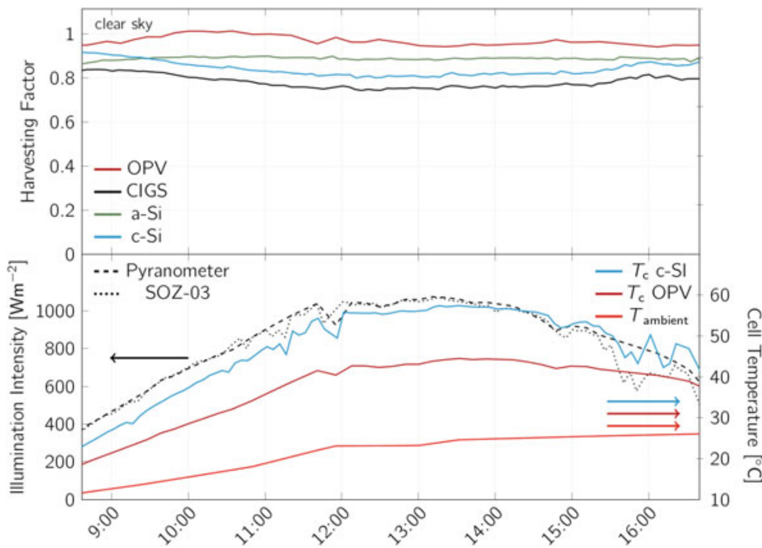


Fig. 10.13 Harvesting factor for clear sky conditions on August 26th, 2015 for CIGS, a-Si, and c-Si modules and the OPV solar cell and the corresponding illumination intensity, ambient temperature and cell temperature

Figures 10.13 and 10.14 show the harvesting factor for two representative days, one with clear and one with overcast sky conditions, respectively. A harvesting factor of H = 1 means that the real power conversion efficiency is identical to the power conversion efficiency at STC (10.1). Obviously, all cells and modules operate with a PCE^{real} below the PCE^{STC} most of the time. While at clear sky conditions (Fig. 10.13) the harvesting is reduced at noon for the CIGS and the c-Si module, the a-Si module and the OPV cell show almost no drop in efficiency.

The effect of a reduced harvesting factor at the noon might be explained with a PCE loss caused by high cell temperatures for crystalline technologies [42, 43]. The OPV cell and the a-Si module are much less affected by an increase of the solar cell temperature [44].

Also at overcast sky conditions the OPV solar cell shows the highest harvesting factor and operates partly better than under STC. For overcast conditions in the afternoon, the harvesting factor of the OPV solar cell slightly increases. All other technologies show a decrease in the harvesting factor with the highest loss observed for the CIGS module. These differences might be explained by a robust diffuse light performance of organic solar cells in contrast to the other technologies [17, 45, 46].

Figure 10.15 shows the harvesting factor H_d averaged separately over each day as function of the solar irradiance W_d integrated over the respective day. High values of W_d correspond to clear sky conditions, whereas low values stand for overcast conditions.

The organic solar cells show the highest harvesting factors for all conditions. The a-Si solar module shows a comparable harvesting factor for a large W_d, but

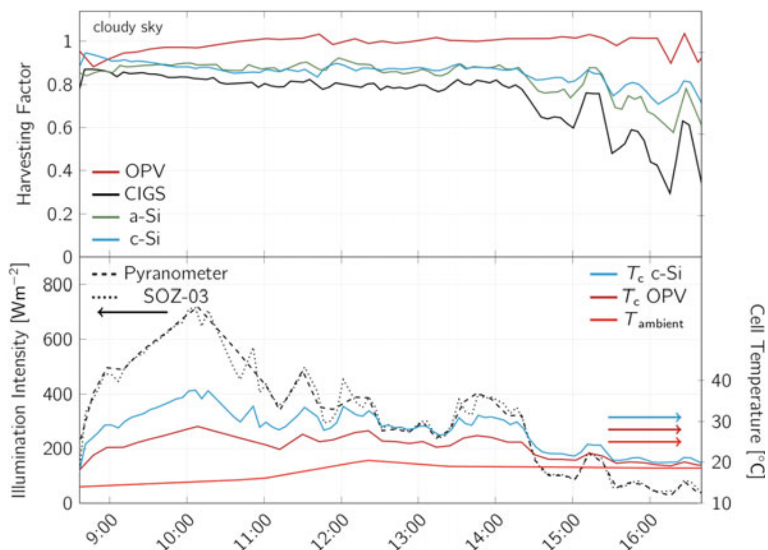


Fig. 10.14 Harvesting factor for overcast sky conditions on September 3rd, 2015 for CIGS, a-Si, and c-Si solar modules and the OPV cell and the corresponding illumination intensity, ambient temperature and cell temperature

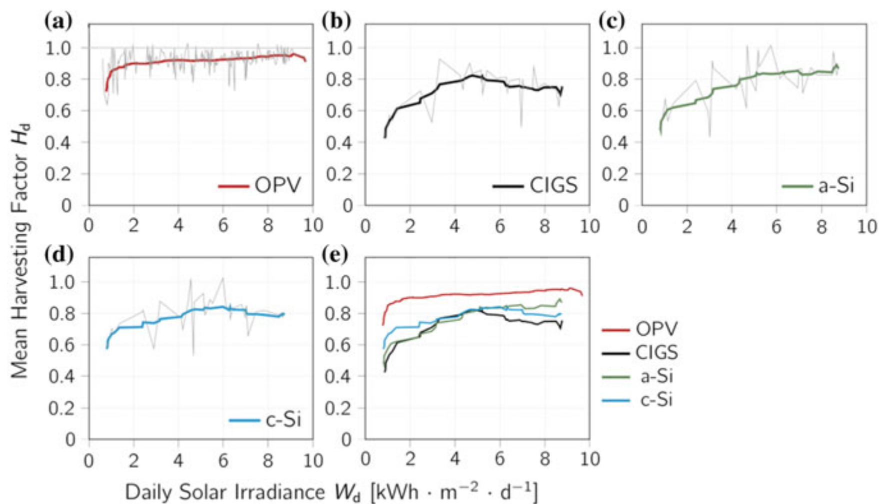


Fig. 10.15 Day by day average of the harvesting factor in dependence of the incident solar irradiance integrated over the respective day. **a–d** show the data of each cell/module separately as thin grey lines, with moving average curves as thick colored lines added to highlight the relevant trends. The OPV cell is measured from May to October, the inorganic modules from August to October. **e** shows a comparison of the moving averages of the four devices

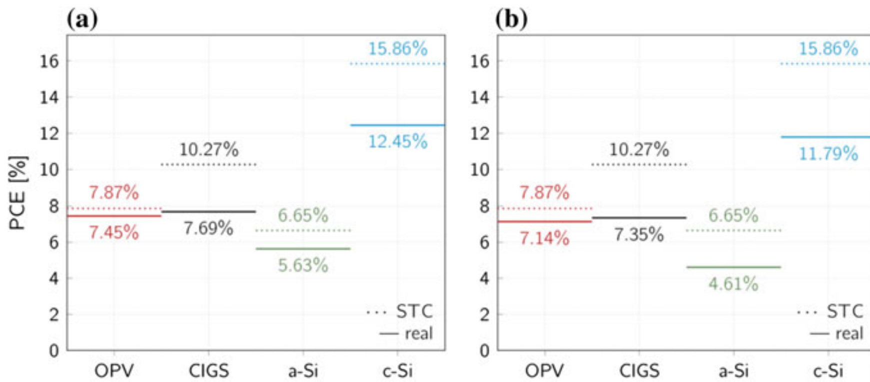


Fig. 10.16 Comparison of all real PCE (*solid lines*) to their STC PCE (*dotted lines*) for the four solar cells and modules. **a** shows the PCE for the clear sky conditions on August 26th, 2015 and **b** shows the PCE for the cloudy conditions on September 3rd, 2015

rapidly drops for a decreasing daily solar irradiance. A similarity of the organic solar cells and the a-Si solar module is the amorphous character of the active material in contrast to the crystalline character of the CIGS and the c-Si technique. The CIGS and the c-Si modules have their maximum H_d at medium W_d around 5 kWh/(m^2d), slightly decrease for higher W_d, and have significant losses at very low solar irradiances, i.e. on very cloudy days.

Figure 10.16 shows the comparison of PCE^{real} and the PCE^{STC} for two different irradiance conditions. In all cases the realistic efficiencies are lower than their expected STC efficiencies. The organic solar cells are less affected by a low daily irradiance, whereas the inorganic modules have a dramatic loss of efficiency, especially at a low irradiance of W_d of 2 kWh/(m^2d). Under non-optimal conditions, the STC values are far away from realistic efficiencies. Thus, for a realistic estimation of the energy yield, solar cells and modules need to be measured with a more sensitive method taking the local environmental influences into account.

For the particular situation described in the previous chapter, organic solar cells show the most advantageous behavior in contrast to the inorganic technologies. The better low light performance leads to an improved solar energy harvesting factor in these situations. Inorganic techniques are particularly suitable in situations with low ambient temperatures and high illumination intensities. In summary, only a realistic characterization of solar cells and modules in varying environmental conditions enables the prediction of the expected solar power generation and the choice of the suitable solar technique.

10.5 Conclusion

This chapter has covered novel thin-film photovoltaic technologies, such as organics and perovskites. While these materials have great potential due to their potential low cost, deposition on flexible substrates etc., a commercial breakthrough has so far not been achieved. The first broad application which might emerge is the usage of organic photovoltaics in building-integrated photovoltaics (BIPV): Here, the adjustable transparency and color, as well as the light weight, are key features which distinguish this technology. For further broad applications, organic PV needs higher efficiencies. The key here is to find improved materials and morphologies for the active absorber materials, to increase absorption and reduce the comparatively high voltage losses of the organic PV technology.

A rather different situation is present for the organic-inorganic technology of perovskite photovoltaic: this technology might in future share the low-cost deposition methods compatible with roll-to-roll processing, but has achieved a rapid efficiency growth which was never observed before in the field of photovoltaics: Today, the cells have achieved efficiencies beyond 20%, and if this development will continue only for a short time, perovskites will be the most efficient thin-film technology. The hurdles to achieve a commercial breakthrough are to realize stable and hysteresis-free cells, ideally without lead or other poisonous elements.

Compared to the established crystalline silicon technology, the novel thin-film technologies are still very young and there will be much more research needed to fully exploit their potential. This potential, however, is excellent and warrants to pursue this research.

References

1. D.M. Chapin, C.S. Fuller, G.L. Pearson, A new silicon pn junction photocell for converting solar radiation into electrical power. J. Appl. Phys. (1954)
2. P. Jackson, D. Hariskos, R. Wuerz, O. Kiowski, A. Bauer, T. M. Friedlmeier, M. Powalla, Properties of Cu(In, Ga) Se2 solar cells with new record efficiencies up to 21.7%. Phys. status solidi—Rapid Res. Lett. **9**(1), 28–31 (2015)
3. C.J. Mulligan, M. Wilson, G. Bryant, B. Vaughan, X. Zhou, W.J. Belcher, P.C. Dastoor, A projection of commercial-scale organic photovoltaic module costs. Sol. Energy Mater. Sol. Cells **120**(PART A), 9–17 (2014)
4. R. Gresser, M. Hummert, H. Hartmann, K. Leo, M. Riede, Synthesis and characterization of near-infrared absorbing benzannulated aza-BODIPY dyes. Chem. A Eur. J. **17**(10), 2939 (2011)
5. S. Kraner, J. Widmer, J. Benduhn, E. Hieckmann, T. Jägeler-Hoheisel, S. Ullbrich, D. Schütze, K.S. Radke, G. Cuniberti, F. Ortmann, M. Lorenz-Rothe, R. Meerheim, D. Spoltore, K. Vandewal, C. Körner, K. Leo, Influence of side groups on the performance of infrared absorbing aza-BODIPY organic solar cells. Phys. status solidi **212**(12), 2747–2753 (2015)
6. C.W. Tang, Two-layer organic photovoltaic cell. Appl. Phys. Lett. **48**(2), 183–185 (1986)
7. M. Hiramoto, H. Fujiwara, M. Yokoyama, Three-layered organic solar cell with a photoactive interlayer of codeposited pigments. Appl. Phys. Lett. **58**(10), 1062 (1991)

8. R. Meerheim, C. Körner, K. Leo, Highly efficient organic multi-junction solar cells with a thiophene based donor material. Appl. Phys. Lett. **105**(6), 63306 (2014)
9. R. Meerheim, C. Körner, B. Oesen, K. Leo, 10.4% Efficient triple organic solar cells containing near infrared absorbers. Appl. Phys. Lett. **108**(10) (2016)
10. K. Cnops, B.P. Rand, D. Cheyns, B. Verreet, M.A. Empl, P. Heremans, 8.4% efficient fullerene-free organic solar cells exploiting long-range exciton energy transfer. Nat. Commun. **5**, 3406 (2014)
11. P. Wurfel, U. Wurfel, *Physics of solar cells : from basic principles to advanced concepts* (2009)
12. K. Walzer, B. Maennig, M. Pfeiffer, K. Leo, Highly efficient organic devices based on electrically doped transport layers. Chem. Rev. **107**(4), 1233–1271 (2007)
13. M. Pfeiffer, A. Beyer, T. Fritz, K. Leo, Controlled doping of phthalocyanine layers by cosublimation with acceptor molecules: A systematic Seebeck and conductivity study. Appl. Phys. Lett. **73**(22), 3202 (1998)
14. B. Lüssem, M. Riede, K. Leo, Doping of organic semiconductors. Phys. Status Solidi (A) **210** (1), 9–43 (2013)
15. B. Maennig, J. Drechsel, D. Gebeyehu, P. Simon, F. Kozlowski, A. Werner, F. Li, S. Grundmann, S. Sonntag, M. Koch, K. Leo, M. Pfeiffer, H. Hoppe, D. Meissner, N.S. Sariciftci, I. Riedel, V. Dyakonov, J. Parisi, Organic p-i-n solar cells. Appl. Phys. A Mater. Sci. Process. **79**(1), 1–14 (2004)
16. J. Widmer, J. Fischer, W. Tress, K. Leo, M. Riede, Electric potential mapping by thickness variation: A new method for model-free mobility determination in organic semiconductor thin films. Org. Electron. **14**(12), 3460–3471 (2013)
17. M. Riede, C. Uhrich, J. Widmer, R. Timmreck, D. Wynands, G. Schwartz, W.M. Gnehr, D. Hildebrandt, A. Weiss, J. Hwang, S. Sundarraj, P. Erk, M. Pfeiffer, K. Leo, Efficient organic tandem solar cells based on small molecules. Adv. Funct. Mater. **21**(16), 3019–3028 (2011)
18. A. Donges, The coherence length of black-body radiation. Eur. J. Phys. **19**(3), 245–249 (1998)
19. C. Falkenberg, K. Leo, M.K. Riede, Improved photocurrent by using n-doped 2,3,8,9,14,15-hexachloro-5,6,11,12,17,18-hexaazatrinaphthylene as optical spacer layer in p-i-n type organic solar cells. J. Appl. Phys. **110**(12), 124509 (2011)
20. K. Vandewal, K. Tvingstedt, A. Gadisa, O. Inganäs, J.V. Manca, Relating the open-circuit voltage to interface molecular properties of donor: acceptor bulk heterojunction solar cells. Phys. Rev. B **81**(12), 1–8 (2010)
21. K. Vandewal, J. Widmer, T. Heumueller, C.J. Brabec, M.D. McGehee, K. Leo, M. Riede, A. Salleo, Increased open-circuit voltage of organic solar cells by reduced donor-acceptor interface area. Adv. Mater. **26**(23), 3839–3843 (2014)
22. T. Mönch, Exploring nanoscale properties of organic solar cells. Ph.D. thesis, TU Dresden (2015)
23. C. Koerner, Oligothiophene materials for organic solar cells—photophysics and device properties. Ph.D. thesis, TU Dresden (2013)
24. J. Widmer, M. Tietze, K. Leo, M. Riede, Open-circuit voltage and effective gap of organic solar cells. Adv. Funct. Mater. **23**(46), 5814 (2013)
25. R. Timmreck, S. Olthof, K. Leo, M. Riede, Highly doped layers as efficient electron–hole recombination contacts for tandem organic solar cells. J. Appl. Phys. **108**(3), 33108 (2010)
26. R. Schueppel, R. Timmreck, N. Allinger, T. Mueller, M. Furno, C. Uhrich, K. Leo, M. Riede, Controlled current matching in small molecule organic tandem solar cells using doped spacer layers. J. Appl. Phys. **107**(4), 44503 (2010)
27. X. Che, X. Xiao, J.D. Zimmerman, D. Fan, S.R. Forrest, High-efficiency, vacuum-deposited, small-molecule organic tandem and triple-junction photovoltaic cells. Adv. Energy Mater. **4** (18), 1400568 (2014)
28. Heliatek GmbH (2013). www.heliatek.com
29. D.B. Mitzi, *Synthesis, Structure, and Properties of Organic-Inorganic Perovskites and Related Materials* (2007)

30. A. Kojima, K. Teshima, Y. Shirai, T. Miyasaka, Organometal halide perovskites as visible-light sensitizers for photovoltaic cells. J. Am. Chem. Soc. **131**(17), 6050–6051 (2009)
31. M. Liu, M.B. Johnston, H.J. Snaith, Efficient planar heterojunction perovskite solar cells by vapour deposition. Nature **501**(7467), 395–398 (2013)
32. J. Burschka, N. Pellet, S.-J. Moon, R. Humphry-Baker, P. Gao, M.K. Nazeeruddin, M. Grätzel, Sequential deposition as a route to high-performance perovskite-sensitized solar cells. Nature **499**, 316–319 (2013)
33. N.J. Jeon, J.H. Noh, W.S. Yang, Y.C. Kim, S. Ryu, J. Seo, S. Seok, Compositional engineering of perovskite materials for high-performance solar cells. Nature **517**(7535), 476–480 (2015)
34. L.E. Polander, P. Pahner, M. Schwarze, M. Saalfrank, C. Koerner, K. Leo, Hole-transport material variation in fully vacuum deposited perovskite solar cells. APL Mater. **2**(8), 81503 (2014)
35. IEC 60904-3: Measurement Principles for Terrestrial PV Solar Devices with Reference Spectral Irradiance Data (1989)
36. C. Falkenberg, Optimizing organic solar cells transparent electron transport materials for improving the device performance. Ph.D. thesis, TU Dresden (2011)
37. T. Menke, *Molecular doping of organic semiconductors—a conductivity and seebeck study*. Ph.D. thesis, TU Dresden: Verlag Dr. Hut (2013)
38. F.A. Cotton, N.E. Gruhn, J. Gu, P. Huang, D.L. Lichtenberger, C.A. Murillo, L.O. Van Dorn, C.C. Wilkinson, Closed-shell molecules that ionize more readily than cesium. Science (80-) **298**(5600), 1971–1974 (2002)
39. Y.-L. Tung, S.-W. Lee, Y. Chi, Y.-T. Tao, C.-H. Chien, Y.-M. Cheng, P.-T. Chou, S.-M. Peng, C.-S. Liu, Organic light-emitting diodes based on charge-neutral Os(II) emitters: Generation of saturated red emission with very high external quantum efficiency. J. Mater. Chem. **15**(4), 460 (2005)
40. C. Murawski, C. Fuchs, S. Hofmann, K. Leo, M.C. Gather, Alternative p-doped hole transport material for low operating voltage and high efficiency organic light-emitting diodes. Appl. Phys. Lett. **105**(11) (2014)
41. M.C. Alonso García, J.L. Balenzategui, Estimation of photovoltaic module yearly temperature and performance based on Nominal Operation Cell Temperature calculations. Renew. Energy **29(**12), 1997–2010 (2004)
42. E. Cuce, P.M. Cuce, T. Bali, An experimental analysis of illumination intensity and temperature dependency of photovoltaic cell parameters. Appl. Energy **111**, 374–382 (2013)
43. M.G. Deceglie, T.J. Silverman, B. Marion, S.R. Kurtz, Metastable changes to the temperature coefficients of thin-film photovoltaic modules, in *2014 IEEE 40th Photovoltaic Specialist Conference, PVSC 2014*, pp. 337–340 (2014)
44. E.A. Katz, D. Faiman, S.M. Tuladhar, J.M. Kroon, M.M. Wienk, T. Fromherz, F. Padinger, C.J. Brabec, N.S. Sariciftci, Temperature dependence for the photovoltaic device parameters of polymer-fullerene solar cells under operating conditions. J. Appl. Phys. **90**(10), 5343–5350 (2001)
45. A. Parretta, A. Sarno, L.R.M. Vicari, Effects of solar irradiation conditions on the outdoor performance of photovoltaic modules. Opt. Commun. **153**(1–3), 153–163 (1998)
46. M. Riede, C. Uhrich, R. Timmreck, J. Widmer, D. Wynands, M. Levichkova, M. Furno, G. Schwartz, W. Gnehr, M. Pfeiffer, K. Leo, Optimization of organic tandem solar cells based on small molecules, in *2010 35th IEEE Photovoltaic Specialists Conference* (2010), pp. 513–517
47. R. Fitzner, E. Mena-Osteritz, A. Mishra, G. Schulz, E. Reinold, M. Weil, C. Koerner, H. Ziehlke, C. Elschner, K. Leo, M. Riede, M. Pfeiffer, C. Uhrich, P. Bäuerle, Correlation of π-conjugated oligomer structure with film morphology and organic solar cell performance. J. Am. Chem. Soc. **134**(27), 11064–11067 (2012)
48. Extraterrestrial sun spectrum AM0. http://rredc.nrel.gov/solar/spectra/am0/special.html
49. Extraterrestrial sun spectrum AM1.5. http://rredc.nrel.gov/solar/spectra/am1.5/

Index

A
Absorber, 269
Absorption, 271
Acceptor, 268
Advanced modulation format, 29, 30, 37, 38, 96, 101
Air mass, 280
All optical processing, 108–111, 115, 119, 128
All optical sampling, 76, 131, 132
Amdahl's principle, 208
Amorphous silicon, 281

B
Blend layer, 273
Building-Integrated Photovoltaics (BIPV), 285
Bulk Heterojunction (BHJ), 272

C
Cache, 211
Carrier lifetime, 118, 120–122, 126, 232, 241
Cascade solar cells, 269
Cell temperature, 282
Charge transport, 268
Chip design, 139
CIGS, 281
Climate, 215, 216, 219, 225, 255, 280
Climate change, 263
Cloud computing, 203
CMOS, 140, 141, 144–148
CMOS technology, 207
Coherence length, 271
Contention, 209
Cooling, 205
Crystallization, 273
Current matching, 277

D
Data centers, 203
Data communication, 7, 29, 75
Destructive interference, 210
Differential (Quadrature) Phase-Shift Keying (D(Q)PSK), 64–67, 96, 99
Disruption principle, 204
Distributed power delivery, 174, 176–178, 186–190, 192, 196, 198
DML, 141, 144, 147, 151, 155–158, 160, 161, 164, 165
Domino logic, 148, 151
Donor, 268
Doping, 270
Dual pump four wave mixing, 60, 63, 66
Dye, 271
Dynamic power, 207

E
Efficiency, 278
Energy consumption, 203
Energy efficiency, 1–3, 8–12, 17, 111–113, 133, 140, 141, 143, 155, 157, 159, 208, 212, 213, 216, 222
Energy efficient, 1–3, 7–9, 11, 13, 18, 22, 30, 39, 68, 115, 120, 128, 141
Energy level, 266
Excited state, 43, 68, 84
Exciton, 268
External memory, 211
Extinction coefficient, 268

F
Fermi level, 270
Four wave mixing, 37, 39, 57, 58, 109

G
Gain, 4, 12, 18, 19, 26, 32, 38–48, 51–53, 55–66, 75–78, 80–83, 88, 93–95, 207, 213
Gain saturation, 40, 57, 61, 62, 77–79
Gain spectrum, 96

© Springer International Publishing AG 2017

G. Eisenstein and D. Bimberg (eds.), *Green Photonics and Electronics*,
NanoScience and Technology, DOI 10.1007/978-3-319-67002-7

GPU, 210

H
Harvesting factor, 280
Heterojunction, 268
Heterojunction Solar Cells, 233
High-Concentration PV, 236

I
InP, 17, 19, 21, 22, 24, 117, 118, 124, 129, 239, 245
InP VCSELs, 17, 18, 22, 32
Interference, 271

L
Locality, 205
Locality principle, 213
Logic, 139–141, 144
Logic-within-memory, 212
Low voltage, 83, 141, 142, 144–146, 151, 165, 166

M
Memory intensity, 212
Microcrystalline, 247, 248
Microprocessor, 203
Mode-locked semiconductor lasers, 75
Modulation characteristics, 9, 65
Molecular orbital, 267
Mono-crystalline, 281
Moore's law, 205
Multi-core, 206
Multi junction solar cells, 237–239, 241, 243, 244
Multistate Pipeline Register (MPR), 212

N
Nano-photonic, 107
Noise figure, 41, 45, 46
Nonlinear, 39, 40, 42, 56–58, 77, 81, 84, 92, 109–112, 115–117, 119, 121, 124, 126, 129, 133, 174, 180
Non-linearity, 57
Nonlocal nonlinear response, 118
NORA, 144, 148
Np-CMOS logic, 148

O
Open-circuit voltage, 275
Optical interconnects, 2
Organic semiconductors, 267, 276, 279
Organic solar cells, 265, 266, 268–270, 278, 279, 282

P
Parallel computing, 208
Partial Rear Contact Technologies, 230
Pass transistor logic, 140, 144, 152
Phase modulation, 39, 46–48, 52, 57, 58, 78, 88, 96, 97, 99, 100
Photo-current, 271
Photon molecule, 121–123
Photonic crystal resonators, 124
Photonic terahertz signal generation, 75
Photovoltaic, 219, 220, 224, 226, 228, 236, 237, 245, 249, 250, 263–266, 278–280, 285
P-i-n stack, 274
Polarization dependence, 41, 45
Pollack's empirical rule, 206
Power conversion efficiency, 278
Power dissipation, 204
Power efficiency, 203
Power grid, 179–183, 186, 193, 194
Power management, 174, 175, 177, 178, 198

Q
Quantum dot, 109

R
Recombination contact, 276
Reflec, 271
Renewable energies, 263
Resource bottlenecks, 210
Roll-to-roll, 265
Roll-to-roll processing, 285
Roughness, 273

S
Scaling, 205
Semiconductor optical amplifier, 38, 40, 97, 109
Semiconductors, 266
Shockley-Queisser Limit, 235
Signal processing, 30, 39, 40, 56, 107, 109, 130, 133
Silicon photonics, 128, 129
Silicon photovoltaic, 228
Sky condition, 282
Small molecules, 267
Small, signal modulation bandwidth, 17, 19, 23, 28, 29, 50
Solar energy, 219, 224, 263, 284
Solution processing, 267
Speedup, 208
Stability criterion, 180, 187, 189, 192, 196, 198
Standard Test Conditions (STC), 279
Sub-cells, 276

Sun spectrum, 268
System on chip, 173–175

T
Tandem solar cell, 276
Telecommunication, 81
Thermalization, 275
Thin-film photovoltaics, 278
Thin film PV technologies, 245
Transport layer, 270
Triple junction, 277
Tunnel junction laser, 22

U
Ultrafast response, 68
Unified machine model, 210
Uniprocessor, 206, 208

V
Vacuum deposition, 266
VCSEL arrays, 29, 31
Vertical-Cavity Surface-Emitting Lasers (VCSELs), 1

W
Wavelength conversion, 39, 56–58, 60, 63–67, 111, 115, 129, 133
Wind power, 217
Wire capacitances, 208