Osamu Ueda
Stephen J. Pearton *Editors*

# Materials and Reliability Handbook for Semiconductor Optical and Electron Devices

Springer

# Materials and Reliability Handbook for Semiconductor Optical and Electron Devices

Osamu Ueda • Stephen J. Pearton
Editors

# Materials and Reliability Handbook for Semiconductor Optical and Electron Devices

Springer

*Editors*
Osamu Ueda
Graduate School of Engineering
Kanazawa Institute of Technology
Tokyo, Japan

Stephen J. Pearton
University of Florida
Gainesville, FL 32611, USA

ISBN 978-1-4614-4336-0 ISBN 978-1-4614-4337-7 (eBook)
DOI 10.1007/978-1-4614-4337-7
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012947361

© Springer Science+Business Media New York 2013
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. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
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.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Preface

This handbook covers reliability procedures and approaches for modern electronic and photonic devices, including lasers and high-speed electronics used in all aspects of our lives, from cell phones to satellites, data transmission systems, and displays. Reliability offers significant experimental and modeling challenges. Typically, parts cannot be tested until they fail under normal operating conditions. Since the target is frequently a decade or longer of useful life, this is impractical. Consequently, accelerated testing is performed. This procedure only works when the physics is well understood, and the failure mechanism is not accelerated by factors not under the control of the testing. Modeling of the failure mechanism is crucial in making extrapolated predictions of lifetime. Computer-Aided Design tools have advanced to the point where multiple factors influencing reliability can be included in testing simulations.

Lifetime prediction for compound semiconductor device operation is notoriously inaccurate due to the fragmented efforts in reliability and the absence of standard protocols. Engineers have usually relied on accelerated testing at elevated temperature with extrapolation back to room-temperature operation. This technique frequently fails for scaled, high-current-density devices. Device failure is driven by electric field or current mechanisms or low activation energy processes that are masked by other mechanisms at high temperature. Device degradation can be driven by failure in either active structures or passivation layers.

In this handbook, we have attempted to provide comprehensive coverage of reliability engineering for III-V device structures, including materials and electrical characterization, reliability testing, and electronic characterization. These latter techniques are used to develop new simulation technologies for device operation and reliability. In turn, these techniques allow accurate prediction not only of reliability but also the ability to design structures specifically for improved reliability of operation. While the characterization and accelerated testing and diagnostics isolate the factors controlling reliability, the simulation capability based on the characterization results can help to predict and better explain device reliability. Therefore, there is extensive feedback between these approaches.

The basic unit of failure in Si device technology is the FIT (failure unit), defined as 1 failure/$10^9$ device hours. For 100 devices on test, a failure rate of 1,000 FIT would mean there would be only 1 failure in 1 year. Given that a relatively small number of devices will actually show failure, it is critical to both enhance the failure rate through accelerated testing (the five common stresses used are temperature, voltage, current, humidity, and temperature cycling) and treat the resulting reliability data correctly. In any study under accelerated aging conditions, different failure mechanisms may be accelerated by different amounts for the same applied stress. Accelerated aging is useful only if we know the failure mechanism. In Si MOS devices, time-dependent failure mechanisms include surface charge accumulation or injection, dielectric breakdown, electromigration, and contact degradation and corrosion due to contamination. In compound semiconductors, we have the added issues of local regions of nonstoichiometry that affect field distributions and increase recombination, oxidation of AlGaAs or AlGaN, high densities of dislocations, and other extended defects in some structures and high surface state densities.

We also cover standard Si reliability approaches to determine the instantaneous failure rate and mean time to failure and therefore the distribution functions most relevant to the specific device technology. It is expected that the Weibull distribution is the most relevant, since the failure rate will likely vary as some power of the age of the device. In this case, the failure rate, $\lambda(t)$, is given by

$$\lambda(\mathrm{t}) = (\beta/\alpha)\mathrm{t}\beta - 1$$

where $\alpha$ and $\beta$ are constants. For $\beta > 1$,the failure rate increases with time, which is likely for compound devices, as defect migration or creation and oxidation or contact degradation occur. In the early part of reliability studies, Duane plotting is also relevant to allow a quick prediction of failure rates when the number of device failures is low. In this approach, the log of the average failure rate (fraction of failed devices at time t divided by the time) is plotted as a function of the log of time. The log-normal distribution function is a more general approach for describing the failure statistics over wide spans of time. It may also be the case that there will be two or more failure populations due to the presence of more than one failure mechanism. This can be resolved on a log-normal distribution that reveals the different time dependence of the various failure mechanisms.

This handbook focuses attention on voltage and current acceleration stress mechanisms. Many studies in Si indicate that the reaction rate of the failure mechanism is proportional to a power of the applied voltage as well as temperature, i.e.,

$$R(T, V) = \mathrm{RO}(T)V\gamma(T)$$

where the coefficient RO($T$) is an Arrhenius function of $T$ and the power dependence varies between 1 and 4.5. This determines how much acceleration occurs with increases in the bias voltage during stressing. If dielectric breakdown is the

dominant failure mode, then at a given field, a fraction of the devices will fail in a short time, with no additional failures until an increased field is applied.

Part 1 of this book covers optical devices such as light-emitting diodes (LEDs) and laser diodes (LDs) including conventional edge-emitting lasers (or edge emitters) and vertical-cavity surface-emitting lasers (VCSELs). Since the earliest stages in the history of R&D of III-V optical devices (mid-1960s), long-term reliability has been one of the key issues. This part of the book provides overall features of life testing of optical devices and results for various changes in device characteristics corresponding to a variety of device degradation modes or root causes. In order to clarify the physical phenomena occurring in the degraded regions, one must do various failure analyses on the degraded devices. In this part, detailed modern failure analyses for optical devices are systematically presented.

Concerning the degradation, three major failure (degradation) mechanisms are distinguished as follows: classical rapid degradation due to "dark-line defects (DLDs) formation" and recombination-enhanced dislocation motion, gradual degradation due to recombination-enhanced point defect reaction which determines an ultimate lifetime of the diodes, and catastrophic failure (catastrophic optical damage (COD) in case of edge emitters). Other degradation modes such as degradation due to ESD/EOS, acceleration of gradual degradation due to local lattice strain, degradation peculiar to high-power laser operation, etc., are also described. It is also worthy of note that the book focuses in detail on materials issues and degradation of InGaN/GaN LDs. The VCSEL chapter is also informative for those working on reliability studies as well as the R&D of VCSEL chips and/or components in which VCSELs are used as key devices. Fundamental physics buried in the degradation phenomena, i.e., theory on recombination-enhanced dislocation glide (REDG) and recombination-enhanced defect reactions (REDR), are systematically reviewed by professionals in this particular field.

The second part of the book focuses on electrical devices such as heterojunction bipolar transistors (HBTs) and high electron mobility transistors (HEMTs). There is tremendous current interest in determining the degradation and failure mechanisms of short gate length high electron mobility transistors in the AlGaN/GaN and AlGaAs/GaAs systems under athermal, voltage, or current-driven conditions and in new methodologies to predict and mitigate against device failure under practical operating conditions. The failure mechanisms have typically involved factors such as metal contact reaction with the semiconductor or insulation on a local scale near the high-field regions between the gate and drain, or a new mechanism called the inverse piezoelectric effect in which high biasing leads to lattice distortions and eventually formation of cracks that can then be oxidized or filled with metal by migration from the contacts. The presence of such failure mechanisms means that local probes for determining compositional and bonding changes on a local scale near the contacts are needed. Ultimately, the need is to determine the mechanisms for the observed increases in device current at threshold values of electric field, and then provide that information to a simulation code used for predicting failure.

In sum, this book is the first to cover all aspects of compound semiconductor device reliability. Research results on reliability and materials issues of both optical and electrical devices since 2000 are systematically described. Readers will find characterization techniques needed to understand failure mechanisms in compound semiconductor devices, the statistics and experimental approaches to reliability studies, and finally case studies of laser degradation and HEMT degradation among the key benefits.

Tokyo, Japan — Osamu Ueda
Gainesville, FL, USA — Steve Pearton

# Acknowledgements

The editors wish to thank Sara Kate Heukerott and David Packer of Springer US for guiding this project to completion. Much of the work at UF was supported in part by an Air Force Office of Scientific Research Multi Disciplinary University Research Initiative monitored by Jim Hwang.

Prof. Osamu Ueda, Kanazawa Institute of Technology, Japan
Prof. Steve Pearton, University of Florida, Gainesville, FL 32611, USA

# Contents

**Part I Materials Issues and Reliability of Optical Devices**

**1 Reliability Testing of Semiconductor Optical Devices** . . . . . 3
Mitsuo Fukuda

**2 Failure Analysis of Semiconductor Optical Devices** . . . . . 19
Osamu Ueda and Robert W. Herrick

**3 Failure Analysis Using Optical Evaluation Technique (OBIC) of LDs and APDs for Fiber Optical Communication** . . . . . 55
Tatsuya Takeshita

**4 Reliability and Degradation of III-V Optical Devices Focusing on Gradual Degradation** . . . . . 87
Osamu Ueda

**5 Catastrophic Optical Damage in High-Power, Broad-Area Laser Diodes** . . . . . 123
Aland K. Chin and Rick K. Bertaska

**6 Reliability and Degradation of Vertical-Cavity Surface-Emitting Lasers** . . . . . 147
Robert W. Herrick

**7 Structural Defects in GaN-Based Materials and Their Relation to GaN-Based Laser Diodes** . . . . . 207
Shigetaka Tomiya

**8 InGaN Laser Diode Degradation** . . . . . 247
Piotr Perlin and Łucja Marona

**9 Radiation-Enhanced Dislocation Glide: The Current Status of Research** . . . . . 263
Koji Maeda

**10 Mechanism of Defect Reactions in Semiconductors** . . . . . 283
Yuzo Shinozuka

**Part II Materials Issues and Reliability of Electron Devices**

**11 Reliability Studies in the Real World** . . . . . 319
William J. Roesch

**12 Strain Effects in AlGaN/GaN HEMTs** . . . . . 381
Min Chu, Andrew D. Koehler, Amit Gupta,
Srivatsan Parthasarathy, Mehmet Onur Baykan,
Scott E. Thompson, and Toshikazu Nishida

**13 Reliability Issues in AlGaN/GaN High Electron Mobility Transistors** . . . . . 431
E.A. Douglas, L. Liu, C.F. Lo, B.P. Gila, F. Ren,
and Stephen J. Pearton

**14 GaAs Device Reliability: High Electron Mobility Transistors and Heterojunction Bipolar Transistors** . . . . . 455
F. Ren, E.A. Douglas, and Stephen J. Pearton

**15 Novel Dielectrics for GaN Device Passivation and Improved Reliability** . . . . . 475
F. Ren, Stephen J. Pearton, B.P. Gila, C.R. Abernathy,
and R.C. Fitch

**16 Reliability Simulation** . . . . . 515
M.E. Law, M. Griglione, E. Patrick, N. Rowsey, and D. Horton

**17 The Analysis of Wide Band Gap Semiconductors Using Raman Spectroscopy** . . . . . 545
Sukwon Choi, Eric Heller, Donald Dorsey, and Samuel Graham

**18 Reliability Study of InP-Based HBTs Operating at High Current Density** . . . . . 583
Yoshino K. Fukai and Kenji Kurishima

**Index** . . . . . 611

# Contributors

**C.R. Abernathy** Department of Chemical Engineering, University of Florida, Gainesville, FL, USA

**Mehmet Onur Baykan** Interdisciplinary Microsystems Group, Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

**Rick K. Bertaska** New England Analytical, Nashua, NH, USA

**Aland K. Chin** Somerville Laser Technology, Somerville, MA, USA

**Sukwon Choi** Woodruff School of Mechanical Engineering, Georgia Institute of Technology, 30332-0405 Atlanta, GA, USA

**Min Chu** Interdisciplinary Microsystems Group, Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

**Donald Dorsey** Materials and Manufacturing Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, USA

**E.A. Douglas** Department of Materials Science and Engineering, University of Florida, Gainesville, FL, USA

**R.C. Fitch** Sensors Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, USA

**Yoshino K. Fukai** NTT Photonics Laboratories, NTT Corporation, Atsugi-shi, Kanagawa Pref., Japan

**Mitsuo Fukuda** Department of Electrical and Electronic Information Engineering, Toyohashi University of Technology, Toyohashi, Aichi, Japan

**B.P. Gila** Department of Materials Science and Engineering, University of Florida, Gainesville, FL, USA

**Samuel Graham** Woodruff School of Mechanical Engineering, Georgia Institute of Technology, 30332-0405 Atlanta, GA, USA

**M. Griglione** Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

**Amit Gupta** Interdisciplinary Microsystems Group, Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

**Eric Heller** Materials and Manufacturing Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, USA

**Robert W. Herrick** C8 Medisensors, San Jose, CA, USA

**D. Horton** Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

**Andrew D. Koehler** Interdisciplinary Microsystems Group, Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

**Kenji Kurishima** NTT Photonics Laboratories, NTT Corporation, Atsugi-shi, Kanagawa Pref., Japan

**M.E. Law** Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

**L. Liu** Department of Chemical Engineering, University of Florida, Gainesville, FL, USA

**C.F. Lo** Department of Chemical Engineering, University of Florida, Gainesville, FL, USA

**Koji Maeda** Department of Applied Physics, School of Engineering, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo, Japan

**Łucja Marona** Institute of High Pressure Physics, Semiconductors Laboratory, Polish Academy of Sciences, Warsaw, Poland

**Toshikazu Nishida** Interdisciplinary Microsystems Group, Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

**Srivatsan Parthasarathy** Interdisciplinary Microsystems Group, Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

**E. Patrick** Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

**Stephen J. Pearton** Department of Materials Science and Engineering, University of Florida, Gainesville, FL, USA

**Piotr Perlin** Institute of High Pressure Physics, Semiconductors Laboratory, Polish Academy of Sciences, Warsaw, Poland

TopGaN Ltd., Warsaw, Poland

**F. Ren** Department of Chemical Engineering, University of Florida, Gainesville, FL, USA

**William J. Roesch** TriQuint Semiconductor, Inc., Hillsboro, OR, USA

**N. Rowsey** Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

**Yuzo Shinozuka** Department of Materials Science and Chemistry, Faculty of Systems Engineering, Wakayama University, Wakayama, Japan

**Tatsuya Takeshita** Electrical Engineering and Information Technology, Faculty of Engineering, Kyushu Sangyo University, Matsukadai, Higashi-ku, Fukuoka, Japan

**Scott E. Thompson** Interdisciplinary Microsystems Group, Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

**Shigetaka Tomiya** Advanced Materials Laboratories, Sony Corporation, Asahi, Atsugi, Kanagawa, Japan

**Osamu Ueda** Graduate School of Engineering, Kanazawa Institute of Technology, Minato-ku, Tokyo, Japan

# Part I
# Materials Issues and Reliability of Optical Devices

# Chapter 1
# Reliability Testing of Semiconductor Optical Devices

**Mitsuo Fukuda**

**Abstract** Reliability of semiconductor optical devices used in recent systems and equipment is described from an aspect of the degradation mechanisms observed on various reliability tests. The degradation mechanisms clarified for last three decades still govern the device reliability in recent systems and equipment. The reliability of optical devices is limited with the change rate of the device characteristics in systems and equipment, and the rate is governed by the degradation mechanisms. The correlations between degradation mechanisms and the change rates of device characteristics are also reviewed and discussed in this chapter.

## 1.1 Introduction

Semiconductor optical devices have been widely used in various fields ranging from consumer electronics to communication systems as key components. Recently, these devices have diffused to sensing equipment for medical and environmental monitoring and space communication systems. These devices are now indispensable for our daily lives. In the initial application stage of semiconductor optical devices, 850-nm-band AlGaAs/GaAs lasers and LEDs began to be used in intra-office communication systems in the late 1970s. After that, various communication systems and equipment were developed and serviced. The prominent application fields of semiconductor lasers were optical fiber communication systems and consumer electronics products such as compact disks (CDs) and other optical memories. In optical fiber communication systems, semiconductor lasers mainly operate at 1,300- and 1,550-nm-wavelength bands corresponding to

M. Fukuda (✉)
Department of Electrical and Electronic Information Engineering,
Toyohashi University of Technology, 1-1, Hibarigaoka, Tempaku, Toyohashi,
Aichi 441-8580, Japan
e-mail: fukuda@ee.tut.ac.jp

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_1,
© Springer Science+Business Media New York 2013

the low-loss region of optical silica fiber. In consumer electronics, lasers operating at shorter wavelengths are used, for example, 650 nm in DVDs and 780 nm in CDs. Attempts have been made to further shorten the wavelength for this application by using GaN-based blue/green LEDs and lasers. Semiconductor LEDs and semiconductor lasers lasing at wavelengths ranging from visible to infrared are increasingly being used in sensing equipment for display, sensing, medical equipment, and IT-agriculture fields.

Photodiodes are usually used with LEDs and semiconductor lasers in communication systems and equipment and convert optical power to electric one. Consequently, the material used in photodiodes is different with the wavelength (photon energy). Si is mainly used in visible range; Ge and InGaAs(P) are main materials in near infrared range. Various device structures are developed in corresponding to the requirement of performances in systems and equipment. Semiconductor optical amplifiers and modulators are used in optical fiber communication systems, and some device structures were developed for improving their efficiency.

The most important factor in applying these devices to practical equipment or systems is reliability as well as performance. The clarification of degradation mechanisms and the improvements of the lifetimes of devices have been carried out under various reliability tests since the 1970s. The degradation mechanisms and their lifetimes closely depend on the operating and ambient conditions, for instance, the bias direction of the pn-junction. Semiconductor lasers and optical semiconductor amplifiers are devices operated under forward bias, while photodiodes and most optical semiconductor modulators operate under reverse bias. Degradation gradually occurs even in recent systems and equipment, although its magnitude has been sufficiently suppressed.

The main degradation mechanisms and reliability of semiconductor lasers, photodiodes, optical semiconductor amplifiers, and optical modulators used in optical communication systems are described, and then the failure modes are discussed in this section.

## 1.2 Devices Operating Under Forward Bias

### *1.2.1 Semiconductor Lasers, LEDs, and Optical Semiconductor Amplifiers*

Semiconductor lasers showed performance degradation, which determines their lifetimes in systems, even though they may be able to operate stably over the terms of service of the systems. Operating conditions in current communication systems are becoming more and more severe because the systems are diverging into access and data communication systems, such as 10-Gbps Ethernet and OC-192 short-reach applications, from traditional trunk systems. In these new systems, higher output power and faster operation at temperatures over 85°C are required [1].

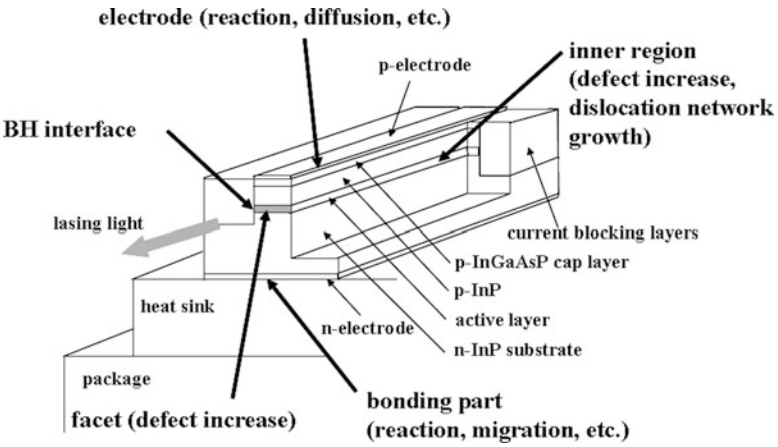



**Fig. 1.1** Schematic diagram of InGaAsP/InP buried heterostructure (*BH*) type semiconductor laser and parts degraded physically during operation

In addition, operation at a very stable lasing wavelength is required for system service terms in dense wavelength-division-multiplexing (dWDM) systems [2]. These requirements have rekindled the old degradation mechanisms or introduced new reliability problems into semiconductor lasers. The parts degraded under operation are summarized in Fig. 1.1. Semiconductor lasers used in optical fiber communication systems have mainly buried heterostructure (BH), and the degradation of BH interface governs the lifetime of lasers.

LEDs are most popular optical devices and widely applied to equipment and systems. LEDs have simple structures and performances when compared with those of lasers. Consequently, their reliability is much higher than those of semiconductor lasers.

Semiconductor optical amplifiers have similar structures to those of semiconductor lasers because the optical amplification can be generated with stimulated emission triggered with input light.

## *1.2.2 Degradation Mechanisms of Semiconductor Lasers, LEDs, and Semiconductor Optical Amplifiers*

### 1.2.2.1 Semiconductor Lasers and LEDs

A schematic diagram of an edge-emitting-type buried heterostructure (BH) type semiconductor laser used in optical fiber communication systems is shown in Fig. 1.1. The parts that degrade during operation are also indicated in the figure. The degradation mechanisms and causes are summarized in Fig. 1.2 [1]. Semiconductor lasers are still governed by these mechanisms in recent communication

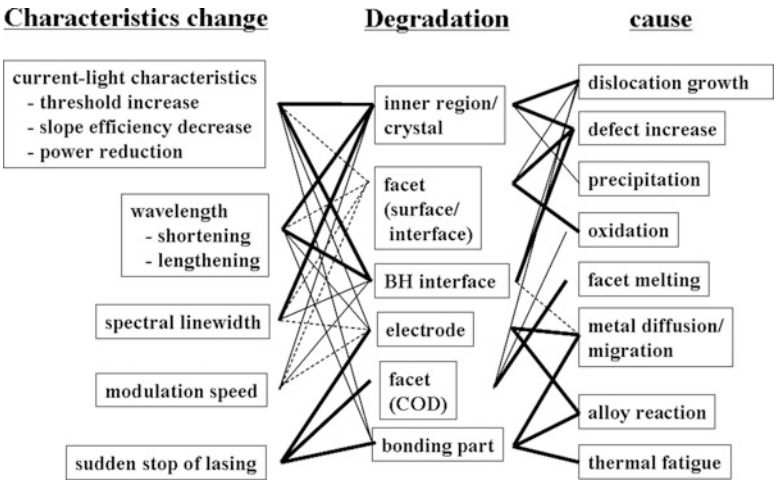



**Fig. 1.2** Main degradation mechanisms of edge-emitting-type semiconductor lasers. The *solid line* and *dashed line* indicates, respectively, strong and weak correlation

systems, in which the degradation rate is determined by the operating conditions [3–5].

The growth of dislocation, which is caused by nonradiative-recombination-enhanced defect motion or mechanical shear stress, induces rapid degradation. Facet oxidation caused by photo-enhanced chemical reaction occurs if the facet is not coated with a suitable film. Semiconductor lasers with coating films are also degraded by an increase in the defect density at the interface between the laser facet and the coating film under relatively high-power operation. Electrode degradation, which results from an alloy reaction of the electrode metal and the semiconductor, is observed in lasers having alloy-type electrodes. At the bonding part between the laser chip and heat sink/package stem, void formation and alloy reaction of the solder metal and the electrode/metallization metal of the heat sink occurs during operation if low-melting-point solder is used for bonding. In BH-type lasers, the increase in defect density at the BH interface occurs during degradation and limits the lifetime of such lasers in current systems. In addition, mechanical stress induced by the difference in thermal expansion coefficient between the laser chip and the heat sink/package stem or by an alloy reaction is an important issue in dWDM systems.

The endurance level against electric surge and electrostatic discharge (ESD) gradually decreases during operation because an increase in defect density during the degradation reduces the endurance level, although sudden failures caused by ESD and electric surges rarely occur in systems. The cause of the most common failure due to surge and ESD is catastrophic optical mirror damage (COMD) in semiconductor lasers.

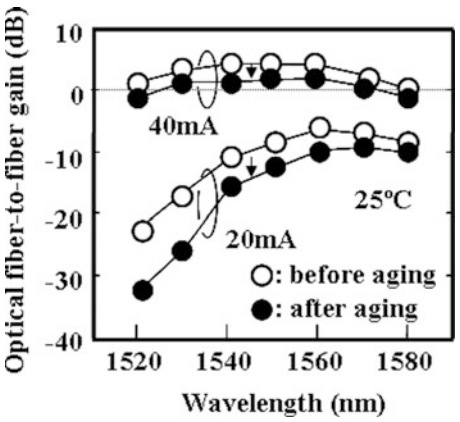



**Fig. 1.3** Change in optical gain during degradation for optical semiconductor amplifier. Aging was carried out under a constant current of 200 mA at 50°C for 2,000 h. The active layer is consisted of InGaAsP strained MQW

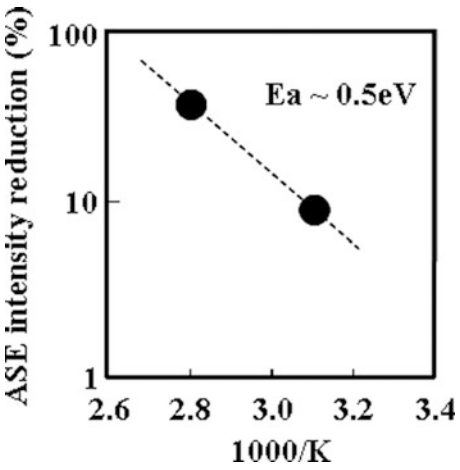



**Fig. 1.4** Change in amplified spontaneous emission (*ASE*) of optical semiconductor amplifier. Aging was carried out under constant current of 200 mA at 50°C for 2,000 h

### 1.2.2.2 Semiconductor Optical Amplifiers

Since the structure of optical semiconductor amplifiers is similar to that of BH-type semiconductor lasers, degradation around the pn-junction perimeter of the BH interface is also generated [6]. The injected-carrier lifetime is shortened, and the optical gain decreases during degradation. Additional current injection is required to keep the gain constant, as shown in Fig. 1.3. This is caused by the increase in nonradiative recombination current, and the activation energy of this degradation is about 0.5 eV, as shown in Fig. 1.4. This activation energy coincides well with that of BH-type semiconductor lasers [3]. The degradation rate depends on the square of the injected current as observed in BH-type semiconductor lasers. The magnitude of the degradation is larger than that of BH-type semiconductor laser because the injected-carrier density is fixed at the threshold value in semiconductor lasers, but the carrier-density pinning is scarcely generated in amplifier.

### *1.2.3 Change in Device Characteristics and Degradation*

Recent optical communication systems require higher performance than conventional communication systems. The physical degradation mechanisms indicated in Fig. 1.2 degrade device performance and determine the lifetime of the devices in optical systems and equipment.

#### 1.2.3.1 Semiconductor Lasers and LEDs

Current-Light Output Power Relation

This relation is the basic characteristic in all applications for semiconductor lasers and LEDs. Most degradation affects this relation and increases the operating current under constant output power operation or decreases the output power under constant current operation. Here, semiconductor lasers are used under constant output power, while LEDs are usually used in constant operating current.

For semiconductor lasers in optical fiber communication systems, the failure criteria used to be set at about 50% increase in operating current from the initial value under a constant output power. However, various other characteristics were necessary for in recent systems and equipment, and thus the failure criteria in such systems were often set with 10–20% increase in operating current from the initial values. Within this increase in operating current, changes in most of the characteristics of semiconductor lasers have been allowable in systems and equipment. LEDs are usually required to operate with stable output power.

Temperature Characteristics

Temperature characteristics are also essential ones for long-term operation in a manner similar to the current-light output power relation. The characteristics are very important for long-term reliability, especially under constant output power operation for semiconductor lasers. As optical communication systems spread to various fields such as access networks and data links, their environmental operating conditions become more and more severe. The semiconductor lasers used in trunk systems have typically been protected from the environment by employing thermal electric coolers, hard metal packaging, etc. In data communication and access systems, these types of protection tend to be omitted in order to reduce the cost of laser/transmitter modules. This introduces new but traditional reliability problems to semiconductor lasers used in recently developed systems.

To improve the temperature characteristics of semiconductor lasers, low threshold current density and high slope efficiency are especially important, and thus, the increase in laser cavity length and facet reflectivity is employed for this purpose. These result in an increase in injected-carrier lifetime. These modifications, however,

deteriorate modulation characteristics in which the reduction of injected-carrier lifetime is important for the quick change from carrier to photon. Consequently, it is difficult to improve temperature characteristics while maintaining high-speed modulation characteristics. This difficult situation is evident in recent access systems and data communication systems. In such systems, operation at 85°C or more and at a bit-rate ranging from 2.5 to 10 Gbps is required. For sophisticated semiconductor lasers, high threshold carrier density and high optical power density appear within the cavity during operation to satisfy the difficult requirements. This severe operating condition brings traditional reliability problems to the semiconductor lasers. The degradation modes observed in AlGaAs/GaAs lasers and initially developed InGaAsP/InP lasers are often generated again, although acceptable reliability is estimated for application to the systems.

Temperature characteristics of LEDs are not so critical because the long-term stability is higher than that of laser diode if the initial temperature characteristics satisfy the requirement of systems and equipment. These result from the simple structure and simple operating conditions when compared with semiconductor lasers.

### High-Speed Performance

High-speed performance is indispensable in high-bit-rate systems of more than 1 Gbps. The relaxation oscillation frequency determined by injected-carrier lifetime and photon lifetime within a laser cavity is a key factor in the high-speed performance because the upper limit of the modulation is governed by the relaxation oscillation frequency. In addition, single longitudinal mode operation is important in high-speed systems in order to separate modulated optical signals during long-distance transmission through optical fiber (suppression of wavelength dispersion in optical fiber), and distributed feedback (DFB) lasers are usually used to maintain single longitudinal mode under modulation. The relaxation oscillation frequency tends to increase as degradation proceeds because defect density increased during degradation reduces the injected-carrier lifetime, and the effect of the change in relaxation oscillation frequency is said to be negligible if the increase in operating current is within 10–20%, as shown in Fig. 1.5.

LEDs cannot operate at high bit-rate because of large parasitic capacitance and large injected-carrier lifetime. Consequently, LEDs are used at low-bit-rate systems, and thus, there is no serious problem in LEDs.

### Wavelength Spectral Characteristics

The lasing wavelength of semiconductor lasers is very sensitive to operating conditions, the magnitude of injected current, ambient temperature, packaging, etc. Fabry-Perot-type laser diodes change their lasing wavelengths due to the band-filling effect and the temperature dependence of band gap energy (shift of

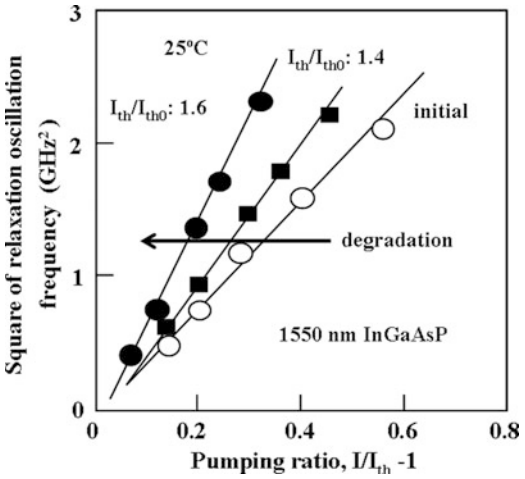



**Fig. 1.5** Change in relaxation oscillation frequency as a function of pumping ratio, $I/I_{th} - 1$, under degradation. Here, $I_{th}$ indicates lasing threshold current. The degradation is indicated with the increase in threshold current, $I_{th}/I_{th0}$ ($I_{th0}$: initial threshold current) [1]

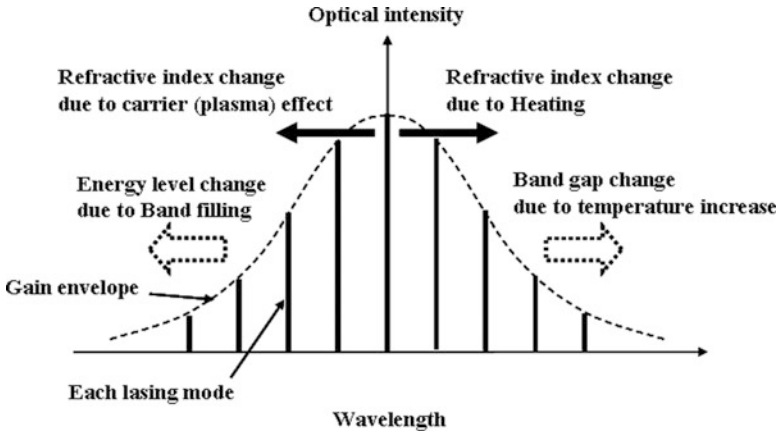



**Fig. 1.6** Schematic diagram of wavelength change mechanism in semiconductor laser under degradation [1]

gain peak) during physical degradation indicated in Figs. 1.1 and 1.2. The lasing wavelengths of DFB-type laser diodes vary with the change in effective grating pitch due to a refractive index change. This refractive index change results from the heating and the plasma effect induced by the injected carrier. These situations are summarized in Fig. 1.6. In addition, mechanical stress from the bonding part is a factor that must be considered in order to stabilize lasing wavelength precisely. During degradation, the mechanical stress originating from the bonding part

introduces a wavelength shift through the change in cavity length and refractive index. This change limits the reliability of DFB lasers in recent dWDM systems, which require precise wavelength control of less than 0.2 nm for about 20 years.

To stabilize wavelength over the long term, external modulator integrated DFB lasers are often used in WDM systems. A typical one is an electroabsorption (EA)-type DFB laser. Another stabilizing method is to use a wavelength monitor such as an etalon plate. For this purpose, wavelength monitor integrated DFB laser modules have been developed and produced. These devices and modules have demonstrated stable operation and high reliability [1]. These wavelength fluctuations might also be a problem in sensing equipment if the lasing wavelength has to be controlled precisely.

Lasing spectral linewidth, which reflects laser coherence, is often defined using the FWHM of the single longitudinal mode for DFB lasers. The linewidth is affected by 1/f-noise, shot (quantum) noise, and the side-mode-suppression ratio (SMSR), which is defined by the ratio of the main peak mode to the second mode. This spectral linewidth gradually broadens as degradation proceeds. This is due to a 1/f-noise increase, shot noise increase, and SMSR reduction during degradation [5]. These noise increases are rarely a problem in current communication systems, which are almost all digital systems. In analog systems, these noise increases are not a critical issue if the failure criterion is set at 10–20% increase in operating current. The lasing spectral linewidth does not directly affect the transmission quality in communication systems or the sensitivity in sensing equipment. The reduction of SMSR, however, deteriorates high-speed performance in transmission systems because the mixture of the second mode with the peak mode in the modulated optical signal reduces transmission quality. This SMSR reduction will also be a problem in sensing equipment because the single mode operation deteriorates, and the optical intensity fluctuation increases, although strict evaluation has not been carried out for sensing applications.

Wavelength tunability is an important characteristic for setting the wavelength precisely in dWDM systems. This wavelength tuning is performed by controlling the laser temperature or the injected current, and the tuning width tends to be narrowed during degradation. The influence of the tunability degradation is negligible in dWDM systems because the tuning range is small. The significant reduction of SMSR and the onset of FP-mode lasing are generated in severe case of degradation at the edge of the tunable range, while a decrease in the tunable range also occurs.

The wavelength stability is never required in LEDs in communication systems because LEDs are used in equipment and systems without precise wavelength requirement.

#### 1.2.3.2 Semiconductor Optical Amplifiers

The performances of semiconductor optical amplifier are similar to those of semiconductor lasers but lasing is suppressed. The physical degradation does not

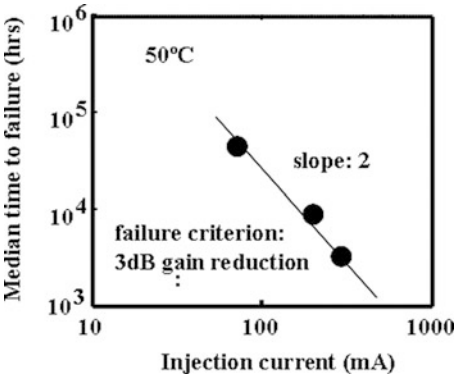



**Fig. 1.7** Median time to failure of optical semiconductor amplifier as a function of injected current. Aging was carried out under constant current (three levels) at 50°C for 2,000 h, and the failure criterion was set at time at which optical gain reduced by 3 dB from the initial value

critically influence device characteristics except for gain. The optical gain gradually decreases during degradation, as indicated in Fig. 1.3. The gain reduction in shorter wavelength range is much larger than that in longer wavelength range. This is due to band-filling effect shown in Fig. 1.6. As the injected current increases, the optical gain increases, especially in shorter wavelength range, and thus, the magnitude of the change is larger in shorter wavelength than in longer wavelength range.

One example of median life estimation is shown in Fig. 1.7 for semiconductor optical amplifier consisted of InGaAsP strained quantum well structure [6]. A feasibility of long-term operation is confirmed in the figure.

### *1.2.4 Failure Modes and Reliability*

Semiconductor lasers operate under constant output power in most communication systems, and thus the operating current, therefore, exhibits different patterns when it increases, as depicted in Fig. 1.8. The mode of current increase—sudden, rapid, or gradual—is determined by the degraded part and by the cause of the degradation (Figs. 1.1 and 1.2). These modes have been eliminated or suppressed to some extent in lasers used in traditional systems but have reappeared in semiconductor lasers operated in currently developed systems, as described above.

In addition, new problems such as precise stability of lasing wavelength have been generated as a reliability factor. Consequently, the lifetimes and failure rates have not improved much from those of the initially used semiconductor lasers if only the lifetimes are compared, and the median lifetimes required are usually a few tens of years. These estimated lifetimes and reliability, of course, satisfy the requirements from communication systems [1].

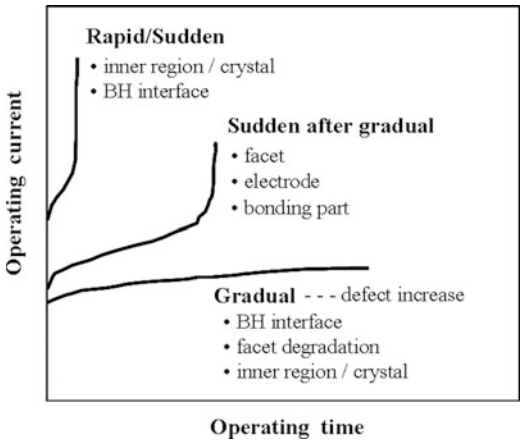



**Fig. 1.8** Failure modes in operating current under constant output power operation for BH-type semiconductor lasers. Main parts of degradation are also indicated (This material is reproduced with permission of John Wiley & Sons, Inc. [5])

For LEDs, a factor limiting reliability is estimated to be the increase in point defects in inner region because of simple structure and simple operating characteristics. The recently developed LEDs, therefore, show no large decrease in operating output power. Here, recently means sophisticated crystal growth and processing, usage of suitable electrode material and bonding solder, and so forth.

## 1.3 Devices Operating Under Reverse Bias

### *1.3.1 Photodiodes and Optical Semiconductor Modulators*

Photodiode converts optical signal to electric signal (photocurrent). Its operating modes can be divided into three categories, solar cell mode without electric bias, photodiode mode, and avalanche breakdown mode at which electric bias is set at near the breakdown voltage. Suitable structures are developed for each mode and practically used. In systems and equipment, electric bias is usually applied to photodiode. A common basic structure is shown in Fig. 1.9. Electron–hole pairs are generated at the absorbing layer when light is illuminated, and the electrons and holes are separated with high electric field generated at the pn-junction under reverse bias. The signal-to-noise ratio is determined by the ratio of photocurrent under light illumination to dark current without illumination. If dark current increases during operation, which resulted from the increase in leak current, the signal-to-noise ratio decreases. Photodiode fails when the magnitude of leak current reaches the failure criterion.

Most semiconductor optical modulators operate under reverse bias to generate high electric field at pn-junction. This is required to shift the absorption edge (or the band gap energy) of active layer of modulator. The absorption edge is shifted to longer wavelength range due to Franz-Keldysh effect in bulk semiconductor, and

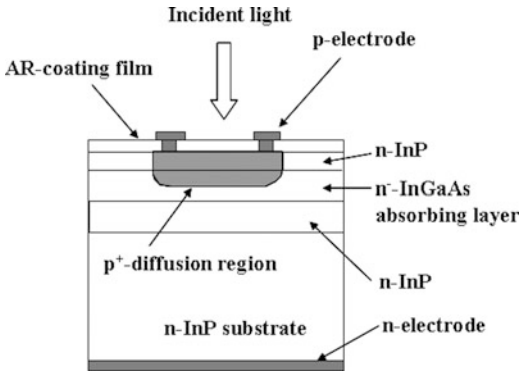



**Fig. 1.9** Cross-sectional view of InGaAs pin-photodiode

the edge shift is introduced by exciton energy reduction in quantum well structure (Quantum Confinement Stark Effect). Electric signals to be transmitted are biased in the reverse direction to the modulator, and the absorption edge is changed in corresponding to the signals. Light guided into the modulator is absorbed in corresponding to the electric signals, and the output light is modulated. In addition, Mach-Zehnder modulator composed of optical cavity and phase modulator is also used as optical intensity modulator. Here, the modulation of light phase is carried out by the change in refractive index under high electric field in semiconductor.

## *1.3.2 Degradation Mechanisms of Photodiodes and Semiconductor Optical Modulators*

### 1.3.2.1 Photodiodes

Degradation of photodiode is usually associated with the high electric field because photodiode operates with high electric field generated at pn-junction under reverse bias. The rate of degradation depends on the device structure and material used. Lifetime of photodiode in systems is determined with the magnitude of leak current (dark current) and thus governed with the weak part at pn-junction [5].

The weak parts are summarized in Fig. 1.10. Microplasma caused by current concentration on crystal defect, and so forth, is generated in inner region. The perimeter of pn-junction at the surface or interface between semiconductor and dielectric film, which is defective part at pn-junction, tends to degrade during long-term operation. This mechanism governs lifetime and reliability of recently developed photodiodes, although the rate is much small. The part at spatially inhomogeneous pn-junction, such as the bending part of pn-junction, is also a weak part because electric field is concentrated.

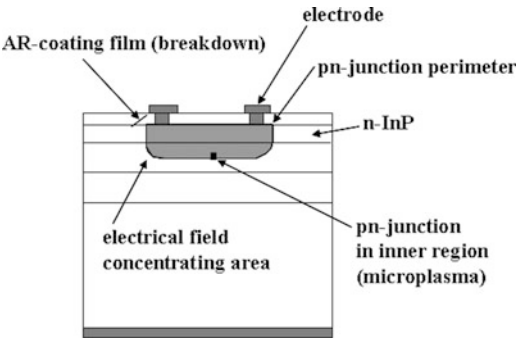



**Fig. 1.10** Parts degraded during operation in InGaAs pin-photodiode

#### 1.3.2.2 Semiconductor Optical Modulators

As previously described, the degradation is usually related to the pn-junction perimeter at the surface/interface of the devices. Optical-absorption modulators have a similar structure to waveguide-type photodiodes, and photocurrent is outputted under reverse bias. The photocurrent tends to concentrate on the part at which the electric barrier is low, such as defective parts within the depletion layer and around the pn-junction perimeter. The photocurrent at the input port of the waveguide is larger than that at the middle or output port along with the waveguide because the intensity of light is highest at the input port. The weak part is therefore around the pn-junction perimeter of the input port or its vicinity [5, 7].

The pn-junction perimeter, especially at around the input port, is also critical for ensuring the reliability of phase modulation type. This degradation tends to be proportional to the input power around the pn-junction perimeter, which is the product of the photocurrent and reverse-bias voltage. The device life is inversely proportional to the input power, as shown in Fig. 1.11.

### *1.3.3 Failure Modes and Reliability*

Degradation rates of devices operating under reverse bias are small, when compared with those of devices operating under forward bias, because the energy enhancing degradation is much smaller. Photodiodes and semiconductor optical modulators usually show stable operation. The degradation caused by the factors described in the previous section is, however, generated during long-term operation, and the physical degradation induces some failure modes through the change in device characteristics as shown in Fig. 1.12.

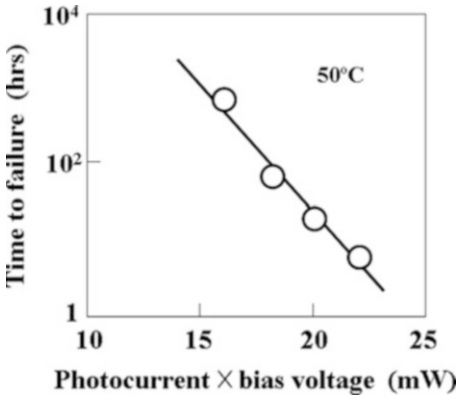



**Fig. 1.11** Median time to failure of phase-modulation-type modulator as a function of input power corresponding to the product of bias voltage and photocurrent. The aging is carried out under a constant input power of +10 dBm and a bias voltage of −4 V at 25°C. The failure is determined at the time at which leak current abruptly increases during operation (This material is reproduced with permission of John Wiley & Sons, Inc. [5])

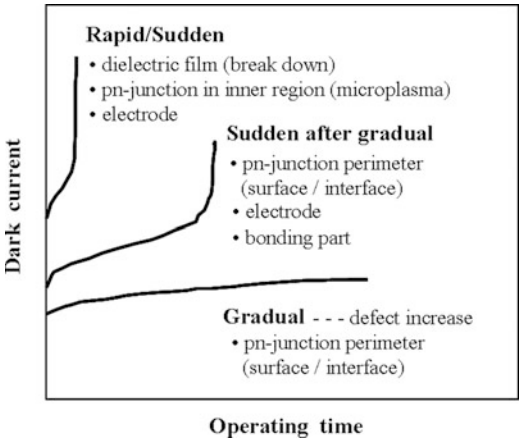



**Fig. 1.12** Failure modes in dark current under operation of pin-photodiodes. Main parts of degradation are also indicated (This material is reproduced with permission of John Wiley & Sons, Inc. [5])

## 1.4 Summary

Semiconductor optical devices have been key components in various fields, such as optical fiber communication systems and optical disk systems. High performance and reliability have been necessary for applications to those systems and equipment. Their reliability has been already examined and analyzed under various kinds

of reliability tests, and the degradation mechanisms and causes were clarified and solved to some extent. Based on those reliability technologies established, semiconductor optical devices will continue to be key components in future systems and equipment and support our daily lives.

## References

1. M. Fukuda, Microelectron. Reliab. **42**, 679 (2002) and **46**, 263 (2006)
2. ITU-T Recommendation. G.692 appendix V.**5**, 28 (1998)
3. M. Fukuda, in *Proceedings of the 18th ESREF 2007*, Paper IP6, Bordeaux, France, 8–12 Oct (2007)
4. M. Fukuda, in *Proceedings of the 2009 ISROS*, Sardinia, Italy, 11–15 May (2009), pp. 46–55
5. M. Fukuda, *Optical Semiconductor Devices* (Wiley, New York, 1999)
6. H. Mawatari, M. Fukuda, K. Magari, Y. Suzaki, N. Yoshimoto, Y. Kondo, H. Takeuchi, Y. Tohmori, in *Proceedings of the CLEO/Pacific Rim'99*, Seoul, Korea, Paper FI3 (1999), pp. 1157–1158
7. M. Yuda, M. Fukuda, H. Miyazawa, Electron. Lett. **31**, 1778 (1995)

# Chapter 2
# Failure Analysis of Semiconductor Optical Devices

**Osamu Ueda and Robert W. Herrick**

**Abstract** In both development and production of semiconductor lasers, failure analysis is crucial to quickly identifying what is responsible for problems once they have been encountered. This chapter gives guidance for how failure analysis is done in an industrial or production setting. It starts with a generalized flow chart and popular techniques for preliminary FA, as well as secondary, and finally in-depth "tertiary" techniques. The details of each of the techniques are presented in sections that give guidance on common uses for each of the techniques, strengths and weaknesses of them, and many examples of data or references with examples.

## 2.1 Introduction

In much of the rest of this book, we discuss how to evaluate whether the parts are reliable enough for your application through accelerated testing. The real problem comes when the parts fall short of your needs or fail early in the field. How do you figure out what the weakness was, and how to prevent failures from happening again?

There are approaches to solving reliability problems that involve just looking at reliability data or just looking at data from parametric measurements. Another approach involves data mining to compare whether the bad devices were all part of a particular epi run or fab run. Or the engineers can speculate on what might be wrong and try experiments to see if they can "guess" what might be wrong.

O. Ueda (✉)
Graduate School of Engineering, Kanazawa Institute of Technology, Tokyo Toranomon Campus, 1-3-4 Atago, Minato-ku, Tokyo 105-0002, Japan
e-mail: oueda@neptune.kanazawa-it.ac.jp

R.W. Herrick
C8 Medisensors, 6375 San Ignacio Ave., San Jose, CA 95119, USA
e-mail: rherrick@c8medisensors.com

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_2,
© Springer Science+Business Media New York 2013

But without good failure analysis techniques, the laser is just a black box, and it is hard or impossible to know what is going on inside. Semiconductor lasers are very simple on one level, with most having just two contacts, but determining root cause is often remarkably complex, due to the root cause often being very small and requiring nearly atomic resolution to resolve.

It has been our experience, having observed a large number of companies in this field, that most have underinvested in failure analysis capabilities. It often works out okay for a year or more, but eventually, problems normally crop up, and then it takes far longer than it needs to for the problems to be resolved, since many wrong directions are often pursued.

In this chapter, we will share many of the techniques that have proven helpful where we have worked in solving reliability problems. We will start with a general flow chart and then discuss the techniques in depth. They are ordered from simplest and easiest to use early in the chapter, all the way to the most difficult and expensive ways of confirming the true cause of failure at the end of the chapter. With time, the analyst should be able to gain confidence in drawing conclusions based on a smaller data set.

## 2.2 Failure Analysis Flow Chart

### *2.2.1 Principles for Developing a Flow*

There are four main principles for how you order the flow of your tests:

1. Perform nondestructive tests before destructive tests. Many tests described in here involve cutting the device open to create sections; doing so will then make it impossible to collect detailed spectra, electroluminescence images, etc.
2. Perform easy/inexpensive tests before expensive or time-consuming tests. These will vary from company to company, since the skills and test equipment available will vary. The principle here is to get the most data with the least resources expended. Once you have developed a "signature" for your most common failure mechanisms, you can often confirm the failure mechanism with just a few tests requiring 5–20 min per device. At a minimum, some basic characterization data is almost always desirable before proceeding on, even if you are already planning on doing extensive TEM characterization later. This should enable you to later correlate what the device looks like inside based on simple characterization techniques.
3. Perform tests which can confirm the most likely suspected failure mechanisms first. If you are seeing a lot of failures due to electrostatic discharge (ESD), you will perform a different set of tests first than if you suspect dark-line-defect failures or premature wearout. This will often change with time, as some new

problems come up, and other ones get solved. Or problems, such as field failures due to ESD, may be specific to a particular customer, so you may use a different flow when looking at parts from them.
4. Distinctiveness of fingerprints. You are usually looking for techniques that can rule out several possible failure mechanisms while confirming others. Many techniques will show a particular result for only one particular type of failure mechanism – for example, on VCSELs, a truly "leaky" reverse I-V almost always indicates EOS or ESD damage and is not observed from any other failure mechanism.

### 2.2.2 Categories and General Flow Chart

A good failure analyst should be able (in many cases) to notice subtle features in the routine characterization scans that correlate with a particular failure mechanism. In time, these should enable the analyst to categorize returns without having to go all the way to submitting the return for a full TEM analysis every time. One of the keys to being able to do this, though, is having a full data set for routine characterization before parts are sent out for the more thorough destructive FA techniques.

In Table 2.1 below, we show three categories of techniques:

1. Routine characterization techniques which can be done inexpensively internally. Most of these should be done every time as the first step in the FA process.
2. Intermediate level FA techniques. One of these should be selected frequently early on or in critical failures. These techniques generally are failure localization techniques with quarter-micron to micron resolution.
3. Deep FA techniques such as TEM or cathodoluminescence (CL). These should be done early on when building an FA library or for critical failures. They have nanometer resolution but are expensive and time consuming. It is typical when developing a new family of lasers to have to spend US $50k and go through 10–20 lasers to develop a library of TEM images for various carefully chosen failure types. In addition, if a particular cause of failure is suspected, it is often desirable to intentionally create the type of defect so that you can see if the failure signature of the known defect matches that of your failed parts from an unknown cause. For example, you might examine parts that were intentionally "zapped" with electrical overstress or have out-of-spec fabrication conditions.

In Fig. 2.1 below, we show the generalized flow chart for failure analysis. We go through each of these categories in sequence. Depending on how much supporting data we have to correlate our FA data with a true root cause, and how critical the return is, we try to close out the return with smaller amounts of required FA work.

**Table 2.1** Table of FA techniques

| Category | When to use | Technique | Labor hours | Cost (US$) |
|---|---|---|---|---|
| Routine characterization | Every time (where applicable) | Spectral scan[a] | 0.2 | $12 |
| | | L-I scan into fiber[b] | 0.2 | $12 |
| | | Optical microscopy in original header[b] | 0.2 | $12 |
| | | Optical microscopy after decapping | 0.2 | $12 |
| | | Reverse I-V[c] | 0.2 | $12 |
| | | EL Imaging[b] | 0.3 | $18 |
| Micron-scale defect localization | Important failures where better defect localization is needed | EBIC | 0.6 | $200 |
| | | Backside EL | 3 | $300 |
| | | Backside TIVA | 4 | $600 |
| | | IR microscopy | 0.3 | $120 |
| | | EMMI | 0.4 | $150 |
| Deep FA; root cause identification | First 2–5 failures of any given type; critical failures (qual failures or customer returns) | Plan view TEM | 7 | $2-4k |
| | | Cross-sectional TEM | 5 | $1.5–2k |
| | | Cathodoluminescence (where DLD structures do not exist) | 5 | $3k |

All applicable techniques in the first category should be performed; one technique for the second category is often chosen. The last category is usually only needed early in the FA process, or for new types of returns, or critical returns

[a]single mode lasers only

[b]not applicable for all lasers

[c]only where ESD suspected

### *2.2.3 Building a Failure Library*

With a new device, it is normally desirable to have perhaps 3–8 characterized failure mechanisms, including how the device reacts to overstress and ESD, a few types of manufacturing flaws, maverick defects (e.g., epi defects), mechanical defects such as cracks or stacking faults, how the device fails due to normal aging, etc. For each of the most important mechanisms, at least a few third-level FAs should have been performed. If such an "FA library" has not been developed during product development, it will need to be developed as problems crop up, something that will normally take many months and involve many incorrect hypotheses about the cause of failure.

## 2.3 Initial Inspection Techniques (First-Level FA)

Every failure analysis will start off with a routine preliminary failure analysis being performed. This collects basic data using techniques that are nondestructive (at least at the die level, if not the header level). Once some basic capabilities are established, this is something technicians would usually be trained to perform.

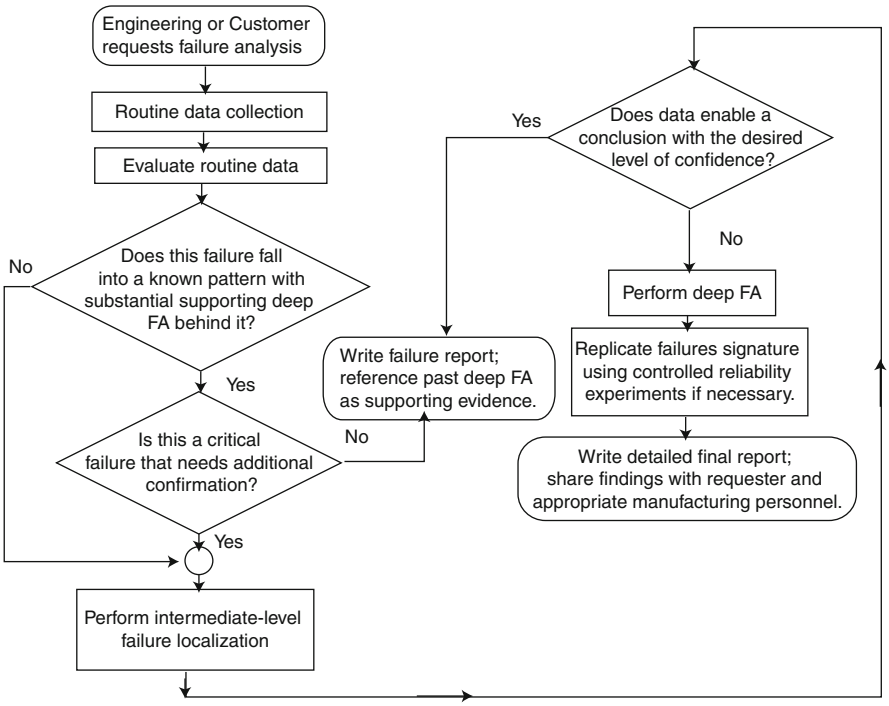



**Fig. 2.1** Generalized flow chart for failure shows how decisions are made for how to escalate to higher-cost failure analysis techniques

### *2.3.1 Visual Inspection*

Visual inspection includes general observation of the surface of the sample using a microscope. For a wafer or epitaxial layers, surface cracks, scratches, particles, roughness, and morphology are readily observable by an optical microscope. Thermal damage from electrical overstress (EOS) is also visible. In the failure analysis of optical devices, this is the first step in a series of evaluation techniques. It is possible to see several failure modes such as catastrophic optical damage (COD) at the mirror surface, electrode degradation due to alloy reaction between metals and semiconductors, and solder migration. Where the devices are in sealed headers with clear windows or low-magnification lenses, it is considered good practice to take a die image of the device before decapping; occasionally, the depackaging process itself causes mechanical damage, and in such cases, referring back to the original photo can determine whether the damage was present in the device as originally received.

One enhancement is to inspect lasers with a Nomarski microscope. This involves interference between beams and is capable of detecting steps tens of nanometers in height that would go unnoticed in bright-field microscopy. It makes it easier to find cracks or epi defects in the failed devices.

### 2.3.2 Device Characterization

Verifying failure is an important next step. For lasers inside a product, for example, for aligned lasers in a fiber optic transceiver, the eye pattern would be captured, and fiber coupled power and spectrum would be measured. For other systems, optical tests would be applied specific to the product. If those point to abnormal operation, the laser would then be removed and characterized. An additional step is often to solder a replacement laser in place, and test the system to see if that is sufficient to restore normal operation.

For lasers that are less embedded in a system, "L-I-V characterization" would be the first step; this is also done after depackaging for lasers that fit into higher level assemblies. L-I-V characterization involves a sweep of current over the full range of operation while monitoring voltage drop and light output. Lower than normal light output would be the common result if the laser is indeed degraded. But how much degradation has occurred is often a clue to what the failure mechanism is. While system failure often occurs at similar device degradation levels, very rapid failure mechanisms often result in severe degradation during the debugging process and will show much more severe signs of degradation than a laser that has been gradually degrading over a long period of time.

In some cases, it will be useful to follow this up with log L-I and log I-V testing. While this is not generally used in production test sets, data is collected using an expensive parameter analyzer, with either logarithmic steps in current or, more commonly, steps in voltage (typically 0.05 V step size). Comparing the light output and voltage below laser turn-on can often help to distinguish between multiple failure mechanisms [1]. Reverse Log I-V is also collected, especially if damage due to electrostatic discharge is suspected. In this case, a good laser will usually have a sharp reverse avalanche breakdown characteristic, while a laser that has suffered thermal meltdown due to damage will have much higher leakage before reaching avalanche breakdown. Also, many types of dislocations, or traps such as copper, will also contribute to reverse leakage, although these are much more subtle than ESD damage and usually only contribute nanoamps of leakage. Note that some types of lasers, such as high-power lasers, often have high reverse leakage even for good devices, preventing this from being a useful screen.

### 2.3.3 Electroluminescence

For many types of lasers, the most useful method of characterization per hour of work put in is electroluminescence. It tends to be the most useful in lasers that are mounted epi-side-up (this usually corresponds to p-doped-side-up) and do not have continuous metal coverage. Lasers with blanket metal over the epi side, or that are mounted epi-side-down, would either require backside electroluminescence or might be better suited to a technique like top-view EBIC. For all edge-emitter

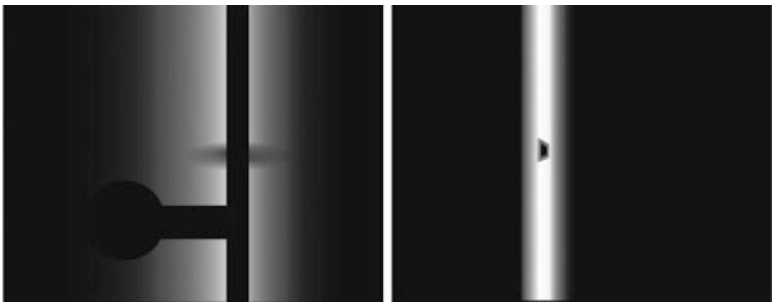


**Fig. 2.2** The topside EL (*left*) allows you to see the general area that the DLD occurred, without seeing much detail; the backside EL (*right*) allows approximately 1 μm resolution, provided degradation is not profound, but requires additional effort in sample preparation

lasers, the analyst will probably want to take facet images with a small applied bias to see if the front and back facets show any dark spots or irregularities, before proceeding on to topside or backside EL.

#### 2.3.3.1 Topside Electroluminescence Imaging

Topside electroluminescence imaging (abbreviated "EL") is usually a very simple process with minimal equipment requirements. All that is used by many firms is a light microscope with objectives that go up to 50–100× magnification, an inexpensive Si-CCD camera (at least for lasers with $\lambda < 1$ μm), and some method of capturing the images. You also need an inexpensive current or voltage source to control the laser bias. You typically would take pictures at currents below lasing and look for dark areas; the dark areas are usually the areas where defects are present. Taking pictures above threshold often confuses matters, since modal structures will have filamentation, and separation between filaments is not easily distinguished from dark-line-defect structures. For top-emitting devices like VCSELs and LEDs, EL imaging is often sufficient to conclusively identify the failure mechanism, although usually not the root cause or point of origin. Please see the VCSEL chapter for additional discussion and examples. For in-plane lasers (i.e., edge-emitters), topside imaging will usually only be sufficient to categorize where along the laser the degradation has occurred: front facet, back facet, or mid-cavity. Once that information has been obtained, one can either submit the lasers directly for TEM if desired (provided the dark area is not too large) or move on to the sample preparation required for backside electroluminescence. Note that special IR-optimized objectives, which are generally twice the price, have the capability of resolving EL features about half the size of visible objectives and are thus a worthwhile upgrade in most cases. Examples of what topside and backside EL might look like are shown in Fig. 2.2.

Note that another class of lasers can give "top-emitting electroluminescence" from the package perspective: devices that are mounted p-down but have only

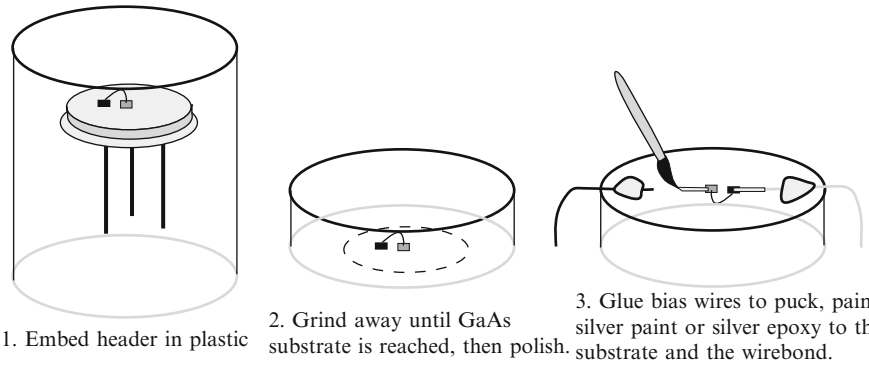



**Fig. 2.3** Sample preparation for backside electroluminescence, or backside laser-scanning microscopy examination

wirebonds on the top of the device or easily removable foil on the top. In this case, wet chemical etching can remove the gold in the n-metal, and ion milling can remove the remaining metal beneath, leaving only clear uncovered substrate, without any of the sample preparation normally required in the bottom-emitting EL section that follows.

#### 2.3.3.2 Bottom-Emitting Electroluminescence

For many edge-emitting lasers, bottom-emitting (or backside) electroluminescence is a favored technique for defect localization and characterization. This technique works best on lasers where the substrate bandgap is larger than the lasing photon energy, since in such cases substrate absorption is minimal. For example, GaAs-based lasers with $\lambda > 0.9$ μm or InP-based lasers with $\lambda > 1$ μm both would work well. However, even in cases where this is not true (e.g., 808-nm lasers on GaAs substrates), some backside EL can often still be observed. In such cases, the laser optically pumps the substrate, and substrate emission in the 870–890-nm range is observed, albeit with lower spatial resolution and requiring higher drive bias than would be needed with a transparent substrate. The laser is normally embedded in a clear epoxy puck, with a top contact accessible after polishing. Then, the header of the laser is mounted on, and the bottom contact is ground away using standard sample preparation techniques, and polished. Probes or contact wires then are used to power the device, and images are taken using the same equipment described earlier for topside EL. The sequence for sample preparation is shown in Fig. 2.3.

Note that the step putting wires on is usually skipped if the system being used for analysis allows the use of microprobes. Also, establishing contact to the bare substrate can be challenging for some sample types and may require etching with a dilute acid or base to remove surface oxides or using a low-melting point alloy (such as an In-Ga mixture) to form an intermetallic compound with the compound

semiconductor. Another popular method is to mask the stripe areas with tape or photoresist, then place the sample in an SEM metal coater, and deposit a thin conductive layer to make n-side contact.

Where the substrate is nontransparent, two options are present. First, other techniques may be better suited to defect imaging, most prominently plan-view EBIC or laser-scanning microscopy (LSM) such as TIVA. The second option would be the indirect emission mentioned earlier that requires no special sample preparation. The third option would be selective substrate removal using spray etching; for example, GaAs etches with high peroxide concentration can be used that stops when an AlGaAs layer is reached [2].

## 2.4 Micron-Level Localization Techniques (Second-Level FA)

Once the first-level FA has been performed, the analyst should examine the data. If the data show a well-known failure signature, and the failure mechanism is understood, the analyst may be able to write the report based solely on the first-level FA data and leverage previous second- and third-level data to support his case. However, before similar failures are well understood, and for most "critical" returns, second-level FA data are usually collected. On occasion, it may be desirable to go straight from first-level to third-level techniques, skipping second-level techniques.

The appropriate second-level FA technique to choose depends mostly on die geometry and defect characteristics – a preliminary defect type identification should have been made based on the first-level data. One common feature among the main techniques listed here is that they are mostly plan view: that is, they provide an image of the damaged or defective areas as they would be seen looking down on the die from the top. This is because, except for a few failure mechanisms like facet damage that can be seen easily, it is generally necessary to identify what area of the laser is degraded before proceeding further. The second-level FA techniques generally take a few hours or cost a few hundred US dollars in lab fees.

### *2.4.1 Electron-Beam-Induced Current (EBIC)*

This method enables us to evaluate electrical properties of semiconductors and defects in semiconductors by detecting current or voltage generated by the irradiation of an electron beam in a scanning electron microscope (SEM), that is, an electron-beam-induced current (EBIC) or an electron-beam-induced voltage (EBIV).

There are two methods for evaluation:

1. Set the electron beam parallel to the p-n junction (cross-sectional EBIC, as shown in Fig. 2.4a).

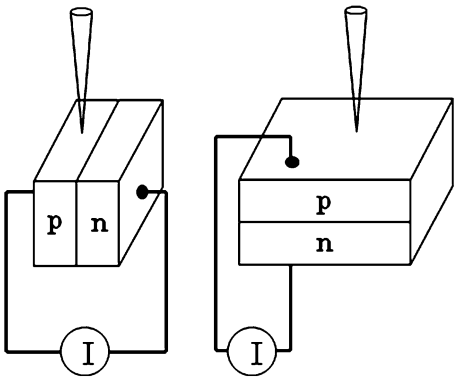



**Fig. 2.4** Schematic diagrams for two types of specimen/electron-beam geometries: cross-sectional EBIC on the *left* and *plan*-view EBIC on the *right*

2. Set the electron beam normal to the p-n junction (plan-view EBIC, as shown in Fig. 2.4b).

In cross-sectional EBIC, the electron beam irradiates a cleaved surface of the sample. By this method, the diffusion length of holes and electrons in the p- and n-type layers are evaluated. The current generated when an electron beam is irradiating on the surface of the sample is given by

$$I = \exp\left(\frac{L_{\mathrm{p}}(L_{\mathrm{n}})}{x}\right) \tag{2.1}$$

where $x$, $L_{\mathrm{p}}$, and $L_{\mathrm{n}}$ are the distance from the p-n junction and the diffusion length of holes and electrons, respectively. Thus, $L_{\mathrm{p}}$ and $L_{\mathrm{n}}$ are obtained by the slope of the 1n[$I$] vs. $x$ plot. The width of the depletion layer at the p-n junction is also characterized by the line profile of $I$. When the electron beam is scanned on the mirror surface of a semiconductor laser, we can evaluate localized defects or the failure mode of the junction or heterointerface by using an image profile based on the current $I$ (i.e., an EBIC image).

On the other hand, with plan-view EBIC, one can evaluate defects in the crystal. For example, dislocations are observed as dark-line defects (DLDs) since minority carriers generated by the penetrating electrons recombine nonradiatively at dislocations [3]. Other defects like stacking faults and precipitates are also observed as dark defects by a similar mechanism [4]. In this method, the resolution power depends on the diffusion length of the minority carriers. One minor limitation of this technique is that the defects must be in the depletion region, which usually means near the active region. So defects with the cladding or contact layer may not be visible with EBIC.

EBIC is available at some service labs. Where the device geometry makes EBIC the favored characterization technique, it should be added to the in-house scanning electron microscope. This is usually not particularly difficult on microscopes where the entire viewing chamber opens and only requires the purchase of a feedthrough

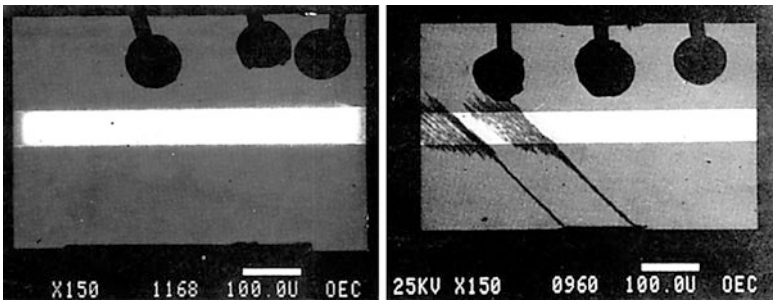



**Fig. 2.5** Plan-view EBIC images of aged stripe lasers (Reprinted from Ref. [5], copyright 1991. With permission from Elsevier)

port for the electrical connections and a current amplifier or picoammeter. For SEMs with load locks, commercial solutions allow the electrical contacts to be fed out.

#### 2.4.1.1 Plan-View EBIC

In edge-emitters with blanket metal on both surfaces and little topography (e.g., no etched ridges), plan-view EBIC will often be the best technique for imaging device degradation. Where the device is mounted epi-up, this can be almost as easy as many first-level FA techniques. An example of a plan-view EBIC image is shown in Fig. 2.5.

The chip dimension is 400 × 600 μm, and the lasing stripe is 60 × 600 μm. The stripe is brighter than the rest of the die because the stripe lacks 0.12 μm of insulating dielectric that reduces the electron penetrating power and prevents simple vertical carrier extraction. The left image shows a normally degraded die, which has degraded due to microloops near the facets. The right die degraded prematurely due to cleave defects on the bottom edge; climb dislocations originated there and killed the device when they grew into the lasing stripe.

#### 2.4.1.2 Cross-Sectional EBIC

Electrical properties of degraded laser diodes can be characterized by EBIC even after the diodes have stopped lasing if the diode still displays some *V-I* characteristics. The EBIC measurement in this case is carried out under the irradiation of an electron beam to the cleaved mirror surface of the degraded laser diode, where the direction of the electron beam is parallel to the junction (see Sect. 6.6.2). From the EBIC image or line-scan EBIC signal, one can characterize local destruction of p-n junctions or heterojunctions. This method is very useful for detecting a leakage current path in a laser with a complicated junction

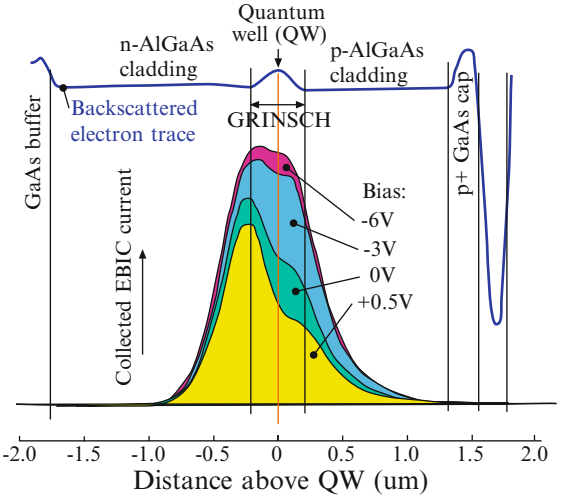



**Fig. 2.6** Line scans of collected current a various biases, overlaid on the backscattered electron intensity scan. It shows that the effective p-doping is far lighter than the n-doping, with the depletion region primarily moving into the p-cladding as the reverse bias is increased

structure (e.g., a BH laser). The method is convenient since the degraded diode that is mounted on a package can be used without any additional preparation. One must only connect the pins of the stem with the socket of the EBIC holder in the SEM stage and align the mirror surface to be normal to the electron beam by tilting the stage.

If you wish to characterize doping near the active region, one of the best available techniques is cross-sectional EBIC with a wide range of applied biases, as shown below in Fig. 2.6.

Note that two types of current amplifiers or picoammeters are available. One type is relatively simple and has a fairly limited ability to stay in range with offset voltages. This type should be adequate for plan-view EBIC, or simple cross-sectional EBIC, but can only compensate for applied bias ranges of at most −1 to +0.1 V. In simple imaging, no applied bias is used, so this is not an important limitation. The second type of amplifier is more versatile and has the ability to offset applied biases on the laser that range from −5 to +0.5 V. This is useful for cross-sectional EBIC in being able to collect data of the sort shown in Fig. 2.6 above. It is difficult to tell what the capabilities of particular amplifier are without trying it; it is best to try a demo of the proposed amplifier type, to verify it has the range to provide an image across the full range of proposed bias voltages.

### *2.4.2 Laser-Scanning Microscopy (LSM; TIVA)*

Laser-scanning microscopy (LSM) has become much more popular as a failure localization technique for silicon ICs over the past decade. In addition, its use for certain types of semiconductor lasers, such as VCSELs, has become somewhat routine. Most commonly, the “thermally induced voltage alteration” or TIVA mode

is used. In this mode, a laser with wavelength below the bandgap is used; 1.06- and 1.34-μm lasers are standard on many commercially available LSM systems. A fixed current source is used, usually with a small reverse or forward current, and the voltage required to obtain that current is recorded as the laser is rastered across the sample. In the vicinity of defects, the voltage required to get the leakage normally drops, and this provides a 2-D map of the defects. A trio of three images is saved for each area examined: the TIVA image, a reflected light image (which shows the metallization and the surface features at the wavelength where the device is transparent), and a combined colorized image with the TIVA defect map overlaid on top of the reflected light image. Please see Sect. 6.7.6 in the VCSEL chapter for sample TIVA images.

The sample preparation is similar to that for backside EL: The sample is embedded in a puck, and the header and back metal is lapped off and then polished. As has been demonstrated for silicon samples, the samples need to be thin to get good resolution; we have obtained 1-μm resolution with 120-μm-thick samples but were able to obtain 0.2-μm resolution when additional thinning was done to take the parts down to only 25-μm thickness.

TIVA does not work for every type of defect – for example, for wearout failures, we studied in VCSELs; the TIVA signature was featureless and looked the same as an undegraded device. Its strengths are that it can provide a defect area map for the entire device (a chip of many millimeters in size can be imaged) and does so with reasonable labor cost. We have used it to trace defect networks back to a wide variety of root causes, including scriber damage at the die edges, cracking from improper pick-and-place, epi defects, and internal device cracking. Typical service lab charges are US $400 per sample for imaging, plus the cost of sample preparation.

### *2.4.3 Mid-IR Microscopy*

Where cracks or mechanical damage can be seen on the surface or are suspected subsurface, it is worthwhile to take microscope pictures with long-wavelength filtering to allow one to see through the device as if looking through a cube of glass with a fracture in it. Of course, the key is to only use wavelengths longer than the absorption cutoff of the layers of interest, as well as ideally the substrate, typically 0.9 μm for GaAs and 1.0 μm for InP. For most Si-CCD cameras, the sensitivity is very marginal with the filtered light, for all but the strongest illuminators. The exception would be cooled CCD cameras that allow longer integration times. And of course, other cameras designed to image longer wavelengths (e.g., those with InGaAs CCDs) can easily do the job.

Besides studying mechanical damage and cracks, another excellent example of where this makes sense is stacking faults. A common cause of stacking faults is concentrated pressure on the device. A piece of grit or hard contamination sticks onto a pickup nozzle, and when the device is mounted onto a header, the force

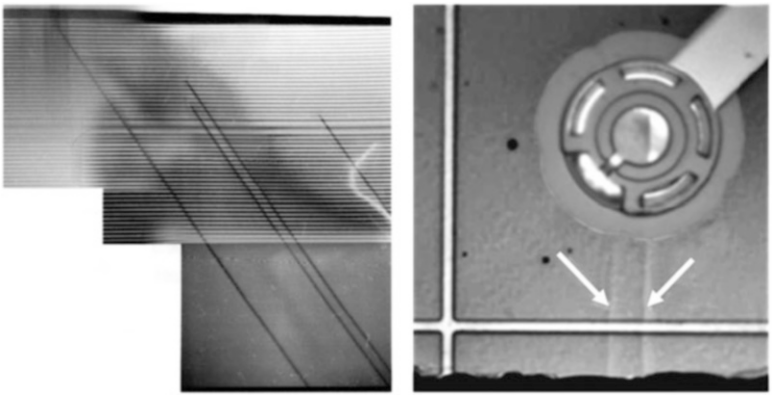


**Fig. 2.7** Cross-sectional TEM image of stacking faults in a failed AlGaAs VCSEL (*left*) and corresponding image obtained by IR-filtered image (*right*) (Reprinted from Ref. [6], copyright 2008. With permission from SPIE)

is concentrated into a very small area, resulting in pressures beyond the yield strength of the brittle crystalline materials used to make semiconductor lasers.

A typical example of evaluation of stacking faults in a failed AlGaAs VCSEL is shown in Fig. 2.7 [6]. Several stacking faults lying on the (111) plane are seen in the cross-sectional TEM image, while the stacking faults are observed as band-shaped regions as shown in the right figure of Fig. 2.7 (see arrowed region). This contrast is assumed to be originated from the interference of backscattered light at the bottom of the chip and that at the stacking fault.

### *2.4.4 Photoemission Microscopy (EMMI)*

Another common failure analysis technique used in silicon integrated circuits (ICs) is reverse-biased photoemission, or emission microscopy (EMMI). Traditionally, this involved applying a reverse bias to the IC and observing it with a microscope inside a light-shielded box, using a camera with a channel multiplier plate (similar to a "starlight scope" used to see in very low-light situations). More recently, a long-integration time with a cooled CCD might be used instead to capture the faint image [7]. Most well-equipped analytical labs offer this capability. However, this technique is not widely used with semiconductor lasers as of this writing.

It has traditionally been used in LED for imaging damaged areas [8]. It has also been used for imaging dislocation arrays [9]. We would not be surprised to see this technique become more popular and more commonly used over the next decade, as cooled CCD cameras become available inside manufacturers labs, and they have the opportunity to turn this from a second-level FA technique into a first-level capability that could be used on a routine basis.

### 2.4.5 Secondary Ion Mass Spectroscopy (SIMS)

Semiconductor lasers are normally grown using metal organic gasses, in a process referred to as MOCVD or OMVPE. One of the main issues that can occur is contamination of the epi, through undesired impurities in the source gasses or reactor vacuum leaks. This usually results in reduced wearout lifetime of the lasers and sometimes in reduced performance as well. This can be caught with regular highly accelerated reliability testing, but when a problem comes up, there are usually a broad range of potential explanations for it, and techniques are needed to narrow down the scope of the problem. For reactor problems of the sort described here, secondary ion mass spectroscopy (SIMS) is one of the best techniques for detecting contaminants, as it is able to detect part per million (ppm) or part per billion (ppb) levels of undesired impurities. We have used SIMS many times to solve line-down production problems, which turned out to be traced to causes like contaminated source gasses, a bad hydrogen or arsine purifier, etc.

SIMS is ordinarily done on unprocessed wafers sampled from the same growth runs as other parts of interest and typically costs a few hundred dollars per wafer analyzed. However, it is possible with a great deal of additional work to do SIMS on processed die. This involves very laborious backside polishing of the substrate off of the bottom of the die: something that takes many hours, a great deal of experience, and needs to be nearly perfectly planar with the growth plane to be effective. At the time of writing, this service costs $5k per die, with nearly all of the cost being for sample preparation.

### 2.4.6 Scanning Electron Microscopy (SEM)

Another easy and popular tool for examining devices is the scanning electron microscope (SEM). While this offers much better depth of field and resolution than the optical microscope, unfortunately, the defects in most lasers are buried far from the surface, and the SEM only shows the surface contours of the device – information that rarely is able to reveal the cause of failure. However, in other sections, we discuss methods where accessories can be added to the SEM to provide information below the surface, from techniques like EBIC or CL.

#### 2.4.6.1 Direct Use of SEM

One example where the SEM might be useful would be in examining cross sections made on semiconductor lasers, as we discuss in the next section. A second example would be examining devices to look at surface contamination or thermal damage. While such features can normally be seen in optical microscopy, it may be useful to see more details at higher resolution. A third example would be to examine a cross

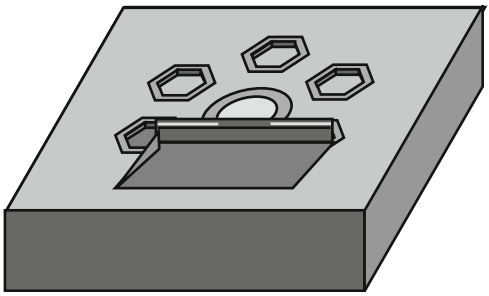


**Fig. 2.8** FIB cutting can prepare a cross section to allow examination of a profile by SEM

section for critical features, suspected cracking, or other issues. For some lasers like VCSELs, this works best when combined with focused ion beam (FIB) cutting to prepare the cross section, as shown below in Fig. 2.8.

Other enhancements can help in using SEM. Examining cross sections with low voltage will show contrast with doping – for example, we can see the proton implant profiles in cross-sectioned VCSELs when imaging voltages of 5 kV or lower are used. High voltages are often considered preferable for high resolution or for limiting effects of surface charging. Another enhancement is the addition of a backscattered electron detector (BSE). BSE gives image contrast based on the atomic composition of the substance being imaged and can thus allow the user to map composition in a flat cross-sectional image.

#### 2.4.6.2 Stain-Etching and SEM

In some type of lasers (e.g., buried heterostructure lasers), the cross-sectional shape of the active buried layer is very important for obtaining high reliability of the diodes. Therefore, it is often required to measure the thickness, width, and geometrical shape of the active layer in cross section. This is carried out as follows (see Fig. 2.9):

1. Stain-etching of the mirror surface using preferential etchants
2. Observation of the etched active region in an SEM

By this process, diodes with an abnormal shape of the active region can be identified.

This method is used to easily reveal the cross-sectional structure of ICs, semiconductor lasers, and MESFETs consisting of heterojunction, p-n junction, and diffused region. In this method, preferential etchants are commonly used. After etching, a step is formed at the interface. For a heterointerface, one can select an etchant with high selectivity as shown in Fig. 2.9b. However, for evaluation of p-n junction or homojunction (i.e., $n^+$ to n junction), special etchant and/or conditions are required. For observation of the cross section, an optical microscope or SEM is used.

A broad range of etches can be used for stain-etching. While "A-B" etches are popular [11], we have had good luck with simple etches like 1:1:10 $H_2SO_4$:$H_2O_2$:

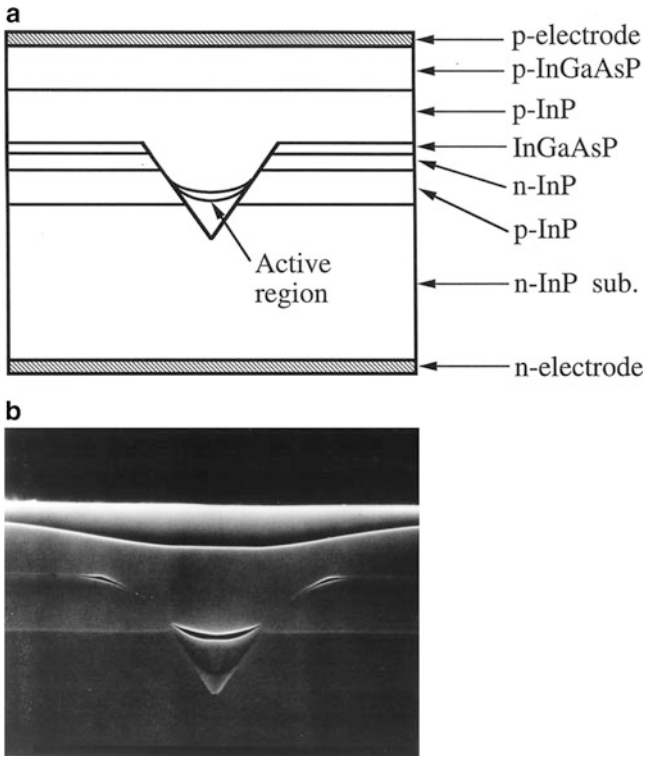



**Fig. 2.9** Result of stain-etching of a VSB laser (**a**) a schematic diagram of VSB laser [10]; (**b**) a cross-sectional SEM image of VSB laser after stain-etching with $HNO_3$/HF solution

$H_2O$. Another popular enhancement is to illuminate the semiconductor while it is in the etch beaker using a microscope illuminator or other intense light source; this generates free carriers at the p-n junction, enhancing the etch rate there [12]. In the early 1980s, the author found that in this method, difference in the contrast between p-InP and n-InP is drastically increased by depositing (evaporation or plating) Au or Au-based alloy on the top surface of the sample [13]. That is, as schematically shown in Fig. 2.10, n-InP regions are observed as brighter ones whereas p-InP regions darker.

The amount of Au from the n-InP region by SEM/EDX measurement is two times higher than that from the p-InP region. It is assumed that during the photo-etching, the surface of the n-InP is negatively charged by the accumulated electrons which are generated by the optical illumination, while the surface of the p-InP is positively charged by the accumulated holes. Therefore, positively charged $Au^+$ or $Au^{3+}$-ions which are dissolved from the Au-film into the stain-etching solution tend to deposit selectively on the surface of n-InP. Based on this assumption, it is concluded that the brighter contrast of n-InP can be explained by either of the following reasons:

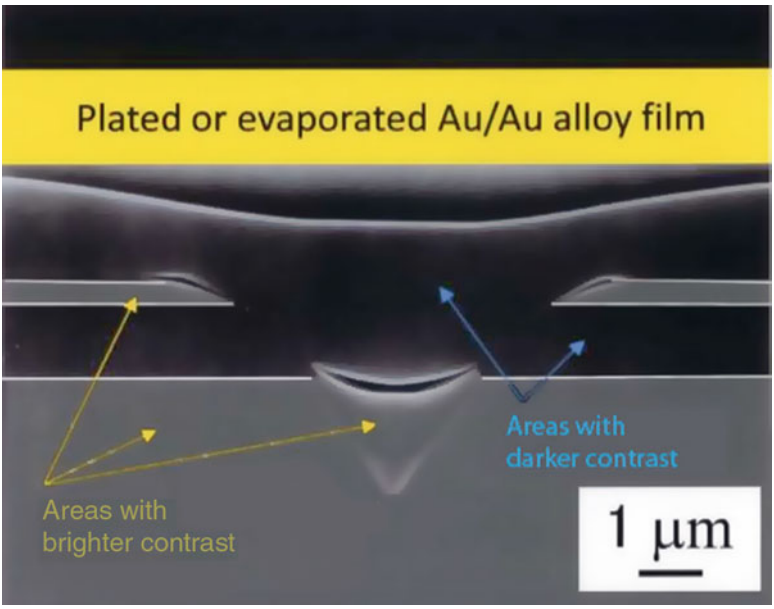



**Fig. 2.10** Schematic diagram of cross-sectional SEM image of stain-etched p-InP/n-InP homojunction in a VSB laser. The etching is carried out with 1 g $K_3Fe(CN)_6$, 1 g KOH in 10 ml $H_2O$ under optical illumination

1. Higher amount of deposited Au on the n-InP generates more secondary electrons than on the p-InP.
2. Due to higher amount of deposited Au, etching rate of n-InP become lower than that of p-InP.

Another unique method by A-B etches is to reveal the shape of stacking faults as shown in Fig. 2.11 [14]. Figure 2.11a shows a typical A-B-etched surface of two AlGaAs epitaxial layer wafer which contains many stacking faults.

Three types of stacking faults were revealed. They are schematically shown in Fig. 2.11b. The A–B etches attacks the two partial dislocations of the stacking fault which reproduces the two-dimensional feature of the stacking fault. From these results, it is concluded that one of the partial dislocation or both change their directions at the interface between epi-layers.

### *2.4.7 Energy-Dispersive X-Ray Spectroscopy (EDX or EDS)*

For identifying contaminant particulates, or studying issues with the packaging, it is often useful to determine the atomic makeup of the contaminants; this is most often done with energy-dispersive X-ray spectroscopy, also know by the acronyms EDX or EDS.

At the start of the analysis, a few points would usually be selected – most often one in the defective region and one in the same layer outside the defective region to

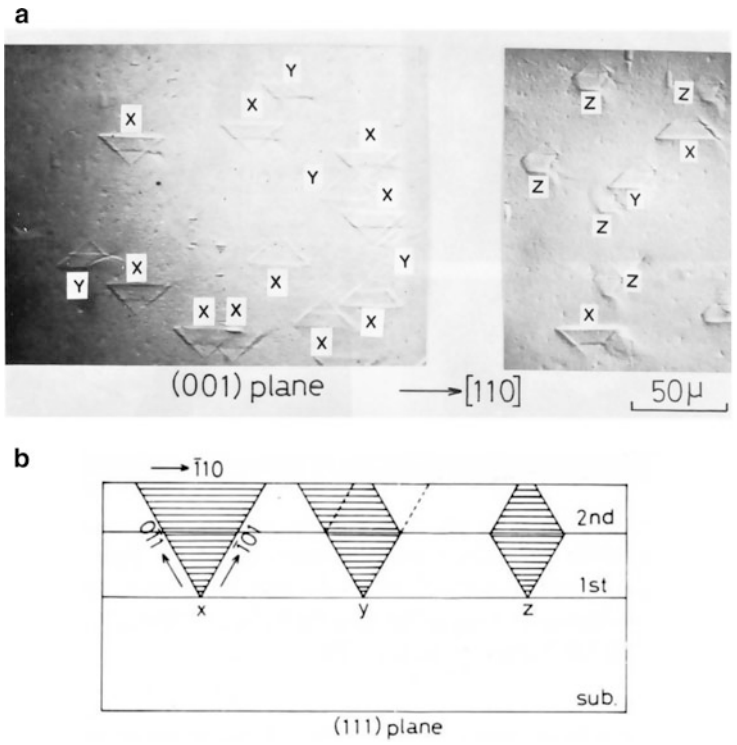



**Fig. 2.11** Characterization of stacking faults originating from the epi/sub interface in an AlGaAs/GaAs multilayered structure by A-B etches. (**a**) An optical micrograph of A-B-etched surface; (**b**) schematic diagram of three types of stacking faults revealed by A-B etches. The mole fractions of the first and the second layer are 0.3 and 0.03, respectively (Reprinted from Ref. [14], copyright 1977. With permission from Elsevier)

allow a comparison. Then, based on what elements are found in the spectra, a line scan or map of that element can be performed, as listed below.

There are three types of analyses:

1. Point analysis: A desired region is analyzed by fixing the incident electron probe on the region. An X-ray spectrum is obtained. Note that this is also beam-energy dependent; at low energies, penetration is relatively low; at times, you may want to look beneath the surface, and higher energies are used for this.
2. Line analysis: A desired region is analyzed by scanning the electron beam in a line. A particular X-ray peak is selected; this would normally correspond to one of the peaks of an element of interest. A line profile of the desired X-ray intensity is obtained by selective detection using an X-ray filter (i.e., "X-ray energy window"); this effectively gives a plot of the concentration of the element of interest versus lateral position in the scan.
3. Two-dimensional analysis (mapping): This technique enables us to visualize the two-dimensional distribution of a desired element with comparison of the SEM

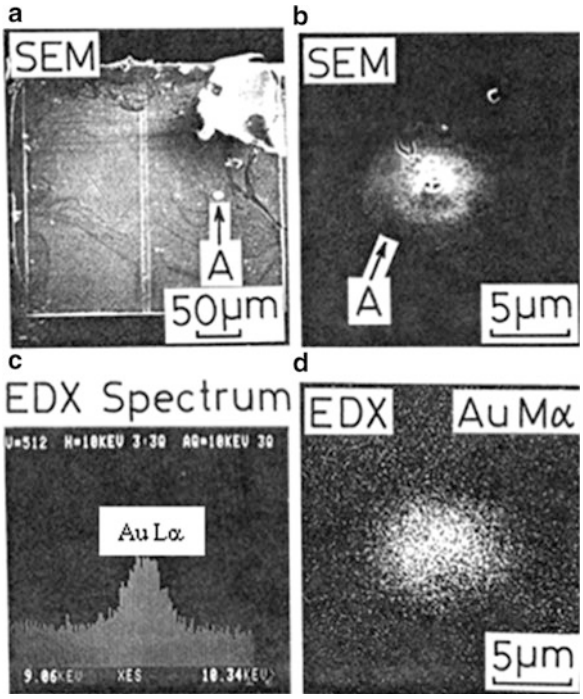



**Fig. 2.12** Results of SEM/EDX analysis on a degraded InGaAsP/InP BH laser under application of a large current pulse. (**a**) SEM image of p-InGaAsP contact layer; (**b**) high-magnification SEM image of region indicated as "A" in Fig. 2.12a; (**c**) EDX spectrum obtained from region A (point analysis); (**d**) EDX image of the region A in Fig. 2.12b (EDX mapping) (Reprinted from Ref. [15], copyright 1984. With permission from American Institute of Physics)

image from the same area. The electron beam is x-y scanned in a similar way to that for the SEM image. This analytical technique is very useful for the analyses of regions degraded by events such as catastrophic failure (due to melting of the crystal), electrode degradation (alloy reaction between the semiconductors and the electrode metals), and solder migration.

Figure 2.12 shows typical examples of EDX spot analysis and EDX mapping for a degraded InGaAsP/InP BH laser (for more detail, see Ref. [15]).

### *2.4.8 Auger Electron Spectroscopy (AES)*

At times, surface phenomena, such as very thin layers of contamination, may be present. In such cases, Auger electron spectroscopy (AES) is often favored. It penetrates no more than 0.5 nm into the sample being studied, and gradual sputtering allows a depth profile to be performed. It also has better lateral resolution

than EDX, with EDX often limited to 0.1–1 μm depending on beam energy and Auger having a minimum resolution of 10 nm (i.e., 0.01 μm).

When an electron beam with a beam energy of 3–5 keV irradiates the surface of the sample, secondary electrons are emitted from the surface through an Auger process. This method gives us the chemical composition of the region of the sample from the surface down to 1 um in depth by analyzing the energies of such secondary electrons, provided the analyst is willing to take the time to gradually sputter the material away during the scan. Since the energies of the Auger electrons are peculiar to each element, one can identify the element. By AES, all elements except for H and He, which do not participate in the Auger process, can be detected. The sensitivity is approximately 0.1–1 at.% of one atomic layer depending on the element. The energy of the Auger electrons is in a range of 10–2,000 eV.

An analysis of depth profile for an element by using ion sputtering during the AES measurement is also commonly performed. Two-dimensional mapping is also possible by using a scanning-type AES apparatus. Such an AES imaging technique is very useful for evaluation of oxidation-induced degradation of laser mirrors as well as compositional analysis of precipitates in heavily degraded devices. Another frequent example of an Auger finding would be the examination of surface contamination such as carbon films on facets or wirebond surfaces.

### *2.4.9 Examination for P-Contact Degradation*

From this process, the layered structure of the diode is examined, following removal of the upper (usually p-side) electrode. There are two methods for the removal of the electrode:

1. Removal of the electrode metals by a chemical process
2. Removal of the electrode metals by ion etching

Method 1 is used for removing metals that can be dissolved easily by some chemical solution. Since Au or Au-related alloys (i.e., Au-Zn) are often used for the electrode metals, these are easily resolved by an Au-stripping solution. Metals like Ti with high melting points are not easily dissolved by chemical solutions (although sometimes they can be peeled off in thinned HF solution under ultrasonic vibration). In such cases, they should be removed physically by Ar+ ion etching. After the electrode metals are removed, the surface of the contact layer can be investigated. The surface morphology is evaluated by optical microscope or SEM. If there is degradation at the contact region, there must be some abnormality in the contact layer (i.e., alloyed region and/or molten region). In such cases, compositional analysis using EDX is advisable. Contact degradation and/or degradation due to current surge can be investigated by this procedure.

## 2.5 Deep FA Techniques (Third-Level FA)

Third-level FA techniques generally are unaffordable for routine characterization, taking 5–20 man-hours per device or costing thousands of dollars in laboratory fees. However, such techniques are essential to truly understanding the root cause of the failure.

In TEM, since the specimens must be transparent to the electron beam, they should be thin, the maximum thickness depending on the material and the accelerating voltage of the electron microscope. For semiconductors such as GaAs and InP, it is recommended that the thickness of the specimens should be less than 1 μm for a 200-kV electron microscope. For high-resolution TEM observation, the thickness should be less than 50 nm. The primary method of sample preparation for the past decade has become the use of a focused ion beam (FIB) using a gallium ion source. Since the die are 100–400+ μm in every dimension and it is only practical to remove material for a sample up to 20–40 μm in thickness, the sample first needs to be reduced before starting thinning. This can be done in one of two ways: mechanical lapping to near the area of interest, or "lift out" where a triangular section is removed from the sample using coarse FIB cutting.

### *2.5.1 Plan-View Transmission Electron Microscopy (TEM)*

In most cases, semiconductor laser degradation begins and stays within ±0.5 μm vertically of the active region where light is created and amplified. Therefore, in most cases, it is safe to thin off the layers above and below this to allow TEM. Plan-view TEM (PV TEM) is generally far more useful than cross-sectional TEM (XS TEM), although it is both more expensive and more difficult to perform. While tedious methods involving selective etching and careful polishing were used previously, now PV TEM is done primarily by FIB [16]. A sample plan-view TEM image is shown in Fig. 2.13.

Such images are most useful where a point of origin can be identified; where did the dislocation network start? An array of low-magnification and high-magnification images would normally be taken, often a multiple tilts to bring out the different contrast of different dislocation types. The images need to be examined very carefully for linear dislocations that might be traveling through the section and have been the cause of failure [17]. One useful approach is to vary the tilt of the sample until the main dislocation array has minimal contrast and then look for features that are not eliminated at that diffraction condition.

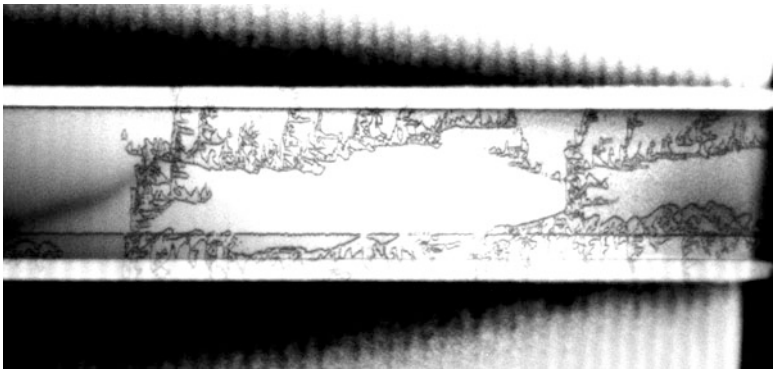


**Fig. 2.13** Plan-view TEM of a dislocation array from an AlGaInAs QW 1.3-μm laser. Such images can often be used to find the point of origin for the dislocation array

## *2.5.2 Cross-Sectional TEM*

### 2.5.2.1 FIB Method

While plan-view techniques are used most often in failure analysis, there are times where it is valuable to look at a sample from the side, usually to examine the suspected point of origin of a dislocation network, and to see what layers the dislocation network is confined to. In our usage, this most frequently comes on a plan-view TEM sample that then uses a precision lift-out technique [17]. To perform precision lift out, the first optional step is to cut alignment marks near the feature of interest with a FIB, then repeat the plan-view TEM. At times, other features will be close enough that this "alignment mark" step can be skipped. Then, on the TEM picture, the customer draws what area he would like the cross section drawn from. The FIB is used to cut along the lines requested by the customer. When the cutting nears completion, a needle is welded onto the sample by depositing metal, and the cross section is then freed from the original TEM piece and then placed on its own TEM grid. Further sample thinning can then be done. The process is shown in Fig. 2.14.

For many types of samples, the FIB works very well and produces no artifacts, or artifacts that can be characterized as arising from the TEM process. (Artifacts would be defects that might be induced by the FIB and might be visible even after preparing an undegraded laser by TEM, making it difficult to distinguish between what was added by the aging process and what was added by the FIB sample preparation.) If the FIB is suspected of producing questionable and unwanted artifacts, the first choice would normally be to perform additional low-kV polishing of the cross-sectioned faces in the FIB to remove the outermost layers of damaged material. If it proves impossible to eliminate objectionable artifacts using FIB, the analyst may wish to use traditional cross-sectional TEM sample preparation techniques as shown in Ref. [18].

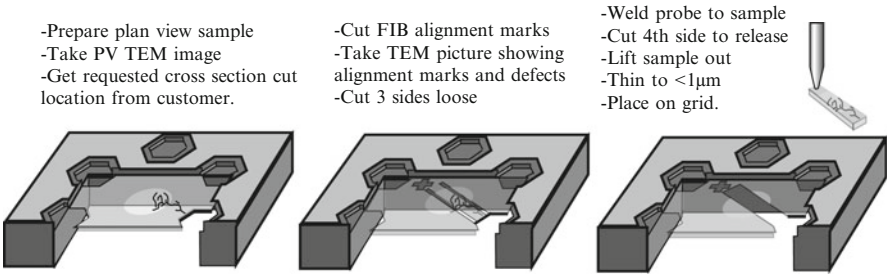



**Fig. 2.14** The process for precision lift out of a cross section from a plan-view sample. At times, the need for special FIB alignment marks and an extra TEM can be avoided if other topographic features in the sample are close enough

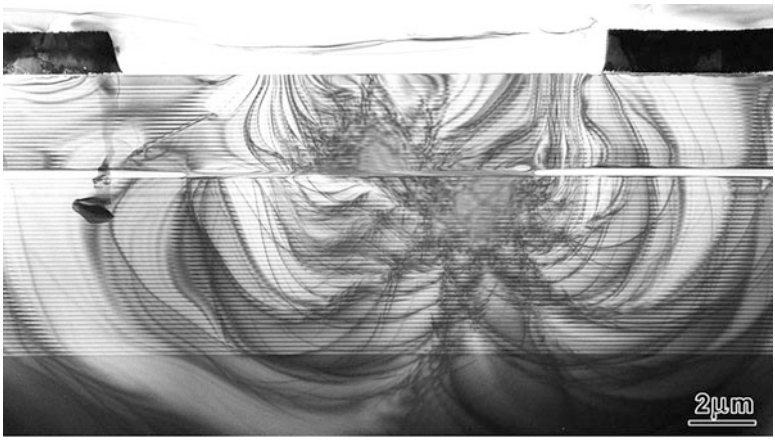



**Fig. 2.15** Typical cross-sectional image from a nonspecialist firm shows extensive bend contours that mask most of the dislocations of interest in this very early oxide VCSEL

An additional problem is frequently present both for cross-sectional and plan-view sample preparation: bend contours. Some labs will leave a "post" of thicker material in the middle of a sample like that shown below in Fig. 2.15 to avoid the extensive bending contours that mask essentially all of the features of interest in the device. For labs that do not specialize in working on this sort of sample, it is commonplace to get an unusable result like this. As we will discuss later, STEM imaging can be used to reduce the impact of bend contours.

### 2.5.2.2 Traditional Method: Dimpling and $Ar^+$-Ion Thinning

A traditional method for preparation of cross-sectional TEM specimen is still done particularly for obtaining high-resolution TEM image. In this method, we must thin

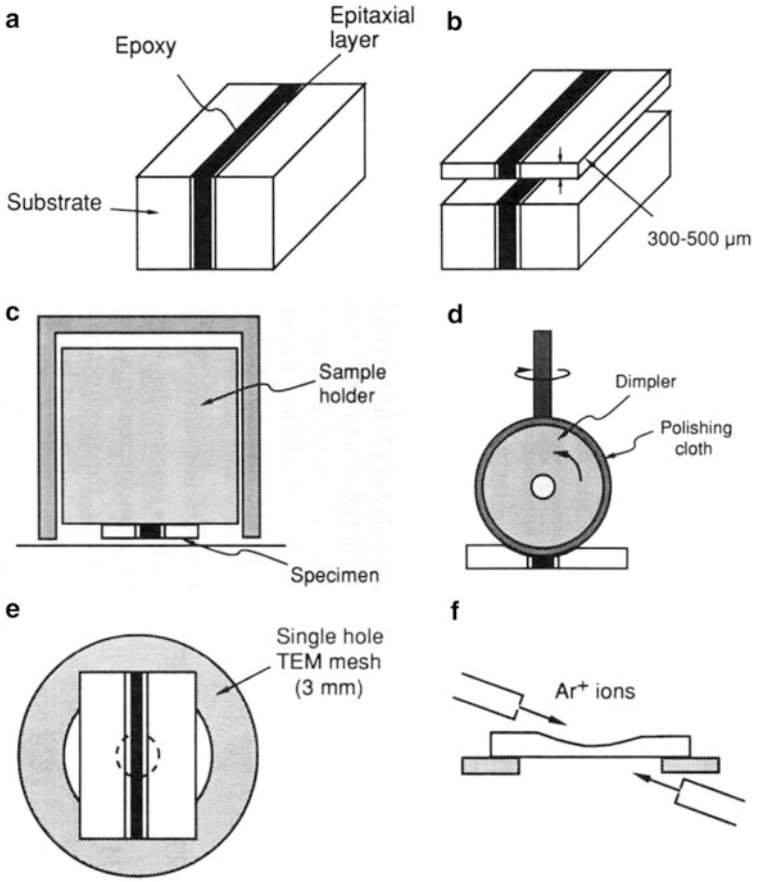



**Fig. 2.16** Schematic diagram for a traditional specimen preparation for crosssectional TEM. (**a**) Adhesion of sample epi-wafers; (**b**) cutting of the sample; (**c**) polishing of the specimen; (**d**) polishing the specimen by dimpling; (**e**) mounting of the specimen onto a TEM mesh; (**f**) final thinning by ion milling

the sample by a more complicated procedure than the FIB case. This method consists of the following steps (see Fig. 2.16):

1. Adhesion of a sample epi-wafer face to face by epoxy glue (Fig. 2.16a)
2. Cutting of the sample processed in step 1 into slices in cross section with thickness of about 0.5 mm (Fig. 2.16b)
3. Polishing of the specimen from both sides down to about 50–70 μm (Fig. 2.16c)
4. Polishing of the specimen from both sides down to about 30 μm by dimpling (making a shallow hole on the surface, Fig. 2.16d)
5. Mounting the specimen onto a single-hole TEM mesh (Fig. 2.16e)
6. Final thinning of the specimen by ion milling from both sides (Fig. 2.16f)

The final thinning must be done by ion milling. However, in the case of specimens consisting of a semiconductor-layered structure, one can perform the final thinning by chemical etching, although the thickness of the specimen is too large for HRTEM. In this case, the etching rate should be low enough to control the final stage of thinning. A $Br_2/CH_3OH$ solution is often used for this purpose.

## *2.5.3 Determination of the Nature of Defects Through TEM*

### 2.5.3.1 Contrast Measurements

Contrast measurements are a good way to identify defect characteristics such as the Burgers vectors of dislocations, stacking faults, and dislocation loops. In this experiment, only one reflection is strongly excited, that is, a two-beam (one Bragg-reflected beam and one transmitted beam) diffraction condition is used. For each contrast experiment, the deviation parameter from the exact Bragg condition is kept slightly positive so as to obtain the best contrast. In order to achieve this condition, the specimen is tilted in a double-tilting goniometer from the zone axis orientation to where the Bragg condition is satisfied. This operation is precisely controlled by the movement of Kikuchi lines on the diffraction pattern. The electron micrograph is obtained by the direct imaging of either the transmitted beam (bright-field image) or the Bragg-reflected beam (dark-field image).

If large area of the specimen is too thin, the specimen can bend due to internal stress in addition to the stress between the specimen and the glue, in which case the incident electron beam is often away from normal to the basal plane. In such a situation, the specimen should be tilted by a large angle to compensate. As a first step, the direction of the electron beam must be identified from the diffraction pattern. The author has found that this procedure becomes much easier if the electron beam is moved to an area of the specimen thick enough to give clear contrast of the Kikuchi lines. It was also realized that the problem of bending of the specimen can be solved by choosing polishing/chemical etching rather than preferential and mirror chemical etching as a specimen thinning method. Note that we have found contrast measurements with controlled Bragg vectors difficult to obtain from service labs.

### 2.5.3.2 Stereoscopic Observation

For determination of the nature of dislocation loops or dislocation dipoles, it is necessary to identify the plane on which they are lying. Such three-dimensional information is obtained by stereoscopic observation. In this procedure, a pair of electron micrographs (either bright-field images or dark-field images) should be taken under the same (two-beam) diffraction condition with a different tilting angle of the incident electron beam. This is also carried out by the use of Kikuchi lines.

## 2.5.4 Structural Determination of Defects by TEM

### 2.5.4.1 Determination of Burgers Vectors of Dislocations

Burgers vector of dislocations in the zinc blende structure are classified into types (a/2)<110> and a<100> for perfect dislocations and (a/3)<111>, (a/11)<113>, or (a/6)<211> for partial dislocations. For perfect dislocations, the diffraction contrast of dislocations is subject to the rule of invisibility as follows:

$$\mathbf{g} \bullet \mathbf{b} = \mathbf{0} \tag{2.2}$$

$$\mathbf{g} \bullet \mathbf{b} \times \mathbf{u} = \mathbf{0} \tag{2.3}$$

where **g**, **b**, and **u** correspond to a reflection vector, a Burgers vector, and a unit vector representing the direction of the dislocation line, respectively. Thus, for determination of Burgers vector of dislocations, a series of bright- or dark-field images are taken under different reflections, for example, 220, −220, 400, and 040 for the (001) reciprocal plane, and the contrast of each dislocation is checked.

### 2.5.4.2 Determination of the Character of Stacking Faults

The characterization of stacking faults is performed by comparing the contrast of fringes at the top and at the bottom of the fault in a bright-and dark-field image under a two-beam diffraction condition. The following effects are observed in the case of "$\alpha = 2\,\pi/3$" type stacking fault where $\alpha$ is an additional phase shift.

1. The dark-field fringe pattern becomes asymmetric, but the bright-field pattern remains symmetric about the foil center. At the electron entrance side (top), the dark- and bright-field fringes are anticomplementary, that is, maxima and minima appear at the same places at the same depth of the stacking faults, whereas at the exit side (bottom), the bright and dark fringes are complementary, maxima occurring in the bright field whereas minima are seen in the dark field and vice versa. This contrast effect can be used to decide how the stacking fault is inclined in the foil.
2. The first fringe in bright field is bright for $\alpha = 2\,\pi/3$ and dark for $\alpha = -2\,\pi/3$.

Therefore, if the direction of $\mathbf{g}_{hkl}$ is known with respect to the bright- and dark-field images, the nature of the stacking fault (intrinsic or extrinsic) can be determined.

### 2.5.4.3 Determination of the Character of Dislocation Loops

A dislocation loop will exhibit outside contrast for ($\mathbf{g} \bullet \mathbf{b}$) $s > 0$ and inside contrast for ($\mathbf{g} \bullet \mathbf{b}$) $s < 0$. An interstitial (extrinsic) loop will have Burgers vector **b** in an

upwards direction, while a vacancy (intrinsic) loop will have **b** in a downwards direction. Thus, if the plane on which the loop is lying is determined by stereoscopic observation, then the nature of a loop can be determined by comparing its contrast under the condition ($A$) (or ($A'$)) with ($B$) (or ($B'$)) as follows:

$$(A)\ \mathbf{g} = hkl \text{ and } s > 0 \tag{2.4}$$

$$(A')\ \mathbf{g} = -h - k - l \text{ and } s > 0 \tag{2.5}$$

$$(B)\ \mathbf{g} = hkl \text{ and } s < 0 \tag{2.6}$$

$$(B')\ \mathbf{g} = -h - k - l \text{ and } s < 0 \tag{2.7}$$

## *2.5.5 Advanced TEM and Other Related Techniques*

### 2.5.5.1 High-Angle Annular Dark-Field Scanning Transmission Electron Microscopy (HAADF-STEM)

One of the methods that has become more prominent in the past 5 years is the use of high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM). Such imaging can be performed in a conventional TEM, but the beam is instead rastered in an SEM manner, and scattered electrons are captured by an annular detector. "Clean" material provides few scattered electrons and a dark image, while defects scatter electrons and show up as bright edges. One advantage of this technique is that it is less sensitive to sample bend contours as described earlier in the section on cross-sectional TEM; it is said to use "incoherent" imaging rather than "coherent" imaging [19]. It also requires less skill from the TEM analyst, as the use of twin-beam diffraction conditions is not needed. A sample image is shown below in Fig. 2.17. Another application, requiring a very high degree of skill, is to obtain atomic resolution of threading dislocation defects and clusters in InGaN lasers [19–21].

### 2.5.5.2 3-Dimensional Atomic Probe Tomography (3DAP or APT)

Atomic Probe Tomography (APT) or 3-dimensional atomic probe (3DAP) has been popular for studying clustering at the atomic level in ceramics and metals for the past decade. It requires the preparation of a small sample into a very sharp tip, that is then destroyed during the atom probe process [22, 23]. It has also been applied to studying the structure on "v-shaped pits" in InGaN lasers, and the indium clustering [20, 21, 24].

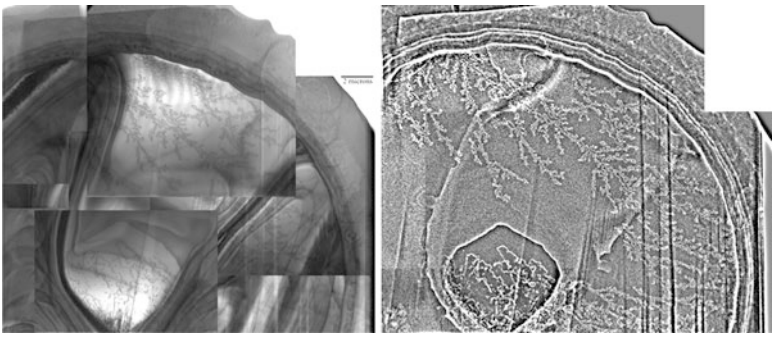


**Fig. 2.17** A conventional TEM plan-view image of an early oxide VCSEL is shown at the *left*, and is heavily obscured by worse-than-normal bend contours on this sample; at the *right*, the HAADF-STEM image shows that it is relatively insensitive to bend contours, and the DLD network can be traced fairly clearly

### *2.5.6 Cathodoluminescence*

When an electron beam irradiates a semiconductor surface, luminescence light is emitted in the crystal by radiative recombination of minority carriers generated by the electrons. This is called cathodoluminescence (CL). This CL light is collected by lens, mirrors, photomultipliers, or pin photodiodes. CL images are projected on a display, and CL spectra may also be obtained using a monochrometer. To enhance the emission efficiency, it is necessary to cool the sample with liquid nitrogen or helium. This method is carried out in a scanning electron microscope (SEM) or transmission electron microscope (TEM). Spectrally filtered CL gives images of just the wavelength peaks of interest, for example, from the active region, and filtering out the CL image from the cladding and waveguiding layers.

CL is particularly useful because unlike most of the other techniques which only allow inspection of the layers near the quantum well, CL allows excitation of a broad area of the laser, including cladding and contact layers well above and below. Furthermore, while TEM may only be capable of imaging coherent defect structures, CL can detect the action of traps due to widely dispersed point defects that cannot be seen in TEM. Finally, unlike other techniques that require the sample to have electrical contacts, CL has no such requirements, giving more flexibility to sample preparation. On the negative side, the cladding layers of some semiconductor lasers have an indirect bandgap that does not efficiently luminesce, for example, $Al_xGa_{1-x}As$ with $x > 0.4$ is indirect. Second, CL appears to only be accessible at universities as of this writing and has only recently become available in one service lab of the many we have contacted. Also, plan-view CL requires difficult sample preparation, with very careful removal of either the p-contact layer (something that would only make sense for planar devices) or careful backside etching. Cross-sectional CL is much easier and would only require a basic cross-sectioning ability.

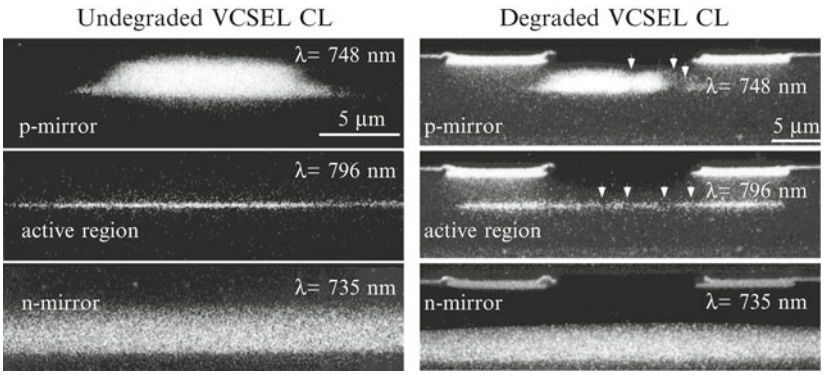



**Fig. 2.18** Spectrally filtered cathodoluminescence of cross-sectioned samples enables us to separately image contributions from each of the three major VCSEL sections and detect degradation in the aged device. The *arrows* in the *right* image show areas of reduced luminescence

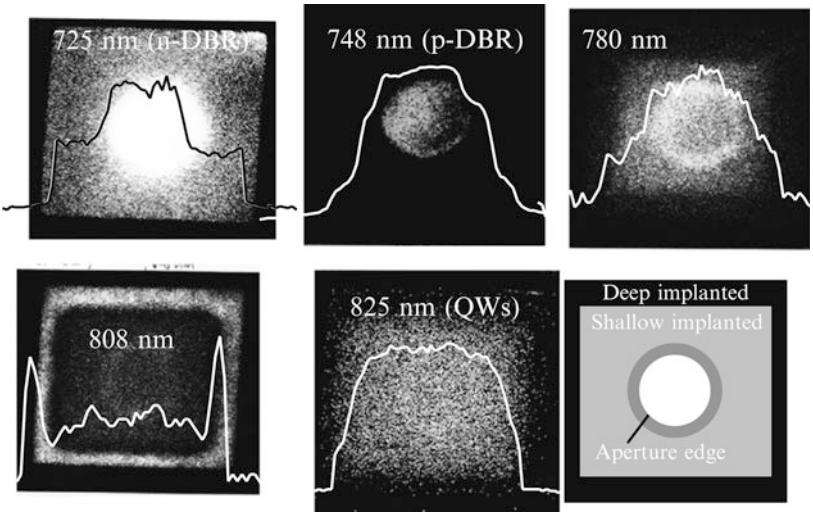



**Fig. 2.19** Spectrally resolved plan-view cathodoluminescence allows us to map degradation in *x* and *y* looking at the device from the *bottom*, with the substrate and some of the n-DBR mirror polished away

To get the best spatial resolution from CL, it is important to understand the excitation volume from the electron impact and how that varies as a function of beam energy. For the best resolution, two approaches can be taken. The first is to use low beam voltage on the electron microscope to minimize the excitation volume. The second is to use a TEM-type sample preparation and thin the sample to where the beam goes through it [25].

Sample cross-sectional [26] and plan-view CL [27] images are shown below in Figs. 2.18 and 2.19.

### 2.5.7 Deep-Level Transient Spectroscopy (DLTS)

DLTS measurement is a classical method for evaluation of characteristics of deep energy levels, which is useful for determining the exact type of defect present. This could, of course, be very useful in determining what defect you need to eliminate to improve device reliability (e.g., a specific type of heavy metal contamination). This method studies trap filling and emptying speeds at various temperatures. However, it is not widely available and is considered challenging to do. Furthermore, its usefulness is limited to defects that are present along the swept edges of the depletion region. Defects in the center, or outside the depletion region, cannot be detected. In addition, it can be difficult to use DLTS on narrow stripe lasers, since the aged stripe represents only a small portion of the entire chip area being probed. In such cases, better data could be collected by using a special structure where the aged portion of the chip was increased to include a majority of the die area. DLTS is used for research purposes but very rarely used in industry [18, 28–30].

### 2.5.8 Scanning Probe Microscopy

Scanning probe microscopy involves variations of an atomic force microscope (AFM) looking at a cleaved cross section of the device to study the nature of the degradation. The AFM has various sharp tips that probe the sample, as it is rastered underneath it. Scanning probe microscopy (SPM) is best suited for studying issues like dopant modifications or the electronic properties of defects at very high resolution. However, it has not been widely used for failure analysis of semiconductor lasers to this point. However, we have used it for studying degraded VCSELs with EFM and SCM as mentioned below, as well as high-resolution scanning thermal microscopy (SThM) [31] and near-field scanning optical microscopy (NSOM) [32].

#### 2.5.8.1 Electromotive Force Microscopy (EFM)

The Atomic Force Microscope (AFM) often operates in “tapping mode,” where the tip slightly vibrates, coming close to touching the sample, where it is attracted to the surface. Depending on the doping of the sample (p-type vs. n-type; highly doped vs. intrinsic), the attraction of the probe will vary. This will show up as “quasi topography” and be mapped as a step of a fraction of a nanometer. Assuming you have a truly flat-cleaved surface, this gives a non-quantitative 2-D map of unbiased Fermi potential. This technique has the advantage of not requiring any special tips or attachments; any AFM that can operate in tapping mode can be used to take EFM images.

#### 2.5.8.2 Scanning Capacitance Microscopy (SCM)

Scanning capacitance microscopy (SCM) involves the use of a special tip and electronics for the AFM. It performs C-V profiling of each point it scans across, allowing it to deduce whether the material below the tip is p-type or n-type and to quantify how high the doping level is. It allows a high-resolution quantitative map of a doping cross section in two dimensions; no reports of plan-view profiling have ever been made, to the best of our knowledge. This has been used to good effect to provide extremely useful maps of doping profiles in silicon microelectronics, as well as a few times in semiconductor lasers [33, 34].

Unfortunately, in addition to the requirement to have some not-widely-available hardware added to the AFM, the sample preparation is very challenging. Silicon samples are prepared using a chemi-mechanical polish as the final step. While this works well on samples with uniform composition like Si or constant III-V compositions, all semiconductor lasers involve a stack of widely varying ternary or quaternary alloys, and it is difficult or impossible to find an etch that has roughly the same etch rate for all the layers; in the absence of such an etch, you either have significant topography in the final sample which interferes with SCM or you have to use purely mechanical polishing, which leaves a subsurface damage layer deep enough that the SCM tip would not work. Our best results were obtained with a weak bromine etch solution on a hard buffing pad; etches that soften but do not remove etch by-products are best, as those allow chemo-mechanical planarization. The other choice is cleaving a sample at the point of interest; if that is able to give a clean surface, then cleaving would be the method of choice. While use of SCM is obviously a challenge in developing the right sample preparation, it offers unprecedented insight into the electric effects of defects and accurate, quantitative information on sample doping profiles.

#### 2.5.8.3 Scanning Spreading Resistance Microscopy (SSRM)

Scanning spreading resistance microscopy (SSRM) solves many of the same issues as scanning capacitance microscopy does and is used for many of the same issues. It would primarily study complex 2-D doping and regrowth profiles for potential problems, especially for InGaAsP lasers with multiple blocking regrowths to define current blocking layers [35, 36]. While useful qualitative information can be obtained more easily by cleaving and using wet chemical stain-etching as mentioned earlier, quantitative results can be obtained by SSRM.

In SSRM, resistance is measured between the tip and the base of the sample. The tip is usually a very hard, conductive substance. Diamond and SiC tips are common. Unlike other forms of scanning probe microscopy, very heavy pressure is often used in a way that makes SSRM "destructive," where it effectively "plows" a square-shaped depression in the sample. If this is done for long enough, 3-D mapping could be done, although such 3-D characterization would make more sense for FET devices than for semiconductor lasers, given the much smaller thickness of the transistors than the lasers.

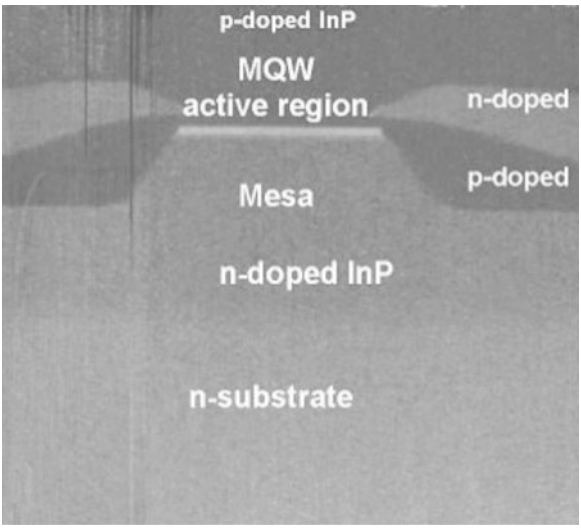



**Fig. 2.20** SSRM image of a buried heterostructure laser (Reprinted from Ref. [37], copyright 2002. With permission from AVS, The Science and Technology Society)

Figure 2.20 is a typical SSRM image of a InGaAsP / InP buried heterostructure laser. With a tip voltage of +0.5V, the current from the tip to the p-doped sections is orders of magnitude smaller than the current to the MQW or n-doped section, and thus the former appears darker than the latter.

## 2.6 Summary

In summary, we have reviewed nearly all of the techniques used for examining semiconductor laser failures, from the most commonplace to relatively rare research techniques. As final advice, many analysts have a background in material science but relatively little understanding of laser physics or device processing. This can make it difficult to understand whether the defect might be responsible for the failure observed or how to fix the defect. So learning more about device operation and fabrication is always recommended. In general, the defects of interest are usually near the active region and in the pumped area of the device. Diagnoses assigning responsibility for failure to a piece of dust on the top of the device are usually wrong. Finally, we have showed in this chapter how you build deep profiles of various failure mechanisms early, see how they correlate with easily available characterization data to create a "failure signature," and then leverage that data later to speed up and streamline the failure analysis process.

**Acknowledgments** The authors would like to thank Dr. David Venables of JDSU and Dr. Michael Salmon on Evans Analytical (Raleigh, NC) for a careful reading of the manuscript and their suggestions.

## References

1. F. Magistrali, E. Mariani, G. Salmini, M. Vanzi, Failure analysis of 980 nm high power lasers, in *Proceedings of ISTFA*, vol. 20, 1994, pp. 335–340
2. J.J. LePore, An improved technique for selective etching of GaAs and Ga1-xAlxAs. J. Appl. Phys. **51**, 6441 (1980)
3. O. Ueda, I. Umebu, S. Yamazaki, K. Oinuma, T. Kaneda, T. Kotani, Characterization of defects in liquid phase epitaxial InP and InGaAsP crystals by scanning electron microscopy and electron beam induced current method. J. Electron Microsc. **33**(1), 1 (1984)
4. O. Ueda, I. Umebu, S. Yamakoshi, T. Kotani, Nature of dark defects revealed in InGaAsP/InP double heterostructure light emitting diodes aged at room temperature. J. Appl. Phys. **53**(4), 2991–2997 (1982)
5. R.G. Waters, Diode laser degradation mechanisms: a review. Prog. Quantum Electron. **15**(3), 153–174 (1991)
6. J. Guenter, D. Mathes, B. Hawkins, J. Tatum, Developments at Finisar AOC, in *Proceedings of SPIE*, vol. 6908, San Jose, USA, 2008, vol. 6908, p. 690805
7. R.W. Herrick, Failure analysis and reliability of optoelectronic devices, in *Microelectronics Failure Analysis Desk Reference*, 5th edn. (ASM International, Materials Park, Ohio, 2004), pp. 230–254
8. R. S. Mann, D. K. McElfresh, Categorizing light output degradation failures in LEDs using the relationship between defect revealing mechanisms responsible for EL, CL, EBIC, and reverse bias photoemission (RP), in *Proceedings IEEE International Reliability Physics. Symposium*, Las Vegas, NV, 1995, pp. 177–186
9. D.K. McElfresh, L.D. Lopez, D. Vacar, Reverse-bias emission sheds light on the failure mechanism of degraded vertical-cavity surface-emitting lasers. J. Appl. Phys. **99**(12), 123113 (2006)
10. H. Ishikawa, H. Imai, T. Tanahashi, Y. Nishitani, M. Takusagawa, K. Takahei, V-grooved substrate buried heterostructure InGaAsP/InP laser. Electro. Lett. **17**(13), 465–467 (1981)
11. G. Olsen, M. Ettenberg, Universal stain/etchant for interfaces in III-V compounds. J. Appl. Phys. **45**(11), 5112–5114 (1974)
12. R.E. Ewing, D.K. Smith, Compositional inhomogeneities in GaAs1-xPx alloy epitaxial layers. J. Appl. Phys. **39**(13), 5943–5948 (1968)
13. O. Ueda, I. Ushijima, T. Tanahashi, Detecting Method for PN Junction in Compound Semiconductor. Japanese Patent S58-563501983
14. T. Kotani, O. Ueda, K. Akita, Y. Nishitani, T. Kusunoki, O. Ryuzan, Direct observation of defects in Si-doped and Ge-doped Ga0. 9Al0. 1As epitaxial layers by transmission electron microscopy. J Crystal Growth **38**(1), 85–92 (1977)
15. O. Ueda, H. Imai, A. Yamaguchi, S. Komiya, I. Umebu, T. Kotani, Mechanism of catastrophic degradation in 1.3um V-grooved substrate buried-heterostructure lasers with the application of large pulsed currents. J. Appl. Phys. **55**(3), 665–669 (1984)
16. J. Mayer, L.A. Giannuzzi, T. Kamino, J. Michael, TEM sample preparation and FIB-induced damage. MRS Bull. **32**, 400–407 (2007)
17. T.J. Stark, P.E. Russell, C. Nevers, 3-D defect characterization using plan view and cross-sectional TEM/STEM analysis. in *Proc of ISTFA*, San Jose, California, USA, 2005, vol. 31, p. 344
18. O. Ueda, Chapter 6: Evaluation techniques for III-V compound semiconductors and degraded optical devices, in *Reliability and Degradation of III-V Optical Devices* (Artech House Publishers, Boston/London, 1996), pp. 45–74

19. M. Shiojiri, High-resolution electron microscopy observations of GaN-based laser diodes, in *III-Nitride Devices and Nanoengineering* (Imperial College Press, London, 2008), pp. 159–188
20. S. Tomiya, M. Ikeda, S. Tanaka, Y. Kanitani, T. Ohkubo, K. Hono, Structural defects in GaN-based materials and their relation to GaN-based laser diodes, in *MRS Proceedings*, Boston, Massachusetts, USA, 2009, vol. 1195
21. S. Tomiya, Y. Kanitani, S. Tanaka, T. Ohkubo, K. Hono, Atomic scale characterization of GaInN/GaN multiple quantum wells in V-shaped pits. Appl. Phys. Lett. **98**, 181904 (2011)
22. T.F. Kelly et al., Atom probe tomography of electronic materials. Annu. Rev. Mater. Res. **37**, 681–727 (2007)
23. T.F. Kelly, M.K. Miller, Invited review article: atom probe tomography. Rev. Sci. Instrum. **78**, 031101 (2007)
24. M.J. Galtrey et al., Three-dimensional atom probe studies of an InxGa1-xN/GaN multiple quantum well structure: assessment of possible indium clustering. Appl. Phys. Lett. **90**(6), 061903–061903 (2007)
25. B. Bastek et al., Nano-scale correlation of structural and optical properties of DBRs using STEM CL. Proc of SPIE **7939**, 7939–46 (2011)
26. Y.M. Cheng, R.W. Herrick, P.M. Petroff, M.K. Hibbs-Brenner, R.A. Morgan, Degradation studies of proton-implanted vertical cavity surface emitting lasers. Appl. Phys. Lett. **67**(12), 1648–1650 (1995)
27. R.W. Herrick, P.M. Petroff, Gradual degradation in 850-nm vertical-cavity surface-emitting lasers. Quantum Electron. IEEE J. **34**(10), 1963–1969 (1998)
28. D. Lang, Deep-level transient spectroscopy: a new method to characterize traps in semiconductors. J. Appl. Phys. **45**(7), 3023–3032 (1974)
29. J.W. Tomm et al., Deep level spectroscopy of high-power laser diode arrays. J. Appl. Phys. **84**, 1325 (1998)
30. M Meneghini et al, Investigation of the deep level involved in InGaN laser degradation by deep level transient spectroscopy. Appl. Phys. Lett. **99**(9), 093506–093506-3 (2011)
31. K. Luo, R. Herrick, A. Majumdar, P. Petroff, Scanning thermal microscopy of a vertical-cavity surface-emitting laser. Appl. Phys. Lett. **71**, 1604–1606 (1997)
32. J.A. DeAro, K.D. Weston, R.W. Herrick, P.M. Petroff, S.K. Buratto, Near-field scanning optical microscopy of cleaved vertical-cavity surface-emitting lasers. Semiconduct. Sci. Technol. **13**, 1364 (1998)
33. J.J. Kopanski, Scanning capacitance microscopy for electrical characterization of semiconductors and dielectrics, in *Scanning Probe Microscopy*, vol. 1 (Springer, New York, 2007), pp. 88–112
34. O. Douheret, K. Maknys, S. Anand, Electrical characterization of III-V buried heterostructure lasers by scanning capacitance microscopy, in *Scanning Probe Microscopy* (Kluwer, Dordrecht, 2005), pp. 413–424
35. P. Eyben, W. Vandervorst, D. Alvarez, M. Xu, M. Fouchier, Dopant profiling in InP with SSRM (a subsection of 'probing semiconductor technology and devices with scanning spreading resistance microscopy'), in *Scanning Probe Microscopy: Electrical and Electromechanical Phenomena at the Nanoscale* (Springer, New York, 2007), pp. 73–87
36. S. Kuntze, D. Ban, E. Sargent, S.J. Dixon-Warren, J. White, K. Hinzer, Electrical scanning probe microscopy: investigating the inner workings of electronic and optoelectronic devices. Crit. Rev. Solid State Mater. Sci. **30**(2), 71–124 (2005)
37. D. Ban et al., Two-dimensional transverse cross-section nanopotentiometry of actively driven buried-heterostructure multiple-quantum-well lasers. J. Vacuum Sci. Technol. B Microelectron. Nanometer Struct. **20**(6), 2401 (2002)

# Chapter 3
# Failure Analysis Using Optical Evaluation Technique (OBIC) of LDs and APDs for Fiber Optical Communication

**Tatsuya Takeshita**

**Abstract** By employing the optical-beam-induced current (OBIC) measurement technique, we have analyzed the sudden and wear-out failure of optical devices. The extent of the degraded region is estimated by using relative OBIC intensity prior to aging. The use of OBIC incident sources at several wavelengths enables us to detect degradation in facets, epitaxial layers, and the device interior.

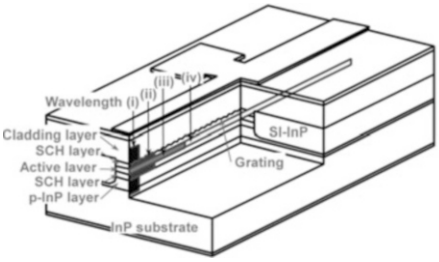
## 3.1 Introduction

Optical-beam-induced current (OBIC) measurement [1–4], as well as other approaches such as electron-beam-induced current (EBIC) measurement [5] and luminescence measurement [6], can determine the existence of deterioration in optical and electronic devices. Of these approaches, the OBIC technique has the advantages of being nondestructive and capable of selecting wavelengths [7].

First, since the OBIC is measured through the window of a transistor outline (TO) can before and after aging, we are able to detect an OBIC change for several periods of aging. When there are nonradiative recombination centers in the degraded region, the OBIC intensity decreases as the recombination density increases [8]. The relative OBIC intensity prior to aging is useful for analyzing the degree of device degradation. Furthermore, the incident light that illuminates the facet is absorbed in the semiconductor. However, the lights incident on the top surface rather than the facet are absorbed in the layer to a depth of not more than a few micrometer. The OBIC measurement technique that involves illuminating the facet is therefore expected to be highly susceptible to degradation.

T. Takeshita (✉)
Electrical Engineering and Information Technology, Faculty of Engineering,
Kyushu Sangyo University, Matsukadai, 2-3-1, Higashi-ku, Fukuoka 813-8503, Japan
e-mail: takeshita@te.kyusan-u.ac.jp

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_3,
© Springer Science+Business Media New York 2013




| λ | Absorption layer | $d_p$ (μm) | $\alpha_0$ ($cm^{-1}$) | $\Gamma$ |
|---|---|---|---|---|
| (i) | All layers | 1 | 10000 | 1.0 |
| (ii) | Active and SCH layers | 3.3 | 10000 | 0.3 |
| (iii) | Active layer | 10 | 10000 | 0.1 |
| (iv) | Active layer | 40 | 2500 | 0.1 |
| (iv) | Active layer | 100 | 1000 | 0.1 |
| (iv) | Active layer | 400 | 250 | 0.1 |

**Fig. 3.1** Penetration distance as a function of wavelength

Second, the incident light can be changed by replacing the optical source in the OBIC measurement setup, which in turn changes the absorption layer and the penetration distance. Figure 3.1 shows the penetration distance as a function of wavelength. The light of an incident beam source illuminates the measured laser facet and is propagated through the laser waveguide. The incident light power ($P_i(x)$) at a distance ($x$) from the AR facet is given by [9]

$$P_i(x) = P_{i0} \exp(-\Gamma \alpha_0 x) \tag{3.1}$$

where $P_{i0}$ is the incident power, $\alpha_0$ is the absorption coefficient, and $\Gamma$ is the optical confinement factor. Light in wavelength ranges (i), (ii), and (iii) is absorbed in all the laser layers, both the active and separate confinement heterostructure (SCH) layers and only in the active layer, respectively. The difference between the absorption regions is an attractive way of clarifying the degraded layers in devices.

We define the penetration distance as the distance at which the strength of the incident power decreases to 1/e. As shown in Fig. 3.1, the penetration distances are estimated to be 10, 3.3, and 1 μm, assuming that $\alpha_0$ is $10^4$ $cm^{-1}$ and the $\Gamma$ values of the measured laser waveguide are 0.1, 0.3, and 1.0 for lights with wavelength ranges of (iii), (ii), and (i), respectively. Since the wavelength lights are absorbed in the laser waveguide around the AR facet, these are useful for detecting degradation around the facet.

Furthermore, light in wavelength range (iv) is absorbed in the semitransparent active layer and is longer than the PL peak wavelength of the active layer. The penetration distance increases as the absorption coefficient ($\alpha_0$) decreases [10]. Lights with long penetration are beneficial for detecting degradation in the interior of a laser waveguide [11]. Thus, the analysis application area changes depending on the light wavelength.

These OBIC properties are useful for clarifying the degradation mechanism and postponing or suppressing degradation thus allowing us to achieve a reliable device. Recently, the introduction of high-speed services for fiber-optic access subscribers has led to a huge growth in data traffic. The rapid diversification of services means that next-generation networks must be built quickly, economically, and reliably [12], and this requires high-performance devices to operate under severe

conditions. However, severe conditions reduce the lifetime of a device [13, 14]. Therefore, a reliable and high-performance device is one of the keys to achieving highly reliable and cost-effective systems.

In this chapter, we introduce some examples using the OBIC technique and discuss the sudden and wear-out failure of laser diodes (LDs) as well as the wear-out failure of avalanche photodiodes (APDs), which contributed to the improvement of device reliability. In addition, in the region that the incident light from the facet does not reach, the planar-view OBIC technique is applied to a degraded device by removing the electrode.

## 3.2 OBIC Analysis

### *3.2.1 Principle of OBIC Generation*

Figure 3.2 shows principle of OBIC generation when using light absorbed in (a) only the active layer, (b) both the active and SCH layers, and (c) all the layers. When a light strikes a semiconductor, photons are absorbed due to the formation of electron–hole pairs. When we use the lights (iii) and (ii) absorbed in only the active layer and both the active and SCH layers, respectively, the generated electron–hole pairs are separated in the depletion region of the pn junction and the separated carriers produce a current through the p and n electrodes of the laser. Next, when we use the light (i) absorbed in all the laser layers (the active, SCH, and cladding InP layers), the obtainable OBICs are both the current generated in the depletion layer and the current that reaches the depletion layer from another region. Even though electron–hole pairs are generated in the InP layer by the light absorption, carriers that not reach the depletion layer are extinguished through a recombination process and we cannot obtain photo-induced current.

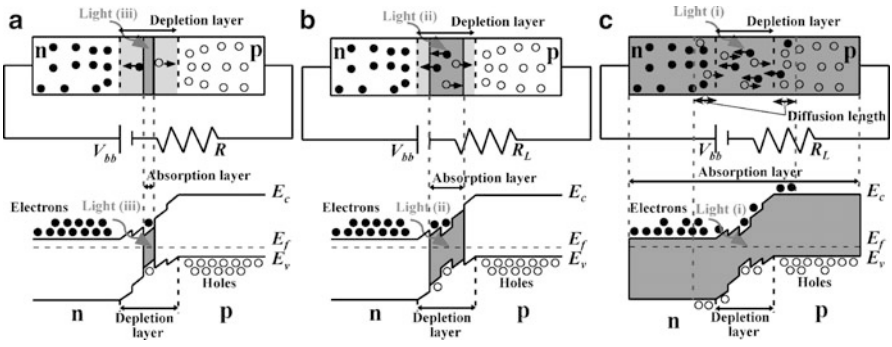



**Fig. 3.2** Principle of OBIC generation when using light absorbed in (**a**) only the active layer, (**b**) both the active and SCH layers, and (**c**) all the layers

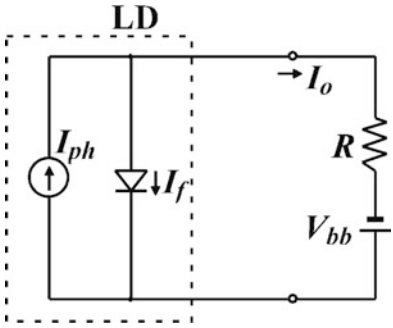



**Fig. 3.3** OBIC measurement method

Figure 3.3 shows a photo-induced current measurement circuit. Provided that the forward current $I_f$ of the LD is mainly diffusion current, the output current $I_o$ is given by [9]

$$I_o = I_{ph} - I_f = I_{ph} - I_{so}\left(e^{qV/kT} - 1\right) \tag{3.2}$$

where $I_{ph}$ is the photo-induced current, $I_{so}$ is the saturation current, $q$ is the electron charge, $V$ is the applied voltage, $k$ is Boltzmann's constant, and $T$ is the absolute temperature. Using a resistance $R$ of almost 0 Ω and an applied voltage of 0 V, $I_o$ is given by

$$I_o \approx I_{ph} \tag{3.3}$$

This indicates that the output current is directly proportional to the photo-induced current.

### *3.2.2 Digital OBIC Scanner*

The OBIC was measured using a digital OBIC scanner (JEOL Ltd.) [15]. Figure 3.4 shows the digital OBIC scanner setup. The incident light wavelength was selected by changing the incident light LD. After the incident light was collimated, the beam was scanned along the x (512 points) and y (512 points) axes by employing two galvanometer mirrors. The point-to-point interval corresponds to about 0.05 μm. The incident light power through the objective lens is adjusted to 1 μW, and the spot size of the incident light is about a wavelength. The OBIC setup is also employed for confocal laser scanning microscopy, and we use it to confirm the position of the laser waveguide.

The light illuminated the measured device facet through the window of the TO can. When light strikes a semiconductor surface, photons are absorbed owing to the formation of electron–hole pairs and directly induce an output current. We adjusted the focus to maximize the peak of the OBIC intensity. To monitor the measurement accuracy, we first measured the OBIC intensity by employing the same reference laser for every measurement and confirmed that the error range of the measured intensity was within ±0.5%.

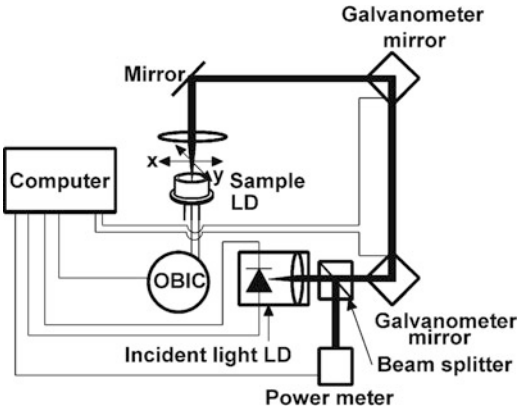



**Fig. 3.4** Setup of digital OBIC scanner

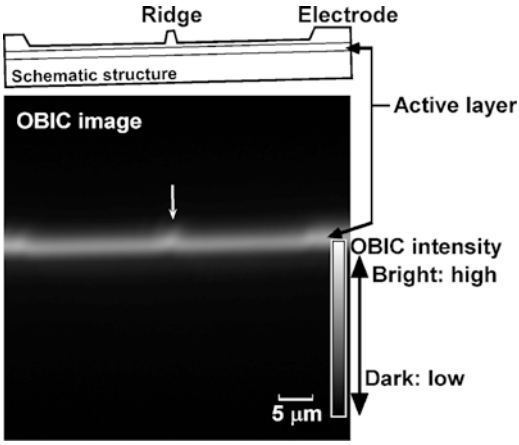



**Fig. 3.5** Digital OBIC scan of the AR facet of an undegraded 0.98-μm wavelength InGaAs/GaAs strained quantum-well ridge-waveguide laser before aging

Figure 3.5 shows a digital OBIC scan of the AR facet of an undegraded 0.98-μm wavelength InGaAs/GaAs strained quantum-well (QW) ridge-waveguide laser before aging. The incident light has a wavelength of 0.97 μm, which is efficiently absorbed in the active layer. The bright regions indicate a high OBIC intensity. A bright horizontal line is obtained around the active layer. The white arrow corresponds to the active layer under the ridge, and there is no degradation with this laser [15].

## 3.3 Sudden Failure

Devices that fail suddenly are not eliminated by screening tests. We must therefore suppress sudden failure if we are to realize more reliable transmission systems. This section addresses sudden failure, which is classified into two OBIC property modes; one is due to degradation at the AR facet (Sect. 3.3.1), and the other is due to degradation in the vicinity of the AR facet (Sect. 3.3.2).

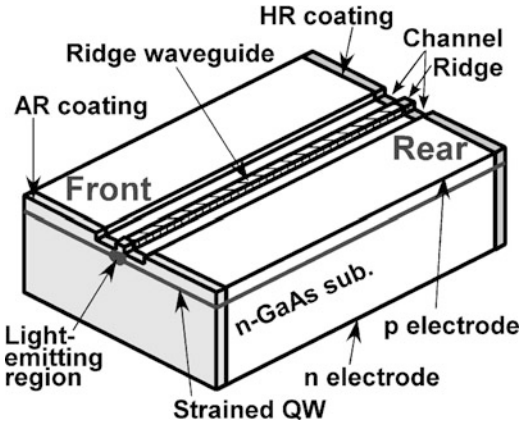



**Fig. 3.6** Schematic structure of 0.98-μm InGaAs/GaAs strained QW laser

### *3.3.1 Facet Degradation*

As the aging time increases, the catastrophic optical damage (COD) [16] level decreases due to facet degradation [17]. Sudden failure occurs when the COD level reaches the operating power level [18]. The facet degradation of InGaAs/GaAs strained-layer lasers has been analyzed with the OBIC technique.

#### 3.3.1.1 Features of COD Failure

We employed 0.98-μm InGaAs/GaAs strained QW lasers, whose schematic structure is shown in Fig. 3.6. The laser has a ridge-waveguide structure formed with the electron cyclotron resonance (ECR) plasma etching technique [19]. The ridge waveguide was 2.5 μm wide and 900 μm long [20]. The active layer was a double quantum well. A silicon dioxide layer was deposited to isolate the laser's epitaxial layers from the p-electrode metal and was removed from the top of the ridge to allow a driving current to be input. The reflectivity of the front facet was reduced to 5% and that of the rear facet increased to 85%. A cleaved laser chip was mounted on CuW heat sink, and then the whole structure was mounted on a 9-mm TO can.

Figure 3.7 shows the driving current dependence of the continuous wave (CW) light power for a laser before aging. The light power is more than 600 mW and exhibits an irreversible decrease. Figure 3.8 shows an AR facet OBIC scan of the laser degraded by the CW *I-L* measurement. The dark region in the center of the bright horizontal line (as indicated by the white arrow) corresponds to the active layer under the ridge and indicates degradation around the AR facet [17].

Figure 3.9 shows a plane-view TEM image around the AR facet of a laser that had suffered a sudden failure. The facet was coated with C and W cover layers to provide a suitable TEM specimen. There are a small protrusion and a recrystallized droplet around the AR facet. At first, a large recrystallized droplet shape [21]

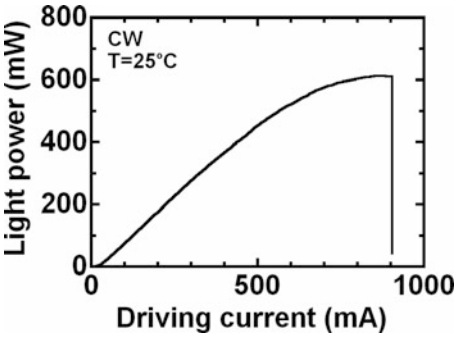



**Fig. 3.7** Driving current dependence of the CW light power for a laser before aging

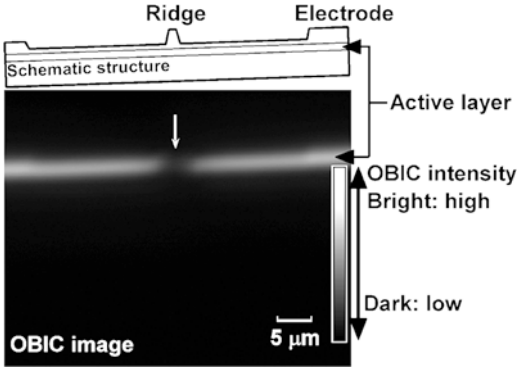



**Fig. 3.8** Digital OBIC scan on the AR facet for a laser degraded by CW *I-L* measurement (Modified from Takeshita et al. [2])

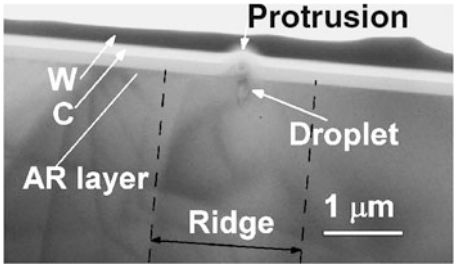



**Fig. 3.9** Plane-view TEM image around the AR facet for a laser that suddenly failed (Modified from Takeshita et al. [2])

appeared in the ridge waveguide. However, it disappeared with additional thinning during TEM sample fabrication. These phenomena indicate that the facet mirror was damaged by heat and that the failure was due to COD around the AR facet.

#### 3.3.1.2 Facet Degradation due to Aging

An aging test was carried out under a constant light power of 150 mW at 50°C. We assumed that the active layer under the electrode does not degrade with aging time,

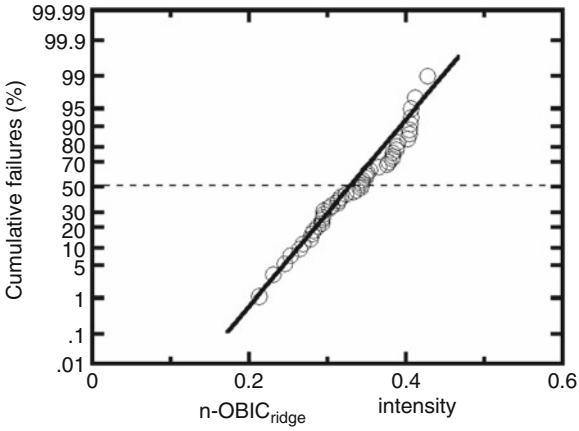



**Fig. 3.10** Lognormal probability plot of *n-OBIC*$_{ridge}$ intensity for lasers that suddenly failed (Modified from Takeshita et al. [2])

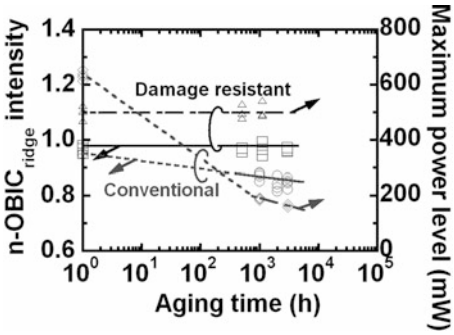



**Fig. 3.11** Changes in *n-OBIC*$_{ridge}$ intensity and maximum power level as a function of aging time for conventional and damage-resistant lasers (Modified from Takeshita et al. [17])

and we normalized the OBIC intensity of the active layer with that under the electrode (*n-OBIC*) to indicate the degree of laser degradation. We focused on the *n-OBIC* intensity under the ridge (*n-OBIC*$_{ridge}$) when estimating facet degradation. Figure 3.10 shows the lognormal probability plot of the *n-OBIC*$_{ridge}$ intensity for the lasers that suddenly failed during aging. The plot is essentially a straight line [22], which indicates that the failure mode is governed by a single degradation mode. In addition, the *n-OBIC*$_{ridge}$ intensity drops significantly below about 0.4.

Figure 3.11 shows the changes in the *n-OBIC*$_{ridge}$ intensity and maximum power level as a function of aging time for lasers with a conventional facet coating layer and a damage-resistant coating layer. These lasers were periodically removed from the aging equipment, and both the *n-OBIC*$_{ridge}$ and the *I-L* characteristics were measured at 25°C. With the conventional lasers, the *n-OBIC*$_{ridge}$ intensity decreased from 0.95 to 0.85 after 3,000 h. This 10% reduction is due to an increase in the defect density, resulting in enhanced nonradiative recombination. In conjunction with this, the COD level dropped from more than 600 to about 170 mW. This reveals that there is degradation around the AR facet of the conventional lasers. On the other hand, with the damage-resistant lasers, there was no change in

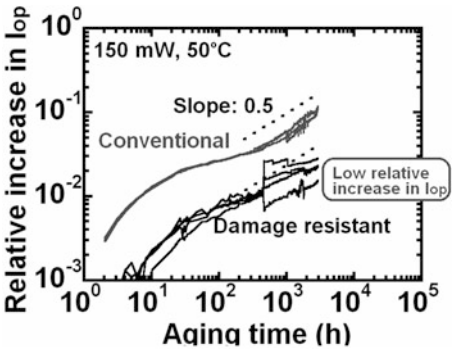



**Fig. 3.12** Aging time dependence of relative increase in operating current for conventional and damage-resistant lasers. The discontinuity of the driving current is caused by the interruption to the accelerated life test that occurs when other samples in the thermostatic chamber are changed (Modified from Takeshita et al. [17])

either the $n\text{-}OBIC_{ridge}$ intensity or the maximum power level, even after 3,000 h. The maximum power is limited by thermal saturation and is as high 400 mW or more. The constant values mean that there is no degradation at the AR facet of the damage-resistant lasers. Hence, this OBIC technique is useful for detecting facet degradation.

The aging time dependence of the relative increase in the operating current for the conventional and damage-resistant lasers is shown in Fig. 3.12. For both lasers, $t^{0.5}$ deterioration appeared after the first-stage degradation. As regards the reaction model [13], a time exponent of 0.5 means that the operating current increase is mainly determined by defect-induced diffusion. On the other hand, the relative increase in operating current for the conventional lasers is about one order of magnitude higher than that for the damage-resistant lasers. Since there was no degradation at the AR facet of the damage-resistant lasers, the difference in the relative increase in operating current must be caused by defects around the AR facet. Consequently, the defects around the AR facet govern the COD level as well as the long-term stability.

### *3.3.2 Facet-Vicinity Degradation*

Sudden failure due to the occurrence of COD during operation is suppressed by suppressing facet degradation. However, even though the deterioration that leads to COD is suppressed, sudden failures frequently occurred in aging tests. In this section, sudden failure independent of facet degradation is analyzed by monitoring the OBIC and by using plane-view transmission electron microscopy around the AR facet.

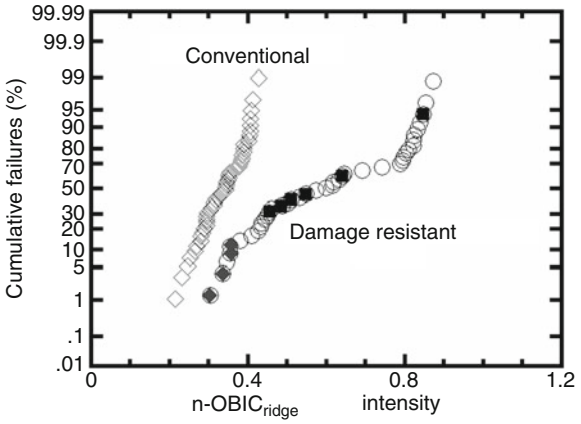



**Fig. 3.13** Lognormal *n-OBIC*$_{ridge}$ intensity statistics for several failed LDs with and without facet treatment (Modified from Takeshita et al. [2])

### 3.3.2.1 Degradation Characteristics

We used a 0.98-μm InGaAs/GaAs strained QW laser with a damage-resistant coating layer. The structure was the same as that described in Sect. 3.3.1.1 except for the reflectivity of the front facet, which was reduced to 2.5%. All the aging tests were carried out at 220 mW and 50°C.

A damage-resistant laser is a laser that suppresses facet degradation. However, sudden failures occurred during the aging tests. The lasers that suddenly failed had a particular characteristic. When an additional aging test was carried out on the failed lasers, the laser operated at a constant power for a few minutes and then suddenly failed again [23]. Unlike the COD failure, this failure was caused by thermal runaway.

Figure 3.13 shows the lognormal statistics of the *n-OBIC*$_{ridge}$ intensity for the LDs with damage-resistant and conventional AR coatings that experienced sudden failure. For conventional lasers, the mean and standard deviations of the *n-OBIC*$_{ridge}$ intensity characteristics were 0.34 and 0.06, respectively. These conventional lasers exhibited COD degradation at the AR facet and did not have an *n-OBIC*$_{ridge}$ intensity distribution of more than 0.45. By contrast, for the damage-resistant lasers, the mean and standard deviations of the *n-OBIC*$_{ridge}$ intensity characteristics were 0.6 and 0.2, respectively. The difference in the *n-OBIC*$_{ridge}$ intensity distribution indicates that the damage-resistant and conventional lasers had different degradation modes. Since the *n-OBIC*$_{ridge}$ intensity was measured from the AR facet, the deterioration region must exist in the vicinity of the AR facet.

### 3.3.2.2 TEM and OBIC Analyses

Three degraded lasers were analyzed to determine their failure modes. Figure 3.14 shows the AR-facet digital OBIC scans for (a) LD 1 with an *n-OBIC*$_{ridge}$ intensity of 0.55, (b) LD 2 with an *n-OBIC*$_{ridge}$ intensity of 0.90, and (c) LD 3 with an

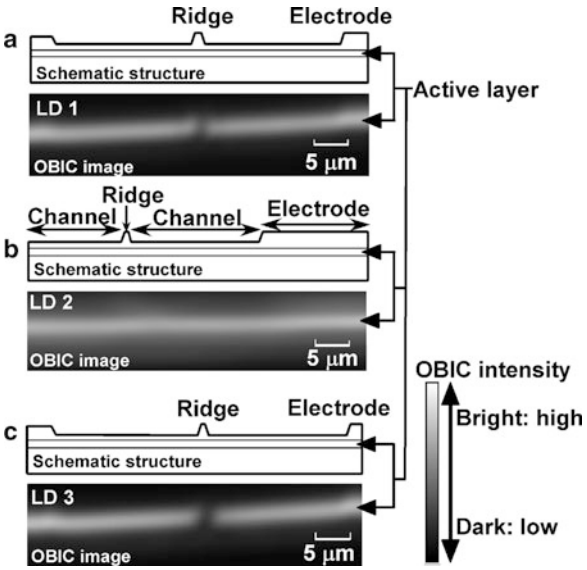



**Fig. 3.14** AR-facet digital OBIC scans for (**a**) LD 1 with an *n-OBIC*$_{ridge}$ intensity of 0.55, (**b**) LD 2 with an *n-OBIC*$_{ridge}$ intensity of 0.90, and (**c**) LD 3 with an *n-OBIC*$_{ridge}$ intensity of 0.35 (Modified from Takeshita et al. [2])

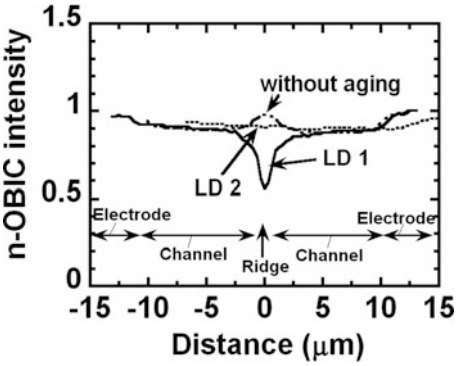



**Fig. 3.15** Profiles of *n-OBIC* intensity at active layer for lasers that suddenly failed (LDs 1 and 2) and for a laser that operated normally (Modified from Takeshita et al. [2])

*n-OBIC*$_{ridge}$ intensity of 0.35. The black region at the center of the horizontal white line found with LDs 1 and 3 corresponds to the deterioration of the active layer under the ridge. However, there is no evident black region at the center of the horizontal white line with LD 2. Here, profiles of the *n-OBIC* intensity in the active layer for lasers that suddenly failed (LDs 1 and 2) and for a laser that operated normally are shown in Fig. 3.15. The *n-OBIC* intensity under the ridge of LD 2 is almost the same as that under the channel. There is some deterioration around the AR facet because there is no *n-OBIC*$_{ridge}$ intensity hump.

Plane-view TEM images for (a) LD 1, (b) LD 2, and (c) LD 3 are shown in Fig. 3.16.

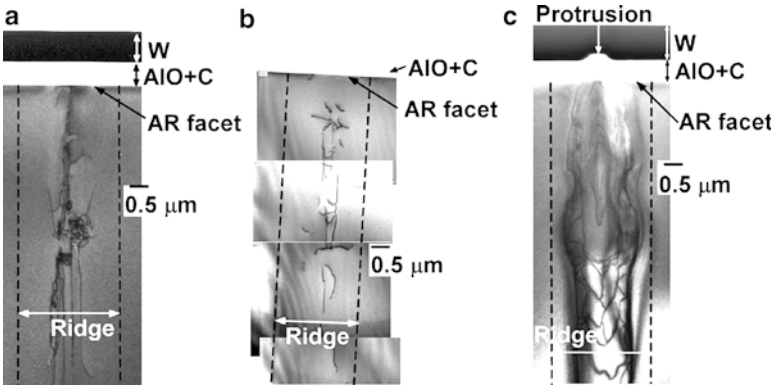



**Fig. 3.16** Plane-view TEM image for (**a**) LD 1, (**b**) LD 2, and (**c**) LD 3 (Modified from Takeshita et al. [2])

First, for LD 1 with an *n-OBIC*$_{ridge}$ intensity of 0.55, the plane-view TEM image indicates both extended <110> direction dislocations and a tangle of dislocations around the center of the active layer within 10 μm of the AR facet. On the other hand, there are no COD-failure symptoms on the facet, which is unlike conventional COD caused by AR facet degradation. Then, the reduction of *n-OBIC*$_{ridge}$ mainly results from the dislocations around the AR facet.

Second, for LD 2 with an *n-OBIC*$_{ridge}$ intensity of 0.90, the facet coating film is flat and the facet exhibits no COD-failure symptoms. On the other hand, a dislocation formed in the active layer around the AR facet. The reduction in the *n-OBIC*$_{ridge}$ intensity was caused by this dislocation. The LD 2 dislocations have a smaller distribution in the ridge waveguide than those of LD 1. Dislocations that extended to the AR facet were measured as an OBIC intensity reduction because they involve a large number of nonradiative recombination centers. Then, the *n-OBIC*$_{ridge}$ intensity responds to the dislocations around the facet.

Third, for LD 3 with an *n-OBIC*$_{ridge}$ intensity of 0.35, there was a projection in the $AlO_x$ coating of the AR facet and the facet was destroyed. Furthermore, there was a melted and solidified region in the active layer around the AR facet, and COD occurred around the facet.

The filled diamonds (◆) and squares (■) in Fig. 3.13 show results for degraded lasers with both COD and dislocations and for those with only dislocations around the AR facet, respectively. The degraded lasers with only dislocations and those with both COD and dislocations generally had *n-OBIC*$_{ridge}$ intensities of more than 0.4 and less than 0.4, respectively. For the COD-failure lasers with a conventional AR coating, the *n-OBIC*$_{ridge}$ intensity was usually less than 0.4. The lasers with an *n-OBIC*$_{ridge}$ intensity of less than 0.4 were therefore presumed to fail due to COD deterioration.

#### 3.3.2.3 Mechanism of Facet-Vicinity Degradation

The LDs with the damage-resistant AR coating had a wide $n\text{-}OBIC_{ridge}$ intensity distribution of 0.3 to 0.9, and dislocations or COD were generated around the AR facet. However, the damage-resistant lasers showed no COD level reduction because the laser facets did not deteriorate. We therefore consider the degradation mechanism to be different from the conventional COD deterioration that causes a defect density increase at the AR facet. As a result, we can explain the deterioration as follows. Dislocations were first generated around the AR facet. They extended to the AR facet and grew around it, and finally COD occurred at the AR facet. For lasers with only dislocations, the dislocations absorbed the light emitted by the lasers. The heat thus generated increased the device temperature, and finally we presume that the laser failed suddenly as a result of thermal runaway.

Around the AR facet region, there were both a tangle of dislocations [24] and <110> direction-extended dislocations as shown in Fig. 3.16a and b. Since dislocations act as nonradiative recombination centers, heat is generated around the AR facet region. Then once dislocations are generated, the temperature increases and a network of dislocations grows in the active region of the LD as a result of the dislocation climb. In addition, dislocation glide is generated by thermal stress caused by temperature increase. Consequently, suppressing dislocation generation around the AR facet is important as regards preventing sudden failure.

The optical power density around the AR facet is the highest in the laser cavity. This mainly generates heat at the AR facet, which increases the surrounding temperature [25]. In addition, the temperature increase around the AR facet increases in the current density around the AR facet. The temperature rise is accelerated locally at the degradation around the AR facet. As a result, we presume that the region around the AR facet reaches the wear-out failure period of the bathtub curve [13], but the region other than that around the AR facet remains in the stable random failure period of the bathtub curve. During the wear-out failure period, certain defects caused by the recombination enhanced defect reaction (REDR) mechanism [26] make it possible for dislocations to occur and grow around the AR facet region under high optical power density, high temperature, and high thermal stress conditions. We therefore believe that it is important to suppress the degradation around the AR facet.

## 3.4 Wear-Out Failure

In the second-stage degradation, some operating lasers exhibit complicated behavior. In Sects. 3.4.1 and 3.4.2, respectively, we use the OBIC technique to analyze lasers that degrade with the properties of conventional $t^{0.5}$ deterioration and deviations from $t^{0.5}$ deterioration.

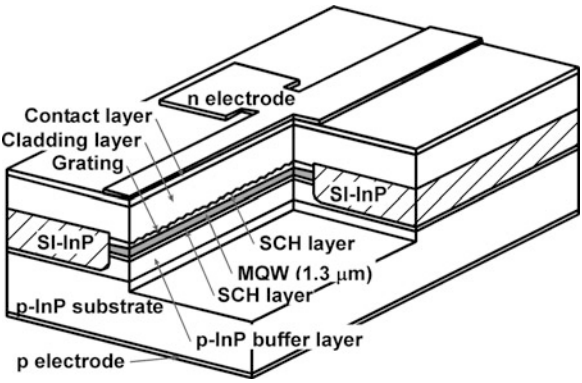



**Fig. 3.17** Schematic diagram of the cross-sectional structure of a 1.3-μm DFB laser

## *3.4.1 $t^{0.5}$ Deterioration*

### 3.4.1.1 Degradation Behavior

High-temperature characteristic InGaAsP/InP strained QW lasers do not suffer from sudden failure but gradually degrade with a current increase of $t^{0.5}$ under a constant power condition. We employ directly modulated 1.3-μm distributed feedback (DFB) lasers, which are promising candidates for use as inexpensive light sources in metropolitan areas as well as in local access optical networks [27]. Figure 3.17 is a schematic diagram of the cross-sectional structure of a 1.3-μm DFB laser with a semi-insulating buried heterostructure (SIBH) using Ru-doped InP. The active layer consists of strained 1.3-μm multiple quantum wells (MQW) composed of InGaAsP. This layer is sandwiched between 1.1-μm SCH layers also composed of InGaAsP. A grating was formed on the upper SCH layer. After growing an n-type InP layer, a mesa structure was formed, and current blocking layers (Ru-doped SIBH-InP) [28] were selectively grown on both sides of the mesa. Finally, n-type InP cladding and contact layers were grown. The laser was cleaved to a length of 200 μm. The front and rear facets were coated with AR and HR layers, respectively. The laser chip was mounted on AlN heat sink, and the entire structure was then mounted on a 5.6-mm TO can.

The operating current ($I_{op}$) change with aging time ($t$) is given by [13]

$$\frac{(I_{op} - I_0)}{I_0} = At^m \exp\left(\frac{-E_a}{kT}\right) \tag{3.4}$$

where $I_0$ is the initial operating current, $A$ is a constant, $m$ is the time exponent, $E_a$ is the activation energy, $k$ is Boltzmann's constant, and $T$ is the absolute temperature. Figure 3.18 shows the aging time dependence of the relative increase in driving current at an ambient temperature of 85°C with a constant output power of 9 mW. These unscreened lasers operated very stably for more than 8,500 h. The degradation of the operating current increase has two stages as regards the $t^1$ and $t^{0.5}$

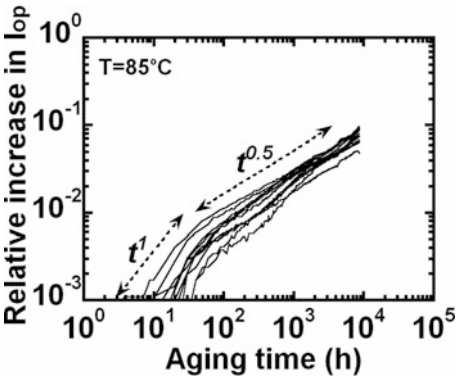



**Fig. 3.18** Aging time dependence of the relative increase in driving current at an ambient temperature of 85°C (Modified from Takeshita et al. [30])

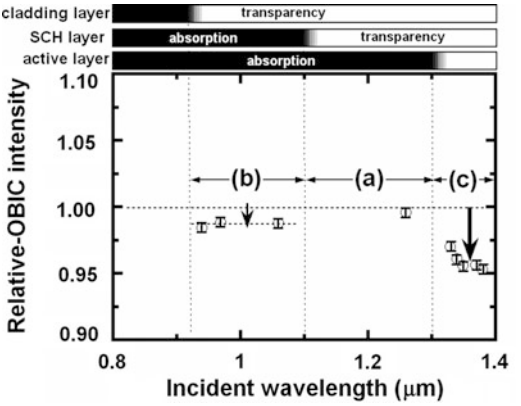



**Fig. 3.19** Wavelength dependence of the relative OBIC intensity (Modified from Takeshita et al. [30])

deterioration characteristics, which resembles the degradation of a conventional laser with a p- and n-type InP buried heterostructure [13, 29]. The slope of the operating current increase is approximately 0.5 after about 60 h of aging. In terms of the reaction model [13], this value means that the operating current increase is mainly determined by defect-induced diffusion.

### 3.4.1.2 SCH Layer Degradation

The peak of the OBIC intensity was normalized at 1.0 prior to aging in order to indicate the degree of laser degradation. The relative OBIC intensity exhibits severe degradation as it decreases below 1.0. The wavelength dependence of the relative OBIC intensity is shown in Fig. 3.19. For light at a wavelength of 1.26 μm, the relative OBIC intensity is approximately 1.0 and there is little degradation in the active layer. On the other hand, for light at wavelengths of 1.06, 0.97, and 0.94 μm, the relative OBIC intensity is less than 1.0 and there is degradation in the active and SCH layers [30, 31]. The results for the two types of light indicate the degradation in the SCH layer, which means that some defects diffuse from the SCH layer to the active layer.

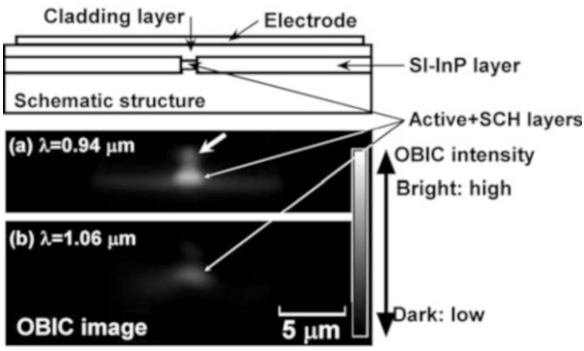



**Fig. 3.20** Digital OBIC scans of the AR facet at wavelengths of (**a**) 0.94 and (**b**) 1.06 μm (Modified from Takeshita et al. [30])

Next, we concentrate on the change in the OBIC image at several wavelengths. Figure 3.20 shows digital OBIC scans of the AR facet at wavelengths of (a) 0.94 and (b) 1.06 μm. The 0.94- and 1.06-μm wavelength lights are fundamentally absorbed in the active and SCH layers. However, an OBIC image is obtained around the upper cladding layer at a wavelength of 0.94 μm (arrow) in Fig. 3.20a but not at a wavelength of 1.06 μm. The 0.94-μm wavelength is only about 20 nm longer than the band gap (1.35 eV) of the InP layer [32]. This is attributed to the spread of the InP band edge, which means that the quality of the upper InP cladding layer is inferior to that of the InP substrate. Some defects were probably generated in the upper InP cladding layer as it was grown. Therefore, we conclude that the epitaxial quality of the upper cladding layer should be improved to suppress the gradual degradation.

### 3.4.1.3 Interior Degradation

We analyzed the interior degradation of a laser waveguide by monitoring the OBIC at several wavelengths. Incident light from the AR facet passes through the waveguide of the laser. We use a wavelength longer than that of the band edge of the active layer to detect the interior degradation of a degraded laser [11].

Figure 3.21 shows a digital OBIC scan of the AR facet at a wavelength of 1.38 μm for a laser that had undergone 8,500 h of aging at 85°C. The incident light wavelength is 80 nm longer than the PL peak wavelength in the active layer ($\Delta\lambda$). In the laser waveguide, the incident light is absorbed in the semitransparent active layer, and a photo current is generated. The bright region indicates a high OBIC intensity, which is obtained in the active region. OBIC scan images were also obtained in the same region at wavelengths of 1.37, 1.35, 1.34, and 1.33 μm.

As shown in Fig. 3.19, for light at wavelengths of more than 1.30 μm, the relative OBIC intensity is less than 1.0. This means that there is degradation in the active layer within the penetration distance. These lights have a penetration distance of more than 10 μm. However, for a 1.26-μm wavelength light, the relative OBIC intensity was approximately 1.0, and there was little degradation in the active layer

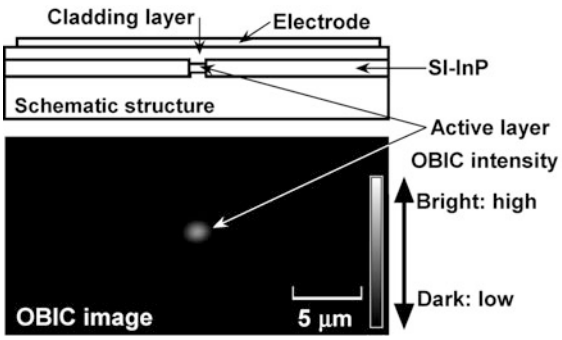



**Fig. 3.21** Digital OBIC scan of the AR facet at a wavelength of 1.38 μm (Modified from Takeshita et al. [30])

within approximately 10 μm (penetration distance) from the AR facet. Therefore, we confirmed that the degradation region of the $t^{0.5}$ deterioration in Fig. 3.18 is located in the active layer but not in the vicinity of the AR facet. Since the DFB laser power density decreases nearer to the AR facet [33], the degradation appears to relate to the power density.

## *3.4.2 Deviation from $t^{0.5}$ Deterioration*

As regards second-stage degradation, we analyze $t^{0.5}$-deterioration-deviation devices in this section.

### 3.4.2.1 Degradation of Highly Strained InAs/InP MQW

We used a 2.3-μm wavelength InAs/InP highly strained DFB laser. This mid-infrared laser has attracted attention as a promising light source for trace gas monitoring applications [34]. Since the laser operates under a low bias voltage of less than 1 V, there is little current in its p-/n-type InP junction. This makes it possible to analyze laser degradation easily.

The epitaxial structure around InAs/InP highly strained multi-quantum wells is shown in Fig. 3.22. The active layer consists of four 5-nm-thick strained InAs wells and 20-nm-thick InGaAs barriers with a lattice matched to the InP substrate [35], which was sandwiched between SCH layers. The lattice mismatch strain of an InAs layer on an InP substrate is as high as 3.2%. The upper and lower SCH layers were 100-nm-thick InGaAsP layers with a band-gap wavelength of 1.3 μm, and the guide layer on the upper SCH layer was a 100-nm-thick InGaAsP layer with a band-gap wavelength of 1.1 μm. First-order grating patterns were formed on this guide layer. After a p-type InP layer had been grown, a 1.5-μm-wide mesa structure was formed [36], and current blocking layers (p- and n-type InP buried heterostructures) were selectively grown on both sides of the mesa. Finally, p-type InP cladding and contact layers were grown. The laser was cleaved to a length of 900 μm.

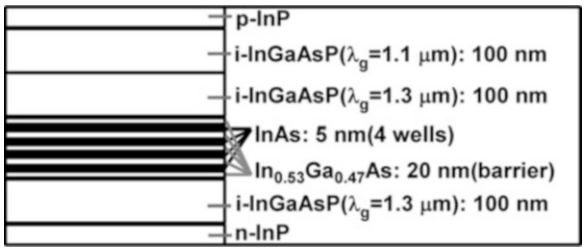



**Fig. 3.22** Epitaxial structure around InAs/InP highly strained multi-quantum wells

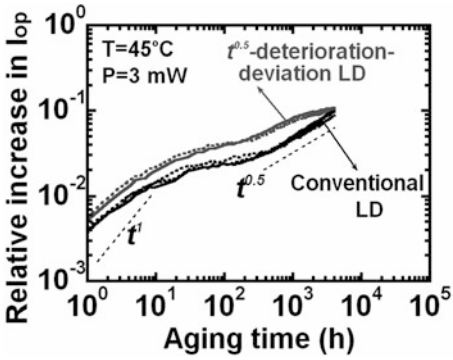



**Fig. 3.23** Aging time dependence on operating current increase ratio at an ambient temperature of 45°C (Modified from Takeshita et al. [39])

The front and rear facets were coated with AR ($TiO_2/SiO_2$) and HR ($SiO_2/TiO_2$) layers, respectively. The laser chip was mounted on AlN heat sink, and the whole structure was then mounted on a 5.6-mm TO can.

The log–log plot of the operating current increase ratio $[(I_{op}-I_0)/I_0]$ over time at an ambient temperature of 45°C and with a constant output power ($P_o$) of 3 mW is shown in Fig. 3.23. These unscreened lasers operate very stably for more than 5,000 h. The first-stage degradation of the $t^1$ deterioration and the subsequent small current increase with a time exponent of less than 0.5 appear with 100–300 h of aging. After that, the second stage of the $t^{0.5}$ deterioration appears. The two main stages are similar to the degradation modes of 2.1-μm lasers [37] and conventional telecommunication lasers [13]. A time exponent of 0.5 means that the operating current increase is mainly determined by defect-induced diffusion [13]. Then, the main degradation mechanism in the DFB lasers with highly strained InAs quantum wells [38] is dominated by a diffusion process as found with conventional telecommunication lasers.

### 3.4.2.2 InP Layer Degradation

The slope efficiency was constant for 100–300 h, which means that there was no change in the internal loss in the laser waveguide. On the other hand, the threshold current increased from the onset of aging. We presume that the pass current around

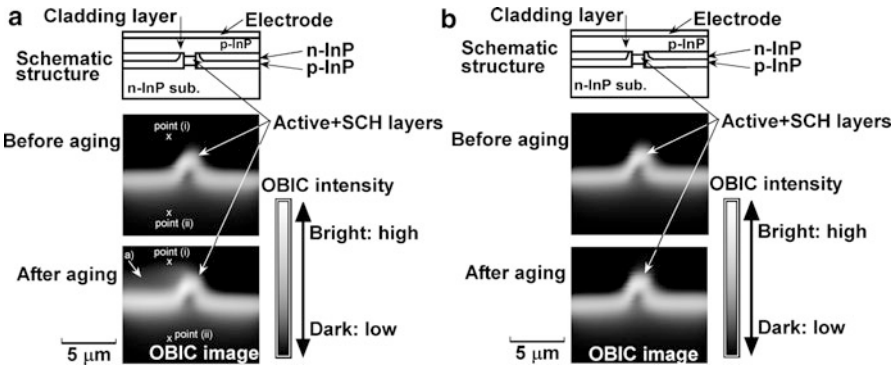



**Fig. 3.24** Digital OBIC scans at a wavelength of 0.85 μm and vertical OBIC profiles for (**a**) a $t^{0.5}$-deterioration-deviation device before and after aging and (**b**) a conventional device before and after aging (Modified from Takeshita et al. [39])

the p-type InP regrowth layer adjacent to the mesa sidewall in 100–300 h is increased by the conventional first-stage degradation.

In addition, lasers that are greatly degraded during the first-stage degradation tend to exhibit a large deterioration in aging time during second-stage degradation. We consider that the p-type InP regrowth layer adjacent to the mesa sidewall affects both the first- and second-stage degradation.

As shown in Fig. 3.23, in the second-stage degradation, the operating current was increased by two different characteristics [39]: namely those of LDs with and without deviation from the $t^{0.5}$ deterioration. We concentrate on $t^{0.5}$-deterioration-deviation and conventional LDs, which are typical lasers with and without deviation from the $t^{0.5}$ deterioration, respectively.

We used a wavelength of 0.85 μm for OBIC measurements to analyze the degradation of the p-type InP regrowth layer adjacent to the mesa sidewall. The measurements were conducted for both lasers before and after an aging time of 2,350 h. Digital OBIC scans for the $t^{0.5}$-deterioration-deviation and conventional LDs are shown in Fig. 3.24a and b, respectively. The bright region indicates a high OBIC intensity, which was mainly obtained around the pn junction in the mesa and the pn junction between the p-type InP buried layer and the n-type InP substrate. After aging, the peaks of the active and SCH layers increase to 1.25 and 1.10 for the $t^{0.5}$-deterioration-deviation and conventional LDs, respectively.

The dark region in the digital OBIC scans indicates a low OBIC intensity. This is because electron–hole pairs generated by the light absorption are extinguished by carrier recombination [40]. For the $t^{0.5}$-deterioration-deviation LD (Fig. 3.24a), the OBIC intensity after aging increases largely in the left region above the pn junction between the p-type InP buried layer and the n-type InP substrate (white arrow a).

Figure 3.25 shows the vertical profiles (i)–(ii) across the left side 2 μm from the center of the mesa in Fig. 3.24a. The left shoulder of the profile before and after aging

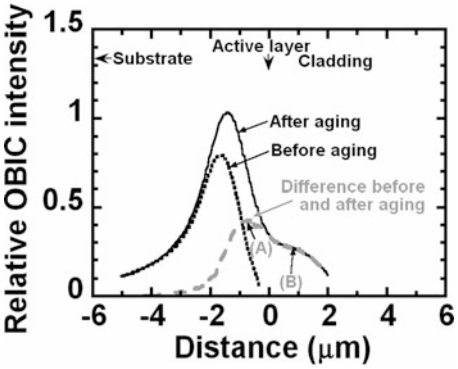



**Fig. 3.25** Vertical profiles (i)–(ii) across the left side 2 μm from the center of the mesa in Fig. 3.24a. The short broken, solid, and long broken lines are the vertical profiles of the relative OBIC intensity before aging, the relative OBIC intensity after aging, and the relative OBIC intensity difference before and after aging, respectively (Modified from Takeshita et al. [39])

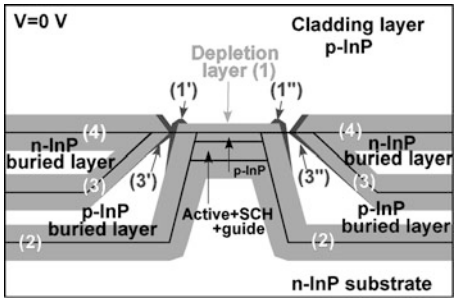



**Fig. 3.26** Schematic depletion layer of a laser at a bias voltage of 0 V (Modified from Takeshita et al. [39])

remains unchanged after aging, whereas the right shoulder exhibits a large change. The long broken line is the vertical profile of the relative OBIC intensity difference before and after aging. The profile mainly consists of two lobes. The peaks of these lobes are located at distances of about −1 and 0.5 μm (denoted as (A) and (B) in Fig. 3.25), which correspond mainly to the depletion layers between the p- and n-type InP buried layers and between the n-type InP buried and p-type InP cladding layers, respectively.

The depletion layer of a laser at a bias voltage of 0 V is shown schematically in Fig. 3.26. The gray regions are schematic depletion layers. When the peak of the incident light for the OBIC measurement is placed in the active layer of the mesa, some of the incident light illuminates depletion layer (3) between the p- and n-type InP buried layers as well as depletion layer (4) between the n-type InP buried and p-type InP cladding layers. However, before aging, the OBIC was not obtained around depletion layers (3) and (4), as shown in Fig. 3.24a and b. This is because the

n-type InP buried layer and the n-type InP substrate are isolated by the p-type InP buried layer, and the electrons generated in depletion layers (3) and (4) do not reach the n-type InP substrate. On the other hand, an additional OBIC generated around depletion layers (3) and (4) is obtained after aging. This means that an electrical connection is realized between the n-type InP buried layer and the n-type InP substrate after aging. The electrical connection is achieved in the p-type InP buried layer by attaching the enlarged depletion layers (3) + (3′) and (1) + (1′) after aging. This indicates that the carrier concentration [9] decreased around the p-type InP buried layer adjacent to the mesa sidewall after aging.

The p-type (Zn dopant) InP buried layer adjacent to the mesa sidewall is exposed to a high current density and a high laser power density during aging. Some defects are mobile at a much lower temperature (about 100°C) than impurities [41], and diffused defects that reach the active layer affect the relative increase of the operating current. We believe that the carrier concentration around the p-type InP buried layer adjacent to the mesa sidewall is decreased by the formation of defect–impurity complexes, which cause deep levels [42–45]. Then, we presume that the carrier concentration around the p-type buried layer adjacent to the mesa sidewall is decreased by the defused defects.

Finally, for the $t^{0.5}$-deterioration-deviation LD, we attempted to determine why the operating- and threshold-current increases are suppressed in the second-stage degradation. Since the depletion layer (1) + (1′) is extinguished by an applied voltage under lasing operation, we consider that the current suppression caused by the enlarged depletion layer (3) + (3′) is improved compared with that caused only by the depletion layer (3). As a result, we believe that the pass current around the p-type InP buried layer adjacent to the mesa sidewall is decreased by the enlarged depletion layer (3) + (3′), which suppresses increases in the operating and threshold currents. Consequently, long-term stability is dominated not only by the defects at the active layer but also by the defects around the p-type InP buried layer adjacent to the mesa sidewall.

## 3.5 Surface Degradation

For InGaAs/InP APDs, an increase in dark current causes a deterioration in the signal-to-noise ratio of optical receivers [46], which dominates device lifetime. The APD degradation region is clarified by using the optical-beam-induced current imaging technique, and we confirmed that the dark-current increase is related to a leakage path near the edge of the InGaAs absorption layer. Highly reliable APDs with a mesa structure can be achieved by preventing dark-current degradation at the mesa edge.

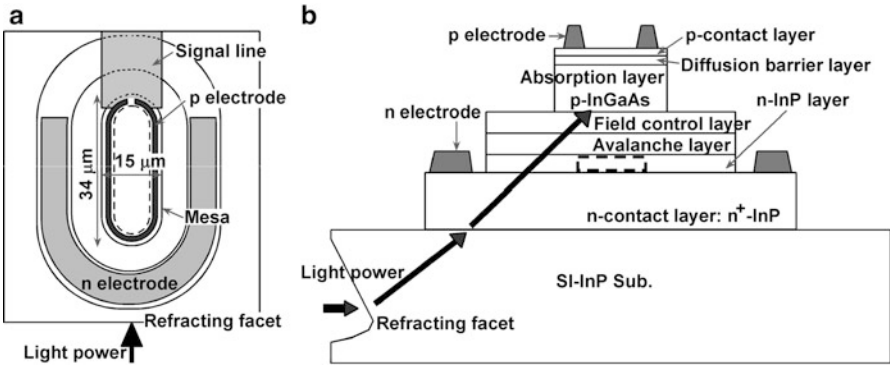



**Fig. 3.27** Schematic diagram of the structure of the InGaAs/InP avalanche photodiode: (**a**) top view and (**b**) cross-sectional view (Modified from Takeshita et al. [55])

### *3.5.1 Degradation Behavior of APD*

Figure 3.27 shows the schematic structure of an InGaAs/InP APD operated in the electron-injection mode. Epitaxial layers were grown on a semi-insulating InP substrate by low-pressure metal-organic vapor phase epitaxy. The APD consists of a Zn-doped p-type InGaAs absorption layer, a Zn-doped InP field-control layer, and an undoped InP avalanche layer (200 nm) [47]. This absorption layer has a graded doping profile, from $2 \times 10^{17}$ to $3 \times 10^{18}$ $cm^{-3}$, which can reduce the carrier transit time by controlling the quasifield of the p-type InGaAs [48]. A refracting facet for edge illumination is used to achieve a high responsivity because the p-type InGaAs absorption layer (570 nm thick) is relatively thin compared with conventional APDs with surface illumination. The stadium-shape is 34 μm long and 15 μm wide. The edge-illuminated refracting-facet structure [49] was formed by wet chemical etching and coated with antireflection $TiO_2/SiO_2$ (reflectivity: less than 0.1%). A cleaved APD chip was mounted on AlN heat sink, and then the whole structure was mounted on a 5.6-mm TO can.

An aging test was performed at 200°C under a constant bias current of 100 μA. The test was interrupted several times to measure the dark current–voltage ($I$–$V$) characteristics at 85°C. We define the breakdown voltage ($V_b$) as the applied voltage giving a dark current of 100 μA. The breakdown voltage was also monitored during aging. The aging time dependence of the dark current at 0.9 $V_b$ for the APDs is shown in Fig. 3.28a. There are two increases in the dark current: the first continues from about 2.5 h, and the second occurs at about 100 h. The aging time dependence of $V_b$ in the stress at 200°C is shown in Fig. 3.28b. The $V_b$ value at 200°C increases from 10 to 100 h for the APD. On the other hand, the $V_b$ value at 200°C begins to decrease after about 100 h, which corresponds to the time of the second dark-current increase in Fig. 3.28a. It was found that the dark current at 200°C after about 100 h is beyond 100 μA before avalanche breakdown.

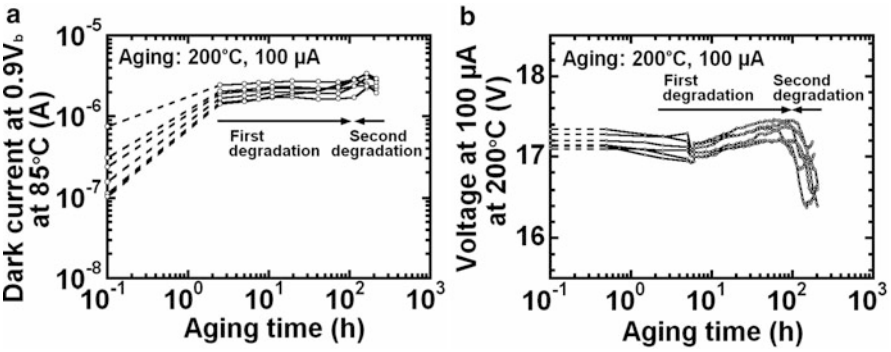



**Fig. 3.28** Long-term aging test results for APDs stressed at 200 °C. (**a**) Aging time dependence of the dark current at 0.9 $V_b$ measured at 85 °C. (**b**) Aging time dependence of $V_b$ measured at 200 °C. The number of devices ($n$) is 6 (Modified from Takeshita et al. [55])

## *3.5.2 Degradation Analysis*

### 3.5.2.1 First Degradation in Dark Current

The photocurrent appears abruptly at a reach-through voltage, and this was 11.9 V for the APD without aging. This voltage is that at which the carriers start to flow into the avalanche layer beyond the potential barrier produced by the field-control layer near the hetero-interface. We examine the first degradation in dark current for an APD aged for 19.5 h at 200°C. The change in the *I–V* characteristics of the APD at various input optical powers is shown in Fig. 3.29. Note that it is not a gentle *I–V* curve from 10.0 to 12.0 V but an additional current increase that occurred from10.0 to 11.5 V due to the first degradation after an aging time of 2.5 h.

A schematic drawing and an OBIC scan of the top surface of the degraded APD are shown in Fig. 3.30. A bright region (OBIC signal) is observed around the curve region on the signal-line side at 10.3 V, which is less than the reach-through voltage (11.9 V) in the initial device characteristics. This region corresponds to the region exposed to oxygen plasma (100 W) for 30 min. Provided that the nonradiative recombination rate increased in the absorption layer exposed to the plasma, the OBIC intensity should have decreased. However, after the aging, the OBIC intensity around the curve region on the signal-line side increased around the reach-through voltage. This may indicate that the plasma damage is not a simple bulk effect; that is, the damage produces a current path on the mesa surface.

When the OBIC analysis was performed from the surface, the incident light was restricted to the electrode window. Then, we analyzed the OBIC through the refracting facet of the APD to measure the deterioration at the mesa edge. The OBIC spot power of 4 μW is equivalent to the input power of 100–200 μW in Fig. 3.29. Figure 3.31 shows OBIC scans obtained around 11 V for APDs (a) without aging and with (b) 19.5- and (c) 114-h aging. The upper and lower figures are OBIC

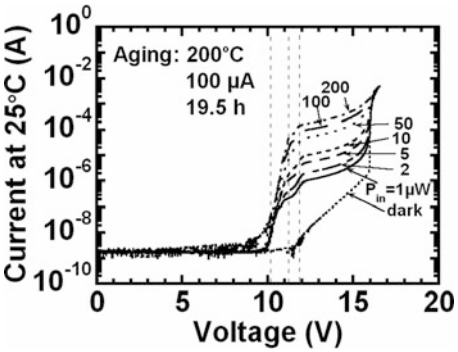



**Fig. 3.29** Change in current–voltage characteristics at various input powers after 19.5-h aging at 200°C (Modified from Takeshita et al. [55])

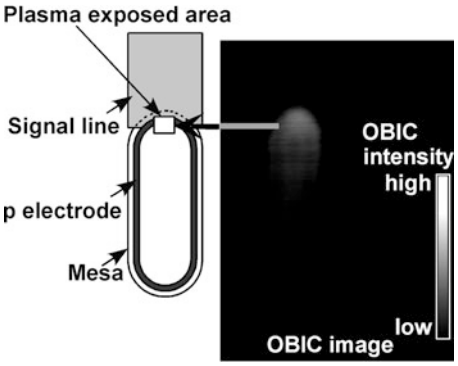



**Fig. 3.30** Schematic drawing and OBIC scan of the surface for the degraded APD after the 19.5-h aging at 200°C (Modified from Takeshita et al. [55])

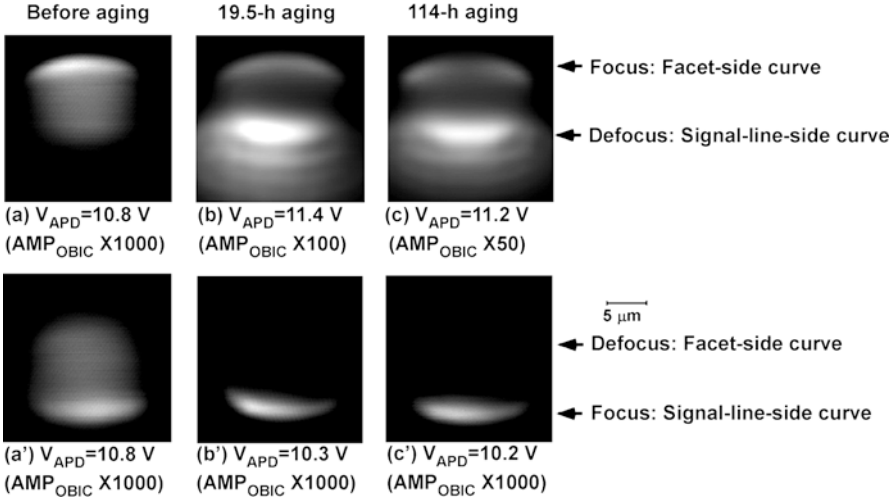



**Fig. 3.31** OBIC scans obtained around 11 V for APDs (**a**) without aging and with (**b**) 19.5- and (**c**) 114-h aging (Modified from Takeshita et al. [55])

scans obtained when the incident beam was focused on the refracting-facet side and the signal-line side, respectively. Before the aging, the OBIC scans were already not uniform and the signal was higher in the curve regions on both sides at 10.8 V. This indicated that there was degradation even for the nonaged device around the mesa edge [50, 51]. On the other hand, the OBIC scan appeared more intensely in the signal-line-side-curve region at 10.3 V for the APD even for 19.5 h of aging. According to the OBIC scans observed from the top surface (Fig. 3.30) and the refracting facet, the first deterioration is mainly correlated with the degradation around the signal-line-side-curve region. The OBIC observation revealed that the reach-through voltage can be lowered locally from its initial value.

One explanation for this behavior would be a decrease in the ionized-Zn concentration in the field-control layer. However, such a decrease would lead to a rise in the local breakdown voltage, and the current concentration would be reduced around this region. If we assume that the local avalanche current accelerates the deterioration, this explanation is inconsistent. Then, we cannot assume that there is any degradation in the field-control layer with aging. Another possibility is that the dark-current path in the signal-line-side-curve region of the p-type InGaAs absorption layer is a surface current channel. It is considered that a change in the surface defect density can make the band profile bend in the degraded region [52–54].

#### 3.5.2.2 Second Degradation in Dark Current

The photocurrent increased around 11.3 V after the aging, as shown in Fig. 3.29. Figure 3.32 shows the correlation between OBIC scans and applied voltage for APDs (a) without aging and with (b) 19.5- and (c) 114-h aging. We focus on the doughnut-shaped OBIC image around 11.6 V after 19.5- and 114-h aging. The OBIC intensity at the mesa edge at 11.6 V after 114 h is about 30% higher than that after 19.5 h. Then, the second deterioration in the dark current $I_{0.9V_b}$ is related to the leakage current around the whole mesa edge. Although the second-degradation dark current is obscured by the main first-degradation dark current for 100 h, we consider that the degradation of the mesa edge has proceeded gradually for 100 h. The OBIC concentration suggests that there is a lateral current flow in the p-type InGaAs absorption layer toward the mesa edge.

Next, OBIC scans obtained around 12 V for APDs (a) without aging and with (b) 19.5- and (c) 114-h aging are shown in Fig. 3.33. The OBIC intensity inside the entire mesa at 11.9 V had hardly changed after the aging test, which means that there was no deterioration inside the mesa except at the edge.

The leakage current concentration at the mesa edge must be suppressed to prevent deterioration of the mesa APD. In Fig. 3.27, the n-type InP layer on the n-contact layer is replaced with an undoped InP layer, and Si ions are implanted in the long broken line region in the undoped InP layer, which is introduced to define the active area inside the periphery of the p-type InGaAs [47]. After treatment to activate the ion-implantation region, the InP avalanche, InP field-control, and InGaAs absorption layers are formed by regrowth. In this

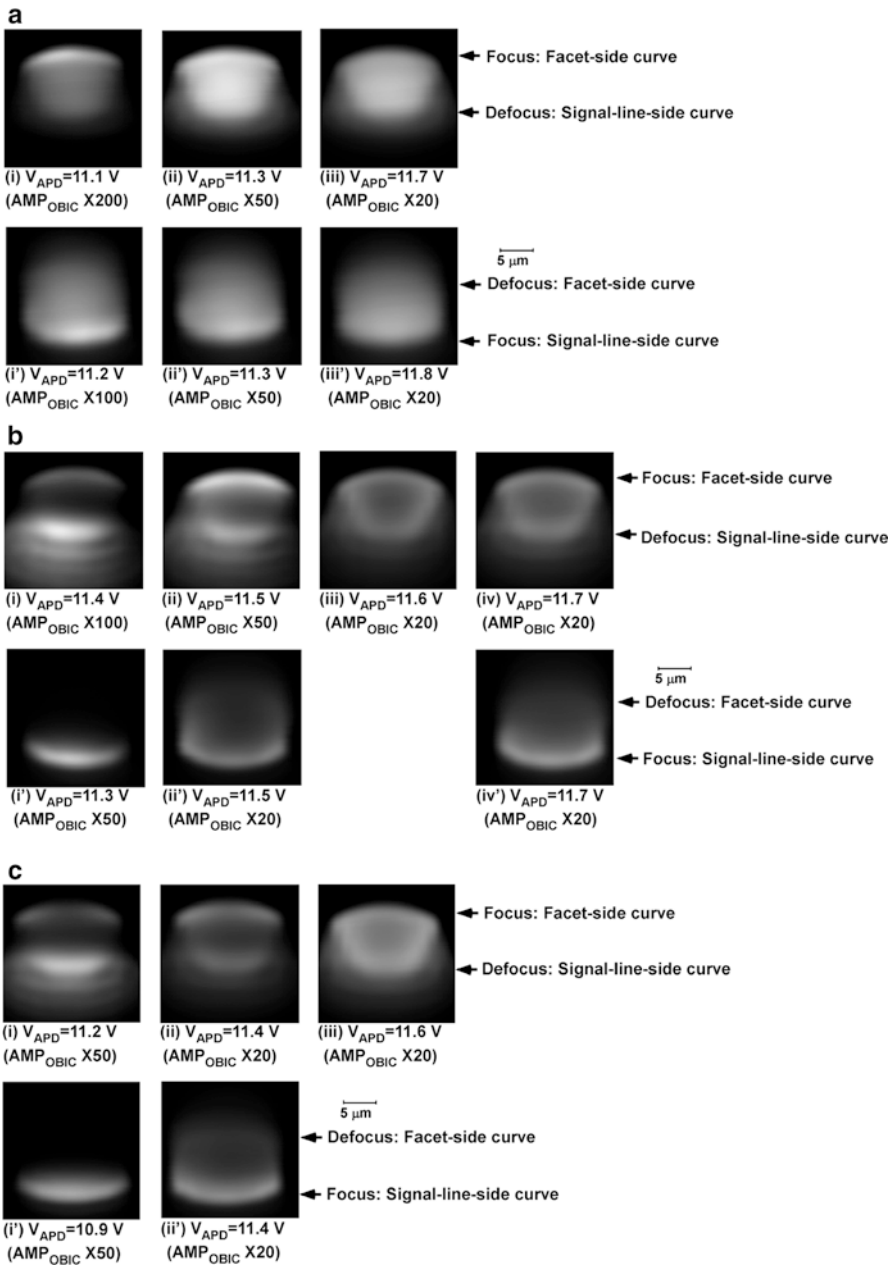



**Fig. 3.32** Correlation between OBIC scans and applied voltage for APDs (**a**) without aging, (**b**) with 19.5-h aging, and (**c**) with 114-h aging. The illuminated power density is the same for APDs at various aging times, even at an out-of-focus location (Modified from Takeshita et al. [55])

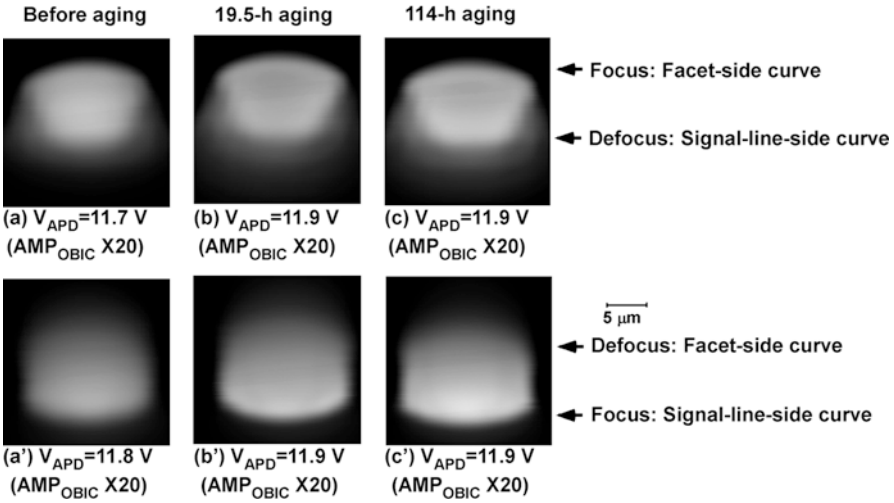



**Fig. 3.33** OBIC scans obtained around 11 V for APDs (**a**) without aging and with (**b**) 19.5- and (**c**) 114-h aging (Modified from Takeshita et al. [55])

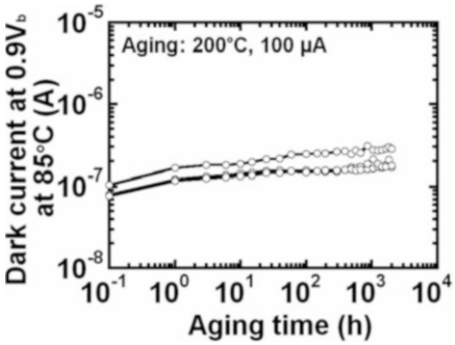



**Fig. 3.34** Aging time dependence of dark current at $0.9V_b$ measured at 85 °C (Modified from Takeshita et al. [55])

APD structure, the mesa edge of the InGaAs absorption layer remains current-blocked even at the breakdown voltage due to the much wider undoped InP layer underneath it. Furthermore, the APD surface was not directly exposed to the plasma. The aging time dependence of the dark current at 0.9 $V_b$ is shown in Fig. 3.34. The dark-current level and its increase rate are very low until 2,000 h [55], and the degradation mode is the wear-out mode. Consequently, highly reliable APDs with a mesa structure can be achieved by preventing dark-current degradation at the mesa edge.

## 3.6 Application to Planar-View OBIC Scan

Degradation around the facet is detected by the OBIC technique by using an incident light absorbed in only the active layer. The penetration distance is about 30 μm for a laser with a Γ of 0.03 and an $\alpha_0$ of $10^4$ $cm^{-1}$. Furthermore, degradation of the laser interior is detected by employing an incident light with a longer wavelength than the PL peak of the active layer.

On the other hand, when there is degradation in both the facet and laser interior, it is hard to detect degradation in the laser interior by using the incident light from the facet. Then, we employed planar-view observation to detect degradation in the waveguide. We removed part of the electrode by ablation etching [23, 56]. A planar-view OBIC scan clarified the location of some degraded regions in the waveguide.

Figure 3.35 shows OBIC and optical images obtained from the surface for a degraded laser when a 150-μm section of the electrode around the AR facet was removed by laser ablation. This degraded laser exhibited a 4-mA increase in the threshold current from its initial value and had an $n\text{-}OBIC_{ridge}$ intensity of 0.84. There are mainly two degradation regions in the OBIC image: one is the active layer under the ridge from the AR facet to 30 μm, and the other is that from 110 to 150 μm. The OBIC intensity of the active layer under the ridge from the AR facet to 30 μm is lower than that from 110 to 150 μm. This indicates that there is a large degradation in the active layer from the AR facet to 30 μm compared with that from 110 to 150 μm. This planar-view OBIC scan is useful for detecting the location of some degraded regions in the waveguide.

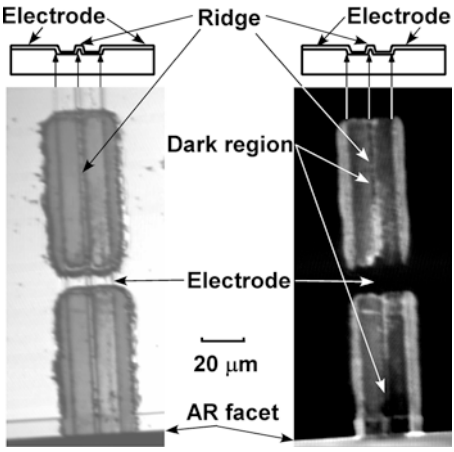



**Fig. 3.35** OBIC and optical images obtained from the surface of a degraded laser when a 150-μm section of the electrode around the AR facet was removed by laser ablation (Modified from Takeshita et al. [2])

## 3.7 Conclusion

By employing the optical-beam-induced current (OBIC) measurement technique, we have analyzed the sudden and wear-out failure of optical devices. The extent of the degraded regions is estimated using the relative OBIC intensity prior to aging. In addition, using the OBIC technique with lights at several wavelengths enables us to detect degradation in the facet, epitaxial layer, and device interior regions.

As a result, we found that increased defects due to the antireflection (AR) facet degradation govern both the long-term stability and the catastrophic optical damage (COD) level, and after suppressing the facet degradation, dislocations generated locally around the AR facet dominate sudden failure due to thermal runaway. Next, for a laser exhibiting wear-out failure, we confirmed that defects diffuse from the separate confinement heterostructure (SCH) layer to the active layer and the degradation region is located in the active layer but not in the vicinity of the AR (low power density) facet. Then, long-term stability is dominated not only by the defects at the active layer but also by the defects around the p-type InP buried layer adjacent to the mesa sidewall. Furthermore, we confirmed that the dark current of APDs is dominated by the leakage current concentration at the mesa edge as well as plasma damage. Finally, a planar-view OBIC scan is useful for detecting the location of some degraded regions. Thus, the OBIC technique is useful for analyzing degraded devices.

## References

1. J. Wallon, P. Devold, in *Proceedings of the European Symposium on Reliability of Electron Devices, Failure Physics and Analysis*, Bordeaux, France, 1995, p. 415
2. T. Takeshita, M. Sugo, T. Sasaki, Y. Tohmori, IEEE Trans. Electron Device. **53**(2), 211 (2006)
3. J. Bajaj, L.O. Bubulac, P.R. Newman, W.E. Tennant, P.M. Raccah, J. Vac. Sci. Technol., A **5**(5), 3186 (1987)
4. R.M. Cramer, W.R. Schade, R. Heiderhoff, L.J. Balk, R. Chin, Microelectron. Reliab. **38**(6), 936 (1998)
5. H. Yonezu, T. Yuasa, T. Shinohara, T. Kamejima, I. Sakuma, Jpn. J. Appl. Phys. **15**(12), 2393 (1976)
6. G.B. Stringfellow, P.E. Greene, J. Appl. Phys. **40**(2), 502 (1969)
7. H.C. Casey Jr., M.B. Panish, *Heterostructure Lasers* (Academic, New York, 1978)
8. H. Kressel, J.K. Butler, *Semiconductor Lasers and Heterostructure LEDs* (Academic, New York, 1977)
9. S.M. Sze, *Semiconductor Devices* (Wiley, New York, 1969)
10. F.R. Bacher, J.S. Blakemore, J.T. Ebner, J.R. Arthur, Phys. Rev. B: Condens. Matter **37**(5), 2551 (1988)
11. T. Takeshita, R. Iga, M. Yamamoto, M. Sugo, Microelectron. Reliab. **47**(12), 2135 (2007)
12. H. Seshake, S. Yamamoto, Y. Matsuoka, NTT Tech. Rev. **4**(2), 66 (2006)
13. M. Fukuda, *Reliability and Degradation of Semiconductor Lasers and LEDs* (Artech House, Boston, 1991)
14. J. Ueda, Electrochem. Soc. **135**(1), 11C (1988)

15. T. Takeshita, H. Oohashi, in *Materials Research Society Symposium Proceedings*, ed. by O. Ueda, M. Fukuda, S.J. Pearton, E.L. Piner, P. Montangero, vol. 1195 (Material Research Society, Warrendale, 2009), p. 15
16. C.H. Henry, P.M. Petroff, R.A. Logan, F.R. Merritt, J. Appl. Phys. **50**(5), 3721 (1979)
17. T. Takeshita, M. Sugo, T. Nishiya, R. Iga, M. Fukuda, Y. Itaya, Microelectron. Reliab. **38**(6), 1211 (1998)
18. M. Fukuda, M. Okayasu, J. Temmyo, J. Nakano, IEEE J. Quantum Electron. **30**(2), 471 (1994)
19. T. Takeshita, M. Okayasu, S. Uehara, Jpn. J. Appl. Phys. **30**(6), 1220 (1991)
20. T. Takeshita, M. Okayasu, S. Uehara, IEEE Photon. Technol. Lett. **2**(12), 849 (1990)
21. C.W. Snyder, J.W. Lee, R. Hull, R.A. Logan, Appl. Phys. Lett. **67**(4), 488 (1995)
22. W.B. Joyce, R.W. Dixon, R.L. Hartman, Appl. Phys. Lett. **28**(11), 684 (1976)
23. T. Takeshita, M. Sugo, T. Sasaki, Y. Tohmori, IEEE Trans. Electron Devices **53**(2), 211 (2006)
24. A. Dietzel, A. Jakubowicz, R.F. Broom, in *Proceedings of the Institute of Physics Conference Series*, Oxford, no. 146, 1995, p. 583
25. E.E. Moser, J. Latta, Appl. Phys. **71**(10), 4848 (1992)
26. L.C. Kimerling, Solid-State Electron **21**(11/12), 1391 (1978)
27. H. Mawatari, M. Fukuda, S. Matsumoto, K. Kishi, Y. Itaya, J. Lightwave Tech. **15**(3), 534 (1997)
28. R. Iga, Y. Kondo, T. Takeshita, K. Kishi, M. Yuda, Electron. Lett. **42**(5), 280 (2006)
29. S.P. Sim, A.P. Skeats, M.R. Taylor, M. Hockly, D.M. Cooper, A.W. Nelson, W.J. Devlin, J.C. Regnault, in *Proceedings of the 14th European Conference and Exhibition on Optical Communication*, Brighton, UK, 1988, p. 396
30. T. Takeshita, M. Yamamoto, R. Iga, M. Sugo, Y. Kondo, K. Kato, IEEE Trans. Electron Device. **54**(8), 1852 (2007)
31. T. Ito, T. Takeshita, M. Sugo, T. Kurosaki, Y. Akatsu, K. Kato, J. Jpn. Appl. Phys. **47**(6), 4523 (2008)
32. P. Lautenschlager, M. Garriga, M. Cardona, Phys. Rev. B: Condens. Matter **36**(9), 4813 (1987)
33. M. Fukuda, *Optical Semiconductor Devices* (Wiley, New York, 1998)
34. H.K. Choi, *Long-Wavelength Infrared Semiconductor Lasers* (Wiley, Hoboken, 2004)
35. T. Sato, M. Mitsuhara, N. Nunoya, T. Fujisawa, K. Kasaya, F. Kano, Y. Kondo, IEEE Photon. Technol. Lett. **20**(12), 1045 (2008)
36. Y. Kondo, K. Kishi, M. Itoh, H. Oohashi, Y. Itaya, M. Yamamoto, *Proceedings of the 8th International Conference on Indium Phosphide and Related Materials* (Schwaebisch Gmuend, Germany, 1996), p. 384
37. T. Takeshita, T. Sato, M. Mitsuhara, Y. Kondo, M. Sugo, K. Kato, IEEE Trans. Electron Device. **54**(10), 2644 (2007)
38. T. Takeshita, T. Sato, M. Mitsuhara, Y. Kondo, H. Oohashi, IEEE Photon. Technol. Lett. **21**(13), 896 (2009)
39. T. Takeshita, T. Sato, M. Mitsuhara, Y. Kondo, H. Oohashi, IEEE Trans. Device Mater. Rel. **10**(1), 142 (2010)
40. K.T. Tan, S.H. Tan, S.H. Ong, in *Proceedings of the 1997 6th International Symposium on the Physical & Failure Analysis of Integrated Circuits*, Singapore, 1997, p. 296
41. M. Wada, M. Seko, K. Sakakibara, Y. Sekiguchi, Jpn. J. Appl. Phys. **28**(10), 1700 (1989)
42. G. Dlubek, O. Brümmer, F. Plazaola, P. Hautojärve, K. Naukkarinen, Appl. Phys. Lett. **46**(12), 1136 (1985)
43. N. Yamamoto, K. Uwai, K. Takahei, J. Appl. Phys. **65**(8), 3072 (1989)
44. J. Slotte, K. Saarinen, A. Salmi, S. Simula, R. Aavikko, P. Hautojärvi, Phys. Rev. B: Condens. Matter **67**(11), 115209–1 (2003)
45. T. Ikegami, K. Takahei, M. Fukuda, K. Kuroiwa, Electron. Lett. **19**(8), 282 (1983)
46. H. Kanbe, G. Grosskope, O. Mikami, S. Machida, IEEE J. Quantum Electron. **QE-17**(8), 1534 (1981)
47. Y. Hirota, S. Ando, T. Ishibashi, Jpn. J. Appl. Phys. **43**(3A), L375 (2004)

48. T. Ishibashi, T. Furuta, H. Fushimi, S. Kodama, H. Ito, T. Nagatsuma, N. Shimizu, Y. Miyamoto, IEICE Trans. Electron. **E83-C**(6), 938 (2000)
49. H. Fukano, Y. Muramoto, K. Takahata, Y. Matsuoka, Electron. Lett. **35**(19), 1664 (1999)
50. P. Skrimshire, J.R. Farr, D.F. Sloan, M.J. Robertson, P.A. Putland, J.C.D. Stokoe, R.R. Sutherland, Proc. Inst. Electron. Eng. **137**(1), 74 (1990)
51. F. Capasso, G.F. Williams, J. Electrochem. Soc. **129**(4), 821 (1982)
52. L.C. Kimerling, Solid-State Electron. **21**(11/12), 1391 (1978)
53. K. Morito, M. Matsuda, H. Ishikawa, O. Wada, in *Proceedings of the International Conference on Indium Phosphide and Related Materials*, Newport, RI, 1992, p. 427
54. S.J. Pearton, U.K. Chakrabarti, F.A. Baiocchi, Appl. Phys. Lett. **55**(16), 1633 (1989)
55. T. Takeshita, Y. Hirota, Y. Muramoto, T. Ito, Y. Tohmori, H. Ito, IEEE Trans. Electron Device. **53**(7), 1567 (2006)
56. T. Takeshita, M. Sugo, Y. Shibata, H. Kamioka, Y. Tohmori, IEEE Photon. Technol. Lett. **16**(1), 42 (2004)

# Chapter 4
# Reliability and Degradation of III-V Optical Devices Focusing on Gradual Degradation

**Osamu Ueda**

**Abstract** This chapter describes studies on the reliability of semiconductor optical devices over more than three decades, dating back to the early 1970s, focusing on gradual degradation. First, three main degradation modes for optical devices, rapid degradation, gradual degradation, and catastrophic failure, are outlined (see also Chaps. 1 and 5). Next, the results of the classical research into rapid degradation that was carried out in the 1970s and 1980s are presented as an introduction to a systematic discussion of the research that followed—remarkable research into gradual degradation. Then, we discuss the influence of local strain to the gradual degradation and methods for eliminating the gradual degradation rate.

## 4.1 Introduction

Semiconductor optical devices (hereafter referred to as optical devices), such as semiconductor lasers and light-emitting diodes (LEDs), come in a great variety of materials and structures, and they are used in an extremely wide range of fields. In addition to being used as light sources in medium-to-high-capacity fiber-optic communication systems, they are also used as light sources in consumer electronic devices, such as audio and digital systems, and in optical printers, as well as being used in sensing systems and medical equipment. It would thus be no exaggeration to say that improving the reliability of optical devices is essential to the development of high-performance, highly reliable products. That is why vigorous reliability studies have been conducted from the very outset and to the present day.

O. Ueda (✉)
Graduate School of Engineering, Kanazawa Institute of Technology, Atago Toyo Bldg. 12F, Tokyo Toranomon Campus, 1-3-4 Atago, Minato-ku, Tokyo 105-0002, Japan
e-mail: oueda@neptune.kanazawa-it.ac.jp

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_4,
© Springer Science+Business Media New York 2013

In this chapter, first we describe three main degradation modes for optical devices—rapid degradation, gradual degradation, and catastrophic failure [1–3]. Then, we present rapid degradation due to recombination-enhanced dislocation climb (REDC) and recombination-enhanced dislocation glide (REDG). Finally, a systematic discussion on the gradual degradation from early research in 1970s–1980s and remarkable research from late 1980s is shown.

## 4.2 Three Degradation Modes of Semiconductor Optical Devices

The three main degradation modes in optical devices are rapid degradation, gradual degradation, and catastrophic failure. When a semiconductor laser is operated with a constant output, there are cases in which the operating current will rapidly increase (as marked "Rapid" in Fig. 4.1a)—within 100 h, for example. When an LED is operated at a constant current, this corresponds to a rapid reduction in the optical output (as marked "Rapid" in Fig. 4.1b) known as "rapid degradation." Even if it is suppressed, however, there will still be the other degradation modes to consider. In semiconductor lasers, there is the phenomenon of a gradual increase in the operating current (see the curved line marked "Gradual" in Fig. 4.1a). In an LED, gradual degradation corresponds to a gradual decrease in the optical output (as shown in Fig. 4.1b). Both device types can also experience instantaneous degradation, due to a current surge, for example. This is referred to as "catastrophic failure" and is marked in Fig. 4.1a, b as "Catastrophic."

The characteristics of rapid degradation are: (1) the optical output decreases rapidly during constant current operations (or, in the case of constant output power operations, there is either a rapid increase in the operating current or a rapid decrease in the internal quantum efficiency), and (2) non-emitting regions are formed within the emitting region (the active region). There are three types of non-emitting regions: (i) dark-line defects (DLDs) [4], (ii) dark-spot defects (DSDs) [5], and (iii) dark regions [6]. Here, the half-life of a component (the

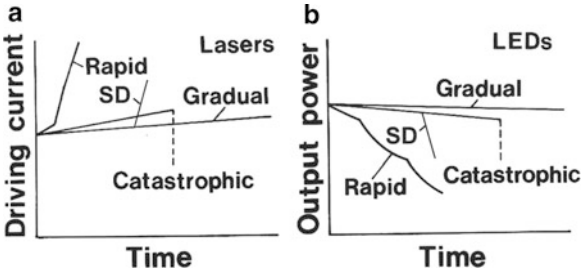



**Fig. 4.1** Three main degradation modes for semiconductor lasers and LEDs. (**a**) Degradation modes for semiconductor lasers; (**b**) degradation modes for LEDs.

time taken for its output to drop to half its initial level) is less than 100 h at room temperature. This type of degradation is due to either dislocation climb motion enhanced by non-radiative recombination of minority carriers, known as REDC (recombination-enhanced dislocation climb, see also Chap. 10 for more detail) or a similarly enhanced dislocation glide motion known as REDG (recombination-enhanced dislocation glide, see also Chap. 9 for more detail) [7].

In contrast to this, gradual degradation is a slow mode of degradation which remains extant even after the rapid degradation mode has been eliminated and which continues over a long term. Ultimately, it is this mode that determines the lifetime of a component. The characteristics of this degradation are: (1) optical output gradually decreases during constant current operations (or, in the case of constant output power operations, there is a gradual increase in the operating current), (2) there is a uniform darkening in the light-emitting region (DSD), and (3) (in the case of AlGaAs/GaAs devices) there is a gradual increase in deep levels [8]. This degradation mechanism is thought to involve point defect reactions due to non-radiative recombination during operations, the generation of point defects, and the formation of point defect clusters or micro-dislocation loops due to condensation of the point defects. (This is discussed in detail later in this chapter).

Catastrophic failure is a mode in which degradation occurs extremely fast, or "catastrophically," due to a current surge during high-output operations. In semiconductor lasers, catastrophic failure occurs due to catastrophic optical damage (COD) [9–11] at a mirror or in a defect region. In AlGaAs DH LEDs, catastrophic failure occurs due to dislocation glide [12]. In addition, catastrophic failure due to melting of the crystal can occur in InGaAsP/InP DH LEDs when Joule heating at the contacts causes a rise in temperature (for a detailed explanation of catastrophic failure, see Chap. 5 of this book and Chapter 12 of Ref. [2]) [13].

Recently, with growing customer demand for increased performance (in terms of higher output power), a new degradation mode has appeared: the "sudden death mode" (see SD in Fig. 4.1a, b), which causes serious damage and which is distinct from the modes described so far. It involves the sudden degradation of a device after a certain period of normal operations. The various causes of it are under investigation, and countermeasures are required for its prevention [14].

The next two sections consist of a discussion of the results of classical research into rapid degradation, followed by a more in-depth treatment of gradual degradation, the study of which has shown recent progress.

## 4.3 Rapid Degradation

As previously noted, rapid degradation is caused by REDC or REDG.

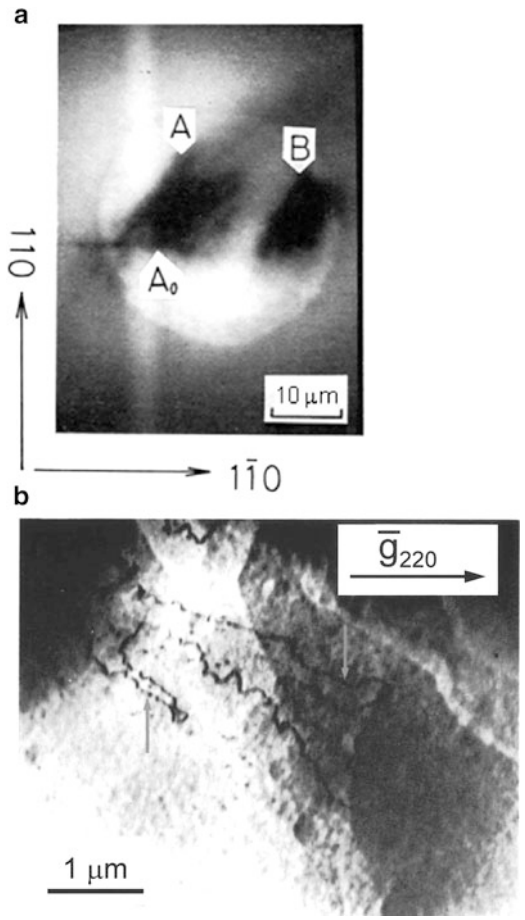



**Fig. 4.2** Rapid degradation of AlGaAs DH LED due to REDC. (**a**) EL image of degraded region; (**b**) TEM image of dislocation dipoles (U-shaped section indicated by the *arrow*) corresponding to a part of the <100> DLD denoted by A in Fig. 4.2a (Reprinted with permission from [15].Copyright 1977, American Institute of Physics)

### *4.3.1 Rapid Degradation Due to REDC*

Figure 4.2b is a transmission electron microscopy (TEM) image of dislocation dipoles (the U-shaped section indicated by the arrow) corresponding to a <100> DLD (see Fig. 4.2a) that was observed in degraded region of an AlGaAs DH LED that had suffered rapid degradation [15]. Detailed analyses under varying diffraction conditions reveal that these dislocation dipoles have Burgers vectors of (a/2) <011> that are inclined at an angle of 45° to the (001) plane and are of the interstitial type. These defects are believed to originate from threading dislocations that propagate from the substrate into the epitaxial layer when the crystal is growing and to develop and extend into the active region as a result of REDC. Prior to the above study, the same results had also been obtained for AlGaAs DH lasers that had suffered rapid degradation [4]. In addition, it was reported that dislocation dipoles were also formed from dislocation clusters [16] generated by the interface between

**Table 4.1** Summarized results of rapid degradation due to REDC in optical devices fabricated from three different materials

| | AlGaAs/GaAs | InGaAsP/InP | InGaAsP/InGaP |
|---|---|---|---|
| DSDs | Small helical dislocations | – | Small helical dislocations |
| <100> DLDs | Dislocation dipoles or climbed dipoles | Dislocation dipoles (very short) | Dislocation dipoles |
| <110> DLDs | Edge dislocation dipoles | – | Edge dislocation dipoles or faulted dipoles |

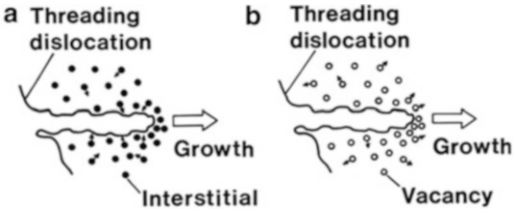



**Fig. 4.3** Two proposed models for dislocation climb motion enhanced by non-radiative recombination of minority carriers. (**a**) Extrinsic defect model; (**b**) intrinsic defect model (Reprinted with permission from [1]. Copyright 1988, The Electrochemical Society)

the substrate and the epitaxial layer and from dislocation loops that occurred during the growth of the crystal or during the impurity diffusion process [17, 18]. However, rapid degradation due to REDC has not been observed in InGaAsP/InP DH LEDs [19]. In addition, although degradation due to REDC has been reported in InGaAsP/InP DH lasers, it is clear that this takes place at a much slower rate (at least three orders of magnitude slower) than in AlGaAs lasers [20]. It has also been revealed that in InGaAsP/InGaP DH lasers in which there is lattice matching with an active layer of GaAs, rapid degradation also occurs if dislocation, dislocation clusters, or dislocation loops are present in the active layer [17]. Results of rapid degradation due to REDC in three different materials are summarized in Table 4.1.

As previously noted, the ease with which rapid degradation can occur in an optical device depends on the material used for the active layer. To understand the effects of this, let us consider two typical models of REDC out of the several that have so far been proposed. One is known as the "extrinsic defect model" [21] and is shown in Fig. 4.3a. In it, only a single type of interstitial atom—for example, a Ga interstitial atom—needs be absorbed into the dislocation core in order for dislocation to climb. The second model, known as the "intrinsic defect model," [22] is shown in Fig. 4.3b. It requires emission of two types of vacancies (those of groups III and V atoms) from the dislocation core—for example, emission of Ga vacancies and As vacancies in GaAs.

**Table 4.2** A summary of correlation between the formation of dislocation loops in heavily doped crystals and the easiness in REDC under carrier injection or optical excitation in various III-V compound semiconductors, alloy semiconductors, and related materials

| Material | Formation of dislocation loops | REDC |
|---|---|---|
| GaAs | Yes | Easy |
| AlGaAs | Yes | Easy |
| GaP | Yes | Easy |
| GaAsP | Yes | Easy |
| InP | No | Difficult |
| InGaAsP on InP | No | Difficult |
| InGaP on GaAs | Yes | Easy |
| InGaAsP on GaAs | Yes | Easy |
| ZnSe (II-VI) | Yes? | Very easy |
| InGaN/GaN[a] | – | Very difficult |

[a]In these materials, we assume that dislocations themselves do not act as non-radiative recombination centers (see Chap. 7 for detail)

It is currently unclear whether either of these two models is valid, but a number of physical parameters are believed to be candidates as factors contributing to REDC. First, as REDC is thought to involve the non-radiative recombination of minority carriers and to correspond to a local transfer of energy from a multi-phonon mode to a lattice vibrational mode excited by light absorption at defects in the crystal [7, 23], REDC is believed to depend essentially on the band-gap energy of the material. This provides a good explanation of why, for example, GaAs and GaP materials are prone to suffer from REDC, while InGaAsP materials are not. Table 4.2 shows a comparison between defect formation in heavy doping and the ease of REDC under carrier injection or optical excitation in various III-V compound semiconductors, alloy semiconductors, and related materials. From this table, it is seen that for almost all of these materials, this model can explain the ease of REDC. However, since InP—with its relatively wide band gap—is also prone to REDC, band-gap energy alone is not a sufficient explanation. Other dominant factors for REDC are deep levels, which are related to such defects as dangling bonds and native point defects, and non-radiative recombination rates at the deep levels. The cases for these factors are supported by the fact that AlGaAs materials (in which such defects form at deep levels during growth) and InGaAsP materials lattice matched to GaAs are also prone to suffer from REDC and by the fact that InGaAsP materials lattice matched to InP (in which such defects do not form at deep levels) are not prone to suffer from REDC. In addition, since REDC is believed to be due to the absorption of interstitial atoms in the dislocation core or to the emission of vacancies, it is thought that the magnitudes of generation energy and migration energy for point defects may also have a major influence on REDC.

As previously noted, rapid degradation due to REDC can essentially be inhibited by reducing such defects as dislocation, dislocation clusters, and dislocation loops during the crystal growth and processing stages. However, even today—in the 2010s—the specific mechanism behind REDC has yet to be fully elucidated, and much of the debate is based on conjecture. This will be the subject of future studies.

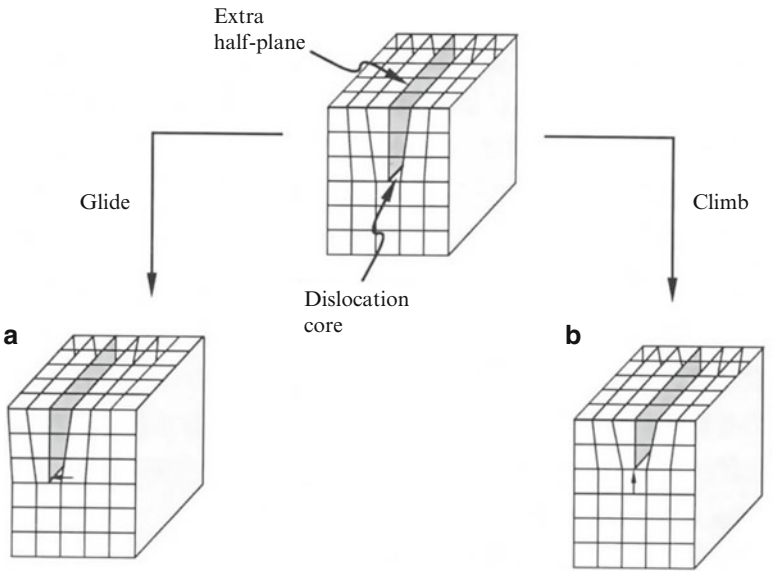



**Fig. 4.4** Schematic diagram of dislocation glide and climb motions. (**a**) A dislocation after glide motion; (**b**) a dislocation after climb motion.

## *4.3.2 Rapid Degradation Due to REDG*

### 4.3.2.1 Thermally Activated Dislocation Glide Motion in III-V Compound Semiconductors

Dislocation motions are classified into two types: glide motion and climb motion. As described in Sect. 4.3.1, the latter one is driven by absorption of vacancies or interstitials into the dislocation core or emission of these point defects from the dislocation core (see Fig. 4.4b). On the other hand, dislocation glide motion occurs when a dislocation with edge component moves on a plane normal to its half atomic plane by jumping an energy barrier called "Peierls potential." For this motion, it is required to cut the atomic bonding along the dislocation core and to combine them again as shown in Fig. 4.4a. In zinc blende structures, the glide system is <110>/{111}. Therefore, in III-V compound semiconductor thin films grown on (001) substrates, dislocation glide motion takes place on the four equivalent {111} planes along the <110> directions. Typical dislocation glide motion in AlGaAs/(001)GaAs substrate is shown in Fig. 4.5. For this motion, we need certain applied stress $\tau$.

The dislocation velocity $v$ is expressed as follows:

$$v = v_0(\tau)^m \exp\left(\frac{-U_p}{kT}\right) \tag{4.1}$$

where $v_0$, $U_p$, and $T$ are constant, Peierls energy for dislocation glide, and temperature, respectively.

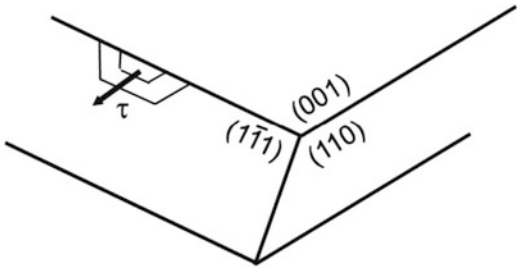



**Fig. 4.5** Schematic diagram of dislocation glide motion in III-V compound semiconductor thin films grown on (001) substrates. The dislocations are formed as half dislocation loops from the surface dislocation source via applied stress $\tau$

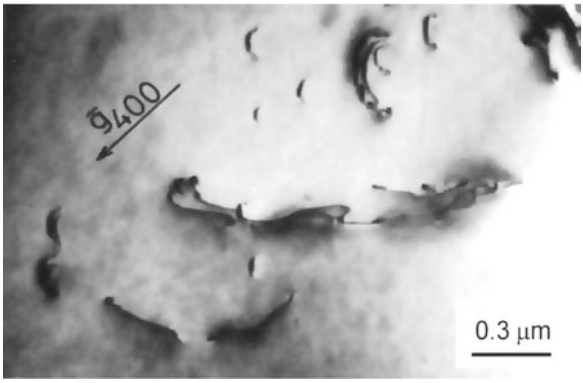



**Fig. 4.6** A plan-view TEM image of dislocation half loops generated by dislocation glide motion

Actually, in order to introduce many fresh dislocations by dislocation glide motion, it is required to deform the crystals, for example, by three or four points bending experiments with introduction of scratches on the sample surface, at high temperature (higher than 550°C for GaAs).

Figure 4.6 shows a plan-view TEM image of dislocations in the regions close to the surface of AlGaAs/GaAs epitaxial wafer. In this case, it is assumed that the dislocations are generated by accidental mechanical damage such as scratches by tweezers.

#### 4.3.2.2 Rapid Degradation of III-V Optical Devices by REDG

In AlGaAs/GaAs DH lasers and AlGaAs DH LEDs, rapid degradation also occurs by REDG, which gives rise to <110> DLDs in the active region. Figure 4.7a shows a typical EL image of <110> DLDs in a rapidly degraded AlGaAs DH LED [24]. The DLDs are usually observed near the edge of the circular active region where stress is concentrated due to the formation of circular p-electrode. Figure 4.7b

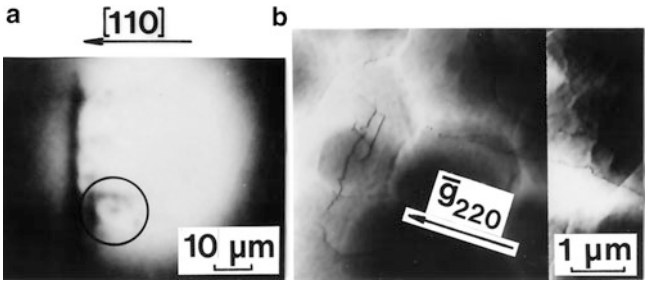



**Fig. 4.7** Rapid degradation of AlGaAs/GaAs DH LED due to REDG. (**a**) An EL image of <110> DLDs; (**b**) a TEM image of half dislocation loops corresponding to L-shaped <110> DLDs in (a) (see *circled region*) (Reprinted with permission from [24]. Copyright 1979, American Institute of Physics)

**Table 4.3** Summarized results of rapid degradation due to REDG

| | AlGaAs/GaAs | InGaAsP/InP | InGaAsP/InGaP |
|---|---|---|---|
| <110> DLDs | (1) Half loops<br>(2) Misfit dislocations<br>(3) Stacking faults | No reports found | Half loops |

shows a plan-view TEM image of defect structure corresponding to the <110> DLDs. These are half dislocation loops with Burgers vector of the type (a/2) <011> caused by dislocation glide motion from the p-side surface deep into the active region. In AlGaAs/GaAs epitaxial layers, glide dislocations were found after optical excitation of mechanical damage [25].

In InGaAsP/InP DH LEDs, no rapid degradation due to the presence of <110> DLDs was found during operation at room temperature. Although dark band defects were occasionally found in EL images, which were found to be associated with mechanical damage induced during LPE growth, no fresh glide dislocations developed from the mechanical damage after the operation at 20°C [26]. On the other hand, there are few reports on <110> DLDs in InGaAsP/InP DH lasers. Only Ishida and others have reported <110> DLDs in InGaAsP/InP DH lasers. However, they found that these DLDs correspond to glide dislocations generated during annealing process of p-electrode [20]. Half dislocation loop corresponding to <110> DLDs was also observed in rapidly degraded InGaAsP/InGaP DH lasers fabricated from crystals grown on (001) GaAs substrates [17].

Results of rapid degradation due to REDG in three different materials are summarized in Table 4.3.

This phenomenon is expected to be caused by both (1) non-radiative recombination of minority carriers in the active region and (2) optical excitation by emitted light in the light-emitted region and in regions where the emitted light is penetrated. Particularly in VCSELs, it is possible to enhance this phenomenon in whole area of the cylindrical waveguide (from the bottom DBR to the top DBR regions) [27], where optical excitation is possible to take place.

#### 4.3.2.3 Two-Step Degradation Process (REDG and REDC)

When REDG occurs during operation, the generated glide dislocations may have chance to climb immediately after they are generated or after certain interval depending on the operation condition. In an aging testing, this degradation phenomenon appeared as "two-step degradation processes," that is, sudden drop of optical output power at the first step, then followed by another decrease of optical power as the second step. Based on this assumption, we can expect with certain possibility to meet this degradation mode as follows:

Ridge-Waveguide-Type Lasers

The first step: REDG occurs at a region where certain amount of stress is accumulated during the operation, for example, bottom part of the ridge on both sides of the stripe; the second step: dislocation networks are generated from these glide dislocations by REDC.

AlGaAs/GaAs VCSELs

The first step: REDG occurs at region similar to the case of (1), for example, edges of the oxide aperture; the second step: dislocation networks are generated from the glide dislocations [28].

Since these degradation modes tend to appear as the most serious degradation mode, "sudden-death (SD) mode" (see Fig. 4.1), one should eliminate all factors for causing this particular degradation mode.

## 4.4 Gradual Degradation 1: Studies from the 1970s to the Mid-1980s

This section will first outline the results of studies conducted at the dawn of the era of research into gradual degradation. Due to the use of the liquid-phase epitaxy (LPE) process in crystal growth, there existed so-called "deep-level" defects that acted as traps for injected carriers in the undoped AlGaAs and GaAs crystals used as the active layers of optical devices. Among five models proposed for gradual degradation were gradual degradation mechanisms due to gradual increase of these deep-level defects (Model 1), as well as (in the case of InGaAsP materials) a gradual degradation mechanism due to native point defects (Model 2) and a diffusion or concentration of electrode atoms (Model 3).

These are outlined in the subsections that follow.

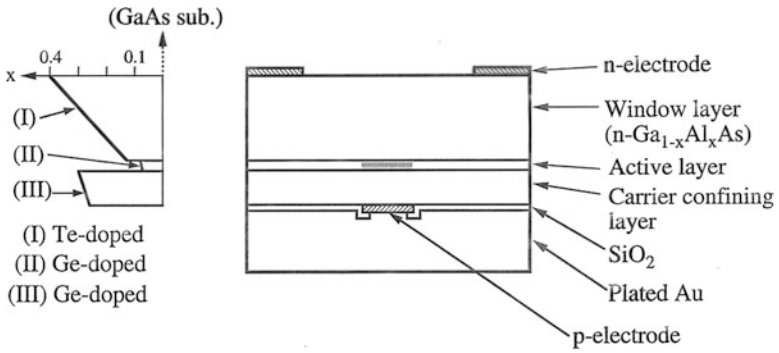



**Fig. 4.8** A schematic diagram of an AlGaAs DH LED

## *4.4.1 Gradual Increase of Deep-Level Defects in AlGaAs DH LEDs (Model 1)*

### 4.4.1.1 Degradation Characteristics of AlGaAs DH LEDs

In this sub-subsection, we describe an early work on the degradation characteristics of AlGaAs DH LEDs [29]. A schematic diagram of the AlGaAs DH LED with a 50 μm thick window layer is shown in Fig. 4.8. The AlGaAs epiwafer was prepared by LPE on (001)GaAs substrates. AlGaAs DH LEDs without DLDs and/or DSDs are easily selected by a short-term reliability testing [30], and such LEDs show gradual degradation over a long term.

The gradual degradation is recognized as a homogeneous decrease in light output with time. Accelerated aging at ambient temperatures of 60, 120, 180, and 250°C has been examined in lightly and heavily doped DH LEDs. The activation energy of gradual degradation was about 0.56 eV and did not depend on the operating current density or the concentration of acceptors, as shown in Fig. 4.9. The results of Zn-diffused DH LEDs are also shown for comparison [30]. An extrapolated half-life in excess of $4 \times 10^7$ h is estimated for operation at room temperature.

### 4.4.1.2 Study of Gradual Degradation of AlGaAs DH LEDs [8]

In order to clarify the gradual degradation mechanism of the AlGaAs DH LEDs, we have fabricated test devices in which we can characterize deep-level defects.

Degradation Characteristics of the Test LEDs

The LED used in this experiment consists of three layers of AlGaAs as schematically shown in Fig. 4.10: $n^+$-$Al_{0.3}Ga_{0.7}As$ window layer, p-$Al_{0.1}Ga_{0.9}As$ active layer, and

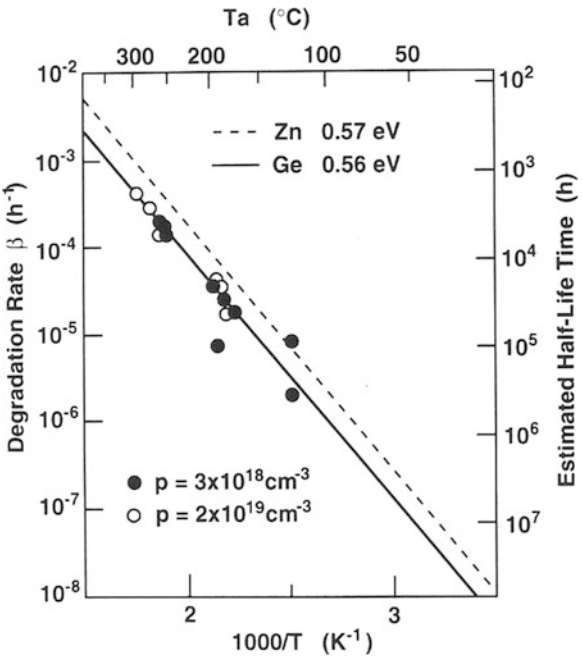



**Fig. 4.9** Temperature dependence of degradation rate of AlGaAs DH LEDs. The results of Zn-diffused DH LEDs are also shown for comparison [30]. The output power $P$ can be expressed as an exponential function with operating time $t$. $P = P_0 \exp(-\beta t)$, where $P_0$ and $\beta$ are initial power and the degradation rate, respectively. The degradation rate depends on the temperature at the emitting area and is given as $\beta = \beta_0 \exp(-E_0/kT)$, where $\beta_0$, $E_0$, and $k$ are the constant, the activation energy of gradual degradation, and the Boltzmann constant, respectively

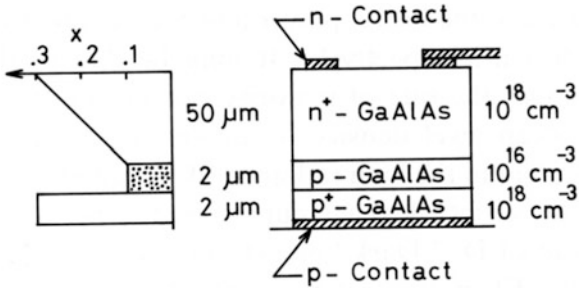



**Fig. 4.10** A schematic diagram of a test LED for the study of gradual degradation of AlGaAs DH LEDs

$p^+$-$Al_{0.3}$ $Ga_{0.7}As$ contact layer. The devices used in this experiment had a broad area p-contact (300 μm in square) without an $SiO_2$ carrier-confining layer, low doping of Ge in the active layer ($p = 1.6 \times 10^{16}$ cm$^{-3}$), and operated at a current density of 556 A/cm$^2$ at junction temperatures of 90, 169, and 230°C. On the other hand, commercial diodes have a confining $p$-contact (about 35 μm in diameter) with an

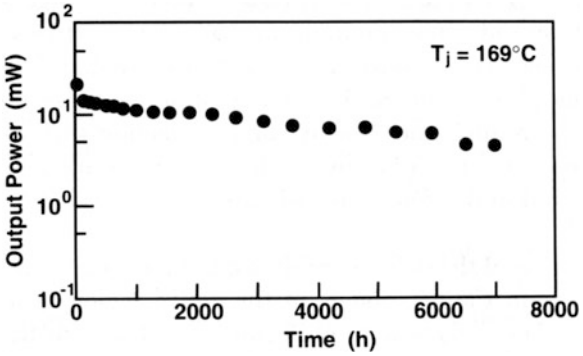



**Fig. 4.11** Output power of AlGaAs DH LED as a function of operating time measured at an ambient temperature of 20°C and at a current density of 556 A/cm$^3$ (Reprinted with permission from [8]. Copyright 1983, IEEE)

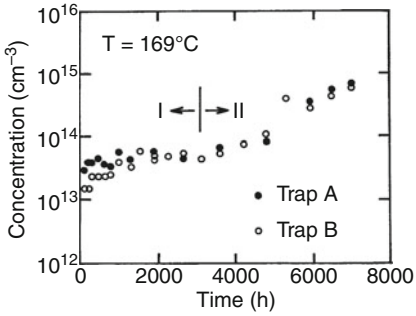



**Fig. 4.12** The concentration of the deep-level defects in the active region of AlGaAs DH LEDs as a function of the operating time (Reprinted with permission from [8]. Copyright 1983, IEEE)

$SiO_2$ layer, high doping of Ge in the active layer ($p = 5 \times 10^{18}\,\mathrm{cm}^{-3}$), and operate a current density of 10 kA/cm$^2$. This difference may not be significant because gradual degradation is probably related to the formation of native point defects and/or defect complexes as is discussed below, and moreover, it has been shown experimentally that the gradual degradation mode does not depend on the injection current density.

Specimens were selected from diodes without DLDs and/or DSDs in the EL images of the light-emitting region. The typical output power of the diodes is shown in Fig. 4.11 as a function of operating time. The diode shows a homogeneous decrease in output level, typical feature of gradual degradation.

### Evaluation of Deep-Level Defects During Aging

Figure 4.12 shows a typical change of defect concentration in the AlGaAs DH LED during operation. A two-stage increase in the number of deep-level defects,

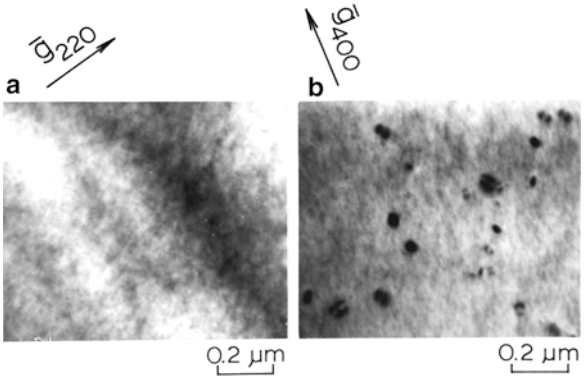



**Fig. 4.13** Bright-field TEM images of the active regions of AlGaAs DH LEDs. (**a**) Before operation; (**b**) after operation for 7,100 h at a current density of 556 A/cm$^2$ at a junction temperature of 169°C (Reprinted with permission from [8]. Copyright 1983, IEEE)

designated A and B, is observed. The defects A ($E_t - E_v = 0.43$ eV) and B ($E_t - E_v = 0.87$ eV) are hole traps similar to those commonly observed in LPE GaAs crystals and are assumed to be associated with point defect complexes [31]. The first stage (stage I) of the change is a rapid increase of [A] and [B] within several hundred hours of operation. The second stage (stage II) is a rather exponential rise over thousands hours of operation.

## Investigation of the Active Region by EL and TEM

Figure 4.13a, b are typical bright-field TEM images of the active region of the diodes before and after operation at 169°C, respectively. Small dislocation loops 15–50 nm in size were mainly observed after more than 7,000 h of operation, and no defects were found in the diodes before operation. This indicates that the defects in Fig. 4.13b were not generated by electron irradiation during TEM observation, but rather by carrier injection during device operation. Furthermore, no defects were observed in the confining (window and contact) layers of the degraded LEDs.

The distributions of the defects are shown in Fig. 4.15a–d. These TEM images were taken at regions indicated by arrows in the EL image shown in Fig. 4.14 (obtained from the active region of a diode operated at 169°C for 7,100 h). The defects were uniformly distributed in the active region and were irregular in size. Since the typical thickness of the TEM specimens was 0.5 μm, the defect concentration was estimated to be in the range of (2–8) $\times$ $10^{13}$ cm$^{-3}$.

The Burgers vector of the dislocation loops shown in Figs. 4.13 and 4.15 was determined from the usual "(**g** · **b**) s" analysis [32, 33]. It is concluded that they are interstitial-type "Frank loops" with Burgers vectors of (a/3) <111>. Since interstitial Frank loops include both the group III and V atoms in the zinc blende structure, these loops can grow by a climb process involving either the absorption of

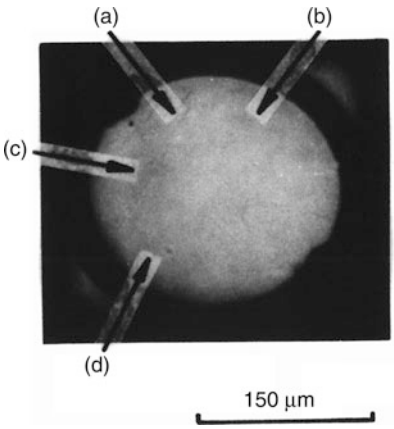



**Fig. 4.14** An EL image of an AlGaAs DH LED after operation at 169°C for 7,100 h (Reprinted with permission from [8]. Copyright 1983, IEEE)

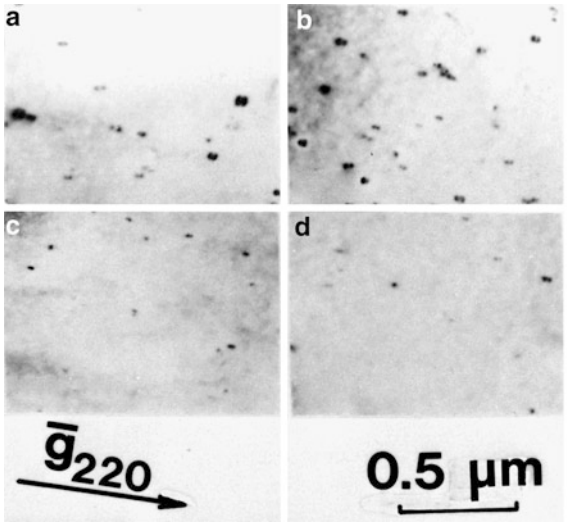



**Fig. 4.15** Bright-field TEM images obtained from the regions indicated by (**a**)–(**d**) in Fig. 4.14, respectively (Reprinted with permission from [8]. Copyright 1983, IEEE)

interstitials ($Ga_i$ and/or $As_i$) or the emission of vacancies ($V_{Ga}$ and/or $V_{As}$) from the loops. From the size and the concentration of the loops, the interstitial atoms ($Ga_i$ and/or $As_i$) contributing to the growth of the loops are estimated to be in the range (2–8) $\times$ $10^{17}$ $cm^{-3}$. This amount may be quite large compared to the concentration of the preexisting interstitial atoms in the material ($10^{16}$–$10^{17}$ $cm^{-3}$) [34, 35]. Therefore, the generation of point defects is required during device operation.

#### 4.4.1.3 Possible Gradual Degradation Mechanism of AlGaAs DH LEDs

On the basis of results described in the previous sub-subsection, the following model was proposed for the gradual degradation mechanism in AlGaAs DH LEDs.

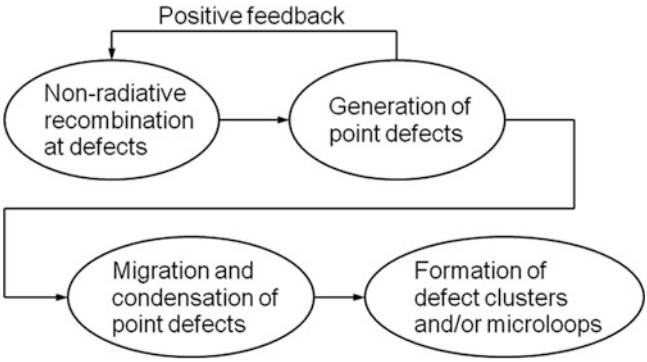



**Fig. 4.16** A model for gradual degradation in AlGaAs DH LEDs

First, non-radiative recombination occurs in some kind of defect (e.g., in an existing point defect or in point defect complex consisting of several existing point defects in an as-grown crystal), and new distinct point defect complexes are formed due to point defect reactions. Possible defect complexes in this context could be $Ga_i$-$V_{Ga}$, $V_{As}$-$Ga_{As}$, or $V_{As}$-$Ga_{As}$-$V_{Ga}$, since growth is carried out under Ga-rich or Al-rich conditions in order to promote LPE growth. These point defect complexes also act as non-radiative recombination centers. Accordingly, reactions generate positive feedback. As a result, point defects that are generated as by-products (individual interstitial atoms, such as $Ga_i$ and $Al_i$) migrate and concentrate into some kind of product nucleus (e.g., residual impurities such as O and C). Finally, point defect clusters and micro-dislocation loops are formed. The above model is schematically shown in Fig. 4.16.

## *4.4.2 Gradual Degradation in InGaAsP/InP Optical Devices (Models 2 and 3)*

In this sub-subsection, we focus the gradual decrease in the optical output of InGaAsP/InP DH LEDs during high-temperature acceleration testing—that is, the results of an investigation of gradual degradation in the device.

### 4.4.2.1 Gradual Degradation Due to Generation and Condensation of Native Point Defects (Model 2)

In a temperature acceleration test, ten diodes were linked together in a nitrogen atmosphere and operated at a constant current of 100 mA (DC). The optical output, cut-off frequency, and EL image of the diodes were then evaluated at 20°C at fixed time intervals. Although DSDs frequently appeared in the active regions during

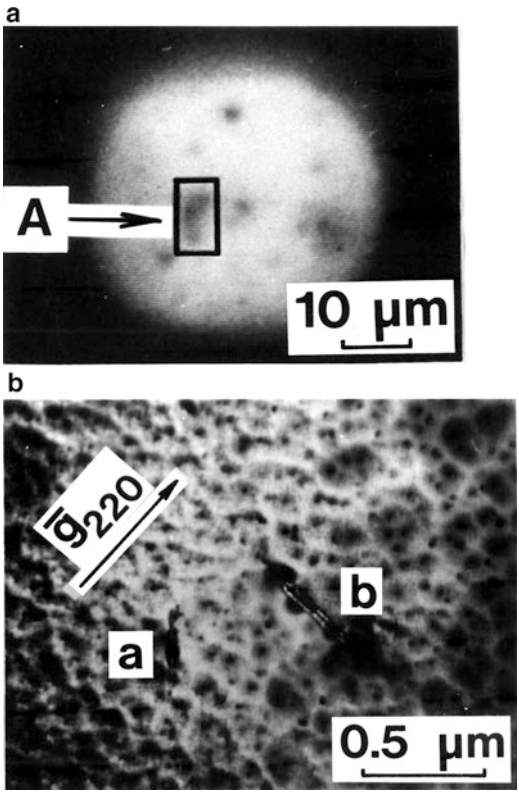



**Fig. 4.17** EL and TEM images of gradually degraded InGaAsP/InP DH LEDs. (**a**) EL image of the light-emitting region; TEM image corresponding to a single DSD observed in region A in (**a**). (*Bar-shaped* defects are observed at **a** and **b**) (Reprinted with permission from [36]. Copyright 1980, American Institute of Physics)

high-temperature operations, above 170°C, no rapid degradation occurred. It was also observed that, even when up to several tens of DSDs appeared in the active regions, optical output remained essentially unchanged.

Figure 4.17a shows an EL image of DSDs that appear in high-temperature operations in an InGaAsP/InP DH LED [36]. The number of DSDs is less than 30 and is on average around 10. DSD diameters are 2–3 μm, and contrast levels among them vary.

Figure 4.17b shows a TEM image of a defect that correspond to one of the DSDs observed in the region shown in A in Fig. 4.17a [36]. Comparing these two images reveals that several bar-shaped defects can be related to a single DSD.

Moreover, since the contrast did not disappear under any diffraction conditions, and since no abnormal spots were seen in the electron diffraction patterns obtained from these defects, it was assumed that the defects were precipitates.

Based on the above results, the following mechanism was suggested as being a factor in the formation of the bar-shaped precipitates (Model 2):

1. The interstitial atoms of the constituent elements of the matrix crystal are formed by a defect reaction caused by the non-radiative recombination of injected carriers at high temperature.

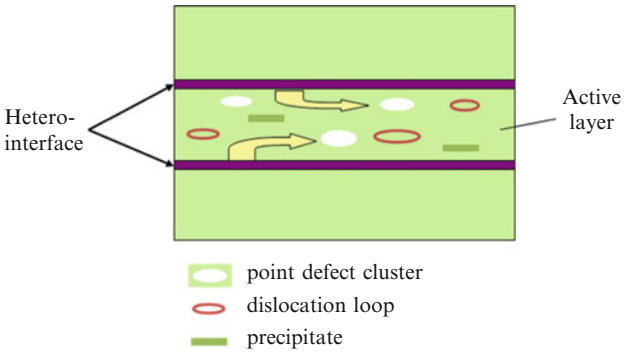



**Fig. 4.18** A model for generation of point defect clusters and dislocation loops (or precipitates) during gradual degradation (cross-sectional view)

2. These interstitial atoms are somehow absorbed into a product nucleus, and bar-shaped precipitates are formed.

Although this model was also suggested for semiconductor lasers, no supporting experimental evidence was obtained. Similar results were obtained by Wakita et al. [37]. They have shown that dislocation loops and platelike precipitates are associated with DSDs. A model related to formation of point defect clusters and dislocation loops (or precipitates) is schematically shown in Fig. 4.18 [38]. In this model, the both hetero-interfaces act as precursor regions for the point defects. Native point defects (vacancies and/or interstitial atoms of host elements) are emitted from the hetero-interface to the active region via certain recombination-enhanced point defect reaction (REDR). The emitted point defects form point defect cluster that eventually coalesce into dislocation loops or precipitates.

#### 4.4.2.2 Gradual Degradation Due to Diffusion of Metallic Atoms from the *p*-Electrode (Model 3)

Apart from the model 2, though interstitial atoms could not be detected from the precipitates using EDX, another mechanism that was regarded as a possible candidate, as discussed below, was the formation of precipitates due to diffusion of metal atoms from the electrodes (such as Au) penetrating into the active layer and concentrating there.

Figure 4.19 shows an Arrhenius plot obtained in a high-temperature acceleration testing of InGaAsP/InP DH LEDs [39, 40]. In the diagram, $\lambda p$ and $\beta$ (degradation rate) represent, respectively, the peak emission wavelength and the degradation rate for LEDs in a constant current operation test.

The degradation rate, $\beta$, is shown in the following equation:

$$P = P_0 \exp(-\beta t) \tag{4.2}$$

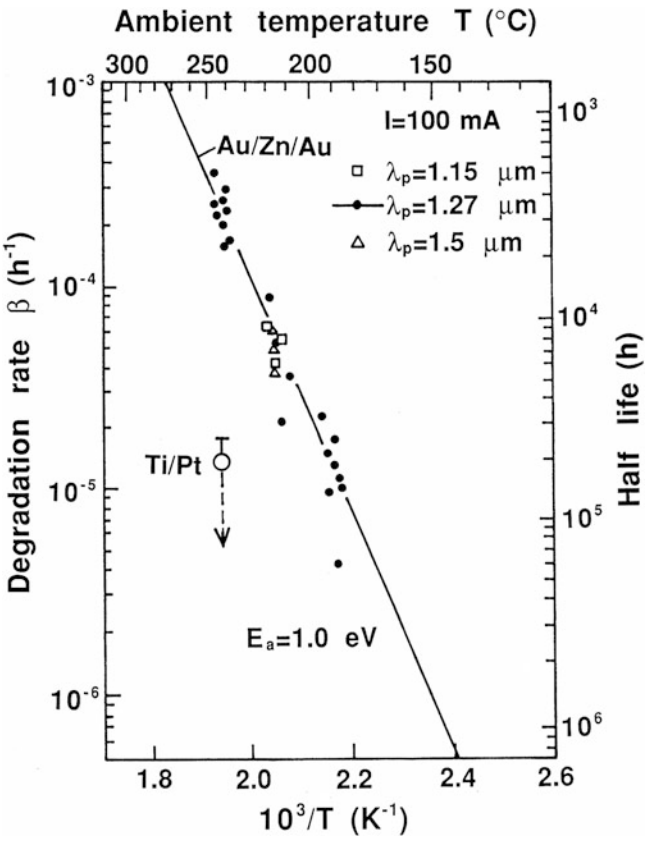



**Fig. 4.19** The temperature dependence of the degradation rates of InGaAsP/InP DH LEDs. The *solid line* shows the calculated result by a least square fit

where $P$ represents the optical output and $P_0$ the initial optical output. The degradation rate depends on ambient temperature (or the junction temperature) that is also given as

$$\beta = \beta_0 \exp\left(\frac{-E_a}{kT}\right) \tag{4.3}$$

where $\beta_0$, $E_a$, and $k$ are the constant, the activation energy of gradual degradation, and the Boltzmann constant, respectively.

The data set connected by straight lines in the figure was obtained from LEDs that use an Au/Zn/Au alloy electrode as the *p*-electrode. The figure also shows one data point for an LED using a Ti/Pt/Au non-alloy electrode. These results can be summarized as follows:

(a) The degradation mechanism is independent of the wavelength of the active layer—that is, of the composition of the quaternary crystal.

(b) The activation energy is 1.0 eV, which is higher than the 0.5–0.7 eV in AlGaAs DH LEDs.
(c) When Ti/Pt/Au non-alloy electrodes (which have strong barrier properties) are used, the lifetime of the device is an order of magnitude greater than when alloy electrodes are used.

The above results suggest that the degradation mechanism might be thought of as a process whereby Au atoms from the *p*-electrode diffuse into, and concentrate in, the active layer during the operation, due to non-radiative recombination or carrier overflow (Model 3) [40–42].

## 4.5 Gradual Degradation 2: Advances in Research Since the Mid-1980s

As the capabilities of InGaAsP/InP long-wavelength lasers grew ever greater, defects in the light-emitting region caused by gradual degradation during high-temperature, high-output operations could clearly be seen using EL (or EBIC) and TEM, and the degradation mechanism became fairly clear. In this section, two extensive studies on gradual degradation mechanisms since the mid-1980s are introduced.

### *4.5.1 Gradual Degradation in InGaAsP/InP DH Lasers 1: Formation of Prismatic-Type Micro-dislocation Loops (Model 4)*

The first of the two studies was by Chu et al. [43]. This group analyzed degradation in CSBH (channeled-substrate buried heterostructure) lasers during acceleration testing in which the operating current was 200 mA and the ambient temperature was 100°C. Dark regions along both sides of the stripe can be seen in the EL image of the degraded diode. <100> DLDs were also frequently observed in components, although these were short. Using plan-view TEM observation, the group first found that high-density interstitial-type micro-dislocation loops had formed on both sides of the stripe.

From cross-sectional TEM observation, they also ascertained that the dislocation loops were not in the active layer but just below the active layer and that they were formed in the vicinity of the n-InP buffer layer just above the p-InP current restriction layer (see (2) in Figs. 4.20 and 4.21b) [44]. As these loops are only present in laser diodes in which both edges of the active layer are in contact with the *p*-InP current restriction layer, the group also proposed a model for the formation of these loops in which carriers (positively charged holes, in this case) that flow to the region outside the active layer become trapped in this region (in the vicinity of

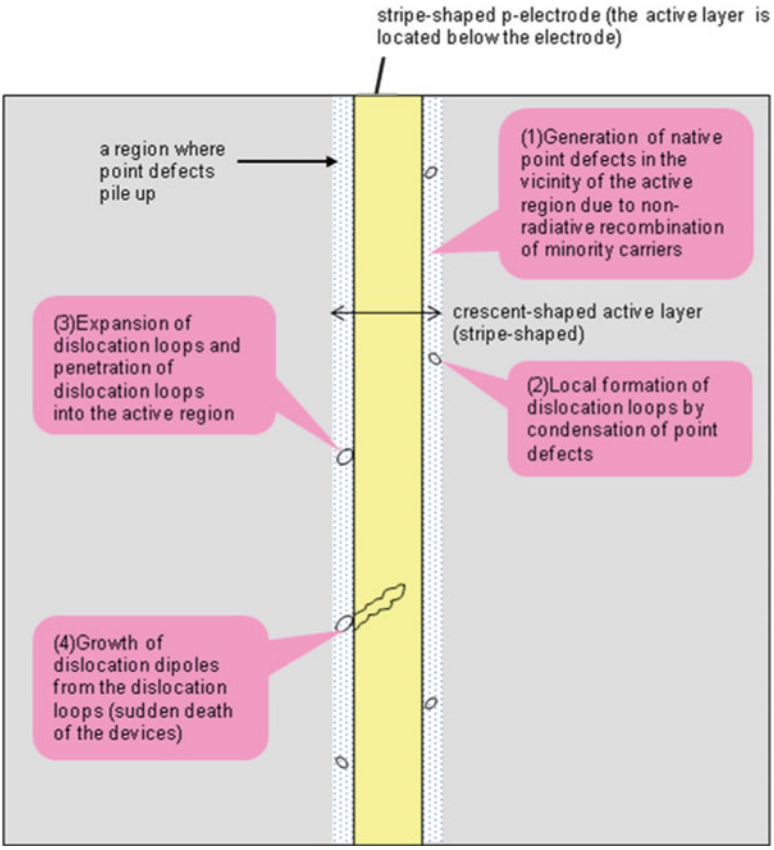



**Fig. 4.20** Plan-view diagram of the mechanism for transition from gradual degradation to sudden death in channeled-substrate buried heterostructure (CSBH) laser (Reprinted with permission from [44]. Copyright 2010, The Japan Society of Applied Physics)

the boundary between the *n*-InP buffer layer and the *p*-InP current constriction layer—the region shown in pink in Fig. 4.21a) [44] by point defects formed during crystal growth, with the interstitial atoms which form as a result of the non-radiative recombination of minority carriers concentrating to form dislocation loops (see (1) and (2) in Figs. 4.20a and 4.21b) [44].

The group also verified that dislocation dipoles can grow from the dislocation loops in the <100> direction, although this is rare. The conjecture is that this is probably the result of rapid degradation that occurs because part of the dislocation loop grows, expanding to make contact with, and penetrate, the active layer (see (3) and (4) in Figs. 4.20 and 4.21c, d) [44]. These results indicate that care is required here because, even in InGaAsP/InP lasers, under severe current flow conditions, gradual degradation may develop into rapid degradation due to the growth of dislocation dipoles, and this may lead to sudden death.

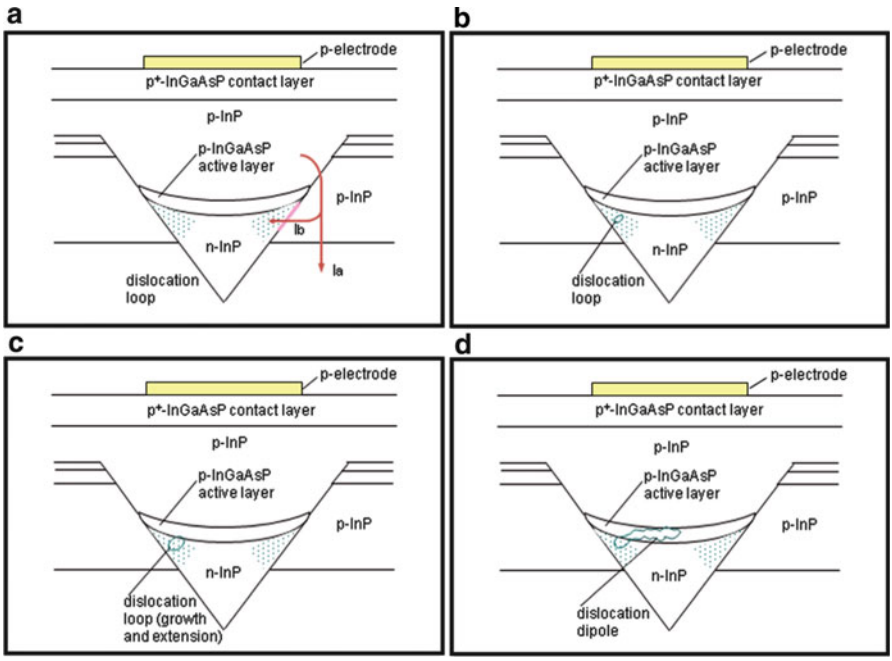



**Fig. 4.21** Cross-sectional diagram of the mechanism for transition from gradual degradation to sudden death in CSBH laser. (**a**) Generation of native point defects; (**b**) formation of dislocation loops; (**c**) expansion of dislocation loops; (**d**) growth of dislocation dipoles

### *4.5.2 Gradual Degradation in InGaAsP/InP DH Lasers 2: Formation of (Frank-Type) Micro-dislocation Loops (Model 5)*

This subsection introduces the model proposed by de Cooman et al., which is currently regarded as the most influential model [45]. The group subjected the so-called DCPBH (double-channel planar buried heterostructure) lasers—embedded-type 1.3-μm InGaAsP/InP lasers [46]—to high-temperature (70°C) acceleration testing and analyzed components that suffered degradation. First, from the cross-sectional EBIC (electron beam induced current) observation of degraded diodes, areas with dark contrast were observed in the regions just outside—and on both sides of—the active layer stripe. Plan-view TEM observations showed that multiple interstitial-type dislocation loops (Frank loops on the (111) plane) were generated which corresponded to these dark regions. Cross-sectional TEM observation has shown that the loops were generated in the p-InP layer just outside the normal mesa sidewall region of the active layer.

Thus, despite the fact that the de Cooman group obtained the similar observational results to the Chu group (although the type of loops identified are different to

each other), the degradation mechanism that it proposed was based on an entirely different view. That is, although in the case of this structure, p-type dopant of Zn atoms (the carrier concentration is only around $1 \times 10^{17}\,\mathrm{cm}^{-3}$) is doped just beside the active layer, the de Cooman group focused on the behavior of these p-type dopant atoms. Their work also explained to some extent the physics behind the ready diffusion of Zn in compound semiconductor crystals subjected to high-temperature heat treatment. For example, if Zn is intentionally diffused into an AlAs/GaAs superlattice structure, Ga and Al atoms are discharged due to the injection of the Zn, and the super-lattice is destroyed as a result. This phenomenon was initially discussed in relation to interstitial Si atoms in Si crystals and has subsequently been called as "kick-out mechanism," in reference to the diffusion of Zn in GaAs [47]. The model proposed by the group is as follows.

In Step 1, the Zn interstitial dopant atoms are energized by some kind of injected-carrier non-radiative recombination process, as shown below, and they diffuse. At this point, group III atoms (such as In atoms) that are those of the constituent element of the matrix crystal are discharged from their position in the lattice.

It is believed that, if this process occurs at a certain frequency, it excites the Step 2 reaction, so as to reduce the free energy in the system. Under these conditions, intrinsic point defects proliferate in the vicinity of the active layer. These somehow concentrate in the nuclei, and a dislocation loop is formed.

$$\text{Step 1}: \ \mathrm{Zn_i}^{+} \rightarrow \mathrm{Zn_s}^{-} + \mathrm{In_i}(\text{or } \mathrm{Ga_i}) + 2\mathrm{h}^{+} \tag{4.4}$$

$$\text{Step 2}: \ \mathrm{P} \rightarrow \mathrm{P_i} + \mathrm{V_P} \tag{4.5}$$

Of all the models proposed so far, this model seems to offer the clearest physical picture and to be best able to explain the phenomenon of degradation.

There is a report by Endo et al., with reference to this model, about lasers which have an active layer of strongly ordered InGaP and which suffer degradation due to disordering of their ordered structure during operation [48]. They proposed a similar model (degradation caused by diffusion of Zn) based on the rate of increase in the threshold current, which is dependent on $\sqrt{t}$. Of major interest, too, is the assumption by Tomiya et al. that in GaN lasers, the rate of degradation also depends on $\sqrt{t}$ and that degradation is due to the diffusion into the active layer of defects related to Mg (see Chap. 7 of this book for more detail) [49].

## 4.6 Gradual Degradation 3: Influence of Internal Stress on the Gradual Degradation

In this section, influence of internal stress on the gradual degradation is described for two cases: acceleration of gradual degradation in (1) CSP-type AlGaAs DH lasers in short wavelength range and (2) strained high-power InGaAs/AlGaAs MQW lasers in wavelength of 980 nm.

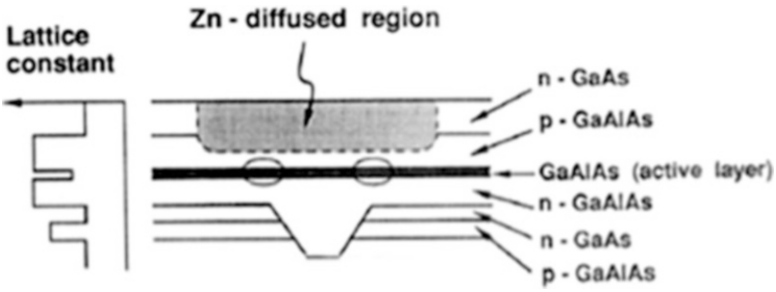



**Fig. 4.22** Schematic diagram of a CSP-type AlGaAs DH laser

## *4.6.1 AlGaAs DH Lasers Emitting in Short Wavelength Range [50]*

### 4.6.1.1 Degradation Characteristics of the Laser

We fabricated CSP-type AlGaAs visible DH lasers emitting at 770 nm using LPE, the structure of which is shown in Fig. 4.22. During life testing under a constant power of 5 mW at 50°C, some of these lasers showed rapid increase in the driving current. The threshold current increased by two or three times of the initial value, 70–80 mA, and dark defects were observed in the active layer by PL imaging. The remaining devices showed only a slow increase in the driving current, and dark defects were not observed in the active layer.

Figure 4.23 shows a PL image of the rapidly degraded AlGaAs active layer observed using a $Kr^+$ ion laser as an exciting source ($\lambda = 676.4$ nm) and a Si vidicon detector, after removing the n-contact layer. Numerous dark defects on both edges of the stripe suggest that the strong stress concentration existed there.

### 4.6.1.2 Investigation of the Degraded Region by TEM

TEM observations revealed that the defects seen in the PL image were precipitate-like microdefects, micro-dislocation loops, and dislocation dipoles [50]. In this work, contrast experiment with the 220, −220, 400, and 040 diffraction vectors, observation with a weak beam, and stereoscopic observation were carried out. Figure 4.24 is a TEM image corresponding to one of the "dark bands" appeared along both edges of the stripe, which consists of precipitate-like microdefects (or point defect clusters), micro-dislocation loops, and dislocation dipoles. They lie also on the two {111} planes inclined toward the (001) surface and become broad in the direction of the epitaxial growth. The clusters are 5–10 nm in size, and their density is about $10^{15}$ $cm^{-3}$. The dislocation loops are 10–20 nm in size with density of about $10^{14}$ $cm^{-3}$. The spacing between the dipoles is 20–30 nm. They originated in the

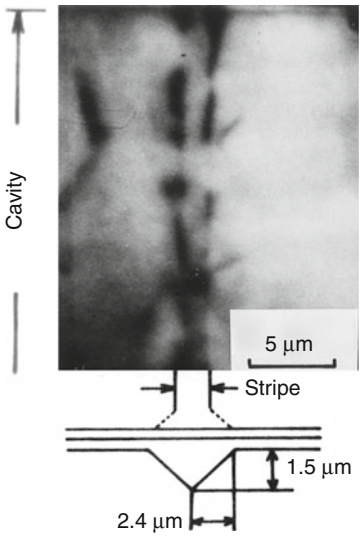



**Fig. 4.23** A PL image of dark defects in a rapidly degraded CSP-type AlGaAs DH laser (Reprinted with permission from [50]. Copyright 1985, American Institute of Physics)

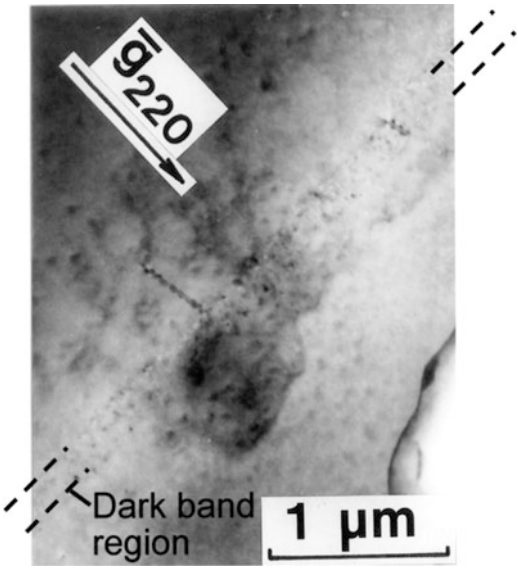



**Fig. 4.24** A TEM image of defects corresponding to the dark defects shown in Fig. 4.23 (Reprinted with permission from [50]. Copyright 1985, American Institute of Physics)

dislocation loops and extended to the inside of the stripe region. As the dislocation dipoles are out of contrast under the 220 diffraction, they may be sessile dislocations with (a/2) <110> or (a/11) <113> Burgers vectors. Since the dipoles are straight line-shaped, they are assumed to be generated as an artifact of the thinning process for TEM specimen.

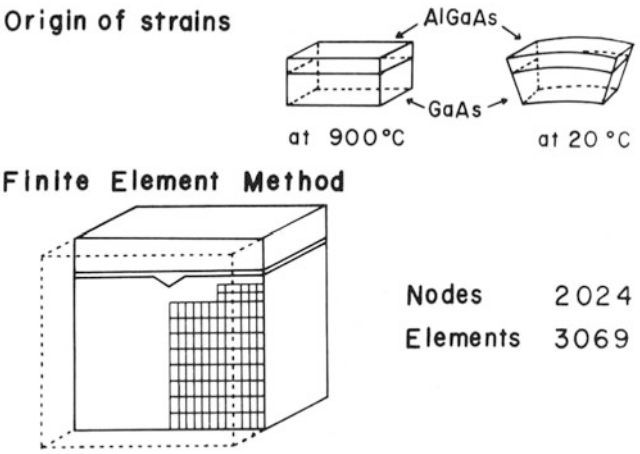



**Fig. 4.25** Two-dimensional finite element mesh of the CSP-type laser, which is used for FEM stress simulation (Reprinted with permission from [50]. Copyright 1985, American Institute of Physics)

#### 4.6.1.3 Analysis of the Two-Dimensional Stress Distribution

Here, we describe the results of the two-dimensional stress distribution for the visible AlGaAs DH laser by finite element method (FEM) [50]. At the interfaces between each layer in the layer structure of the laser, we assume that atom-by-atom alignment is preserved across the interface and that there is no thermal strain at 900°C. Then, internal stresses in the multilayers are a consequence of a difference in thermal expansion coefficients. For $Al_xGa_{1-x}As$, the thermal expansion coefficient is expressed in terms of the GaAs ($\alpha_{GaAs}$) and AlAs ($\alpha_{AlAs}$) [51–53]:

$$\alpha_{AlGaAs} = (1 - x)\alpha_{GaAs} + x\alpha_{AlAs} \tag{4.6}$$

where $\alpha_{GaAs} = 6.86 \times 10^{-6}$ ($^{\circ}C^{-1}$) and $\alpha_{AlAs} = 5.20 \times 10^{-6}$ ($^{\circ}C^{-1}$).

The elastic modulus E and Poisson's ratio ν necessary for calculation are assumed to have the same values for all layers [51, 54] as follows:

$$E = 1 \times 10^{12}\left(\text{dyne}/\text{cm}^2\right), \nu = 0.23.$$

In applying the two-dimensional FEM technique to a channel stripe, we divide the cross section of the visible laser into 2,024 nodes and 3,069 elements, as shown in Fig. 4.25. The values of the stress are calculated by changing the Al content of the active layer while keeping the Al content of other layers constant. In Fig. 4.26a, we plot the stress $S_{xx}$ parallel to the active layer and $S_{yy}$ perpendicular to it and the shear stress $S_{xy}$ in the active layer for $\lambda = 770$ nm ($x = 0.15$) against $X$ where the origin of $X$ is in the center of the stripe. Stress concentrations are present on both edges of the stripe ($X = \pm 2.4$ μm): $S_{xx}$ is about $-3 \times 10^8$ dyne/cm$^2$ and is highly

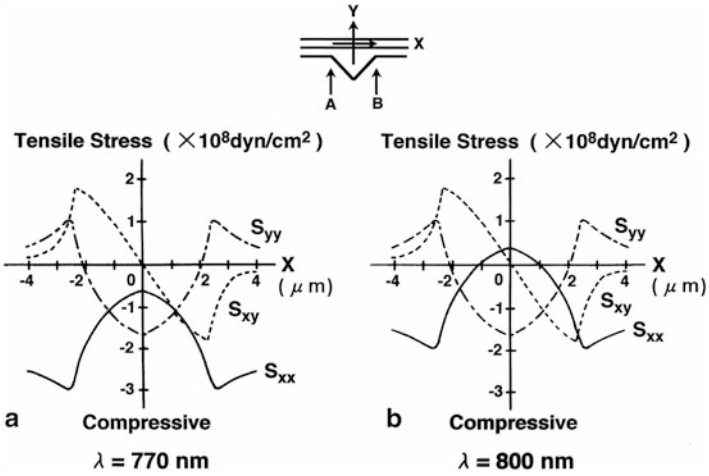



**Fig. 4.26** Calculated stress distribution in the active layer plotted against $X$, where the origin is the center of the stripe region. (**a**) $\lambda = 770$ nm; (**b**) $\lambda = 800$ nm (Reprinted with permission from [50]. Copyright 1985, American Institute of Physics)

compressive. On the other hand, $S_{yy}$ is about $1 \times 10^8$ dyne/cm$^2$ and is tensile. Figure 4.26b shows the stress distribution for $l = 800$ nm ($x = 0.1$). The value $S_{xx}$ at $X = 2.4$ μm is about $-2 \times 10^8$ dyne/cm$^2$ and is less compressive than when $X = 0.15$.

#### 4.6.1.4 Degradation Mechanism: Acceleration of Gradual Degradation Mode by Internal Stress

Here, we describe the degradation mechanism proposed by Ikeda, Ueda, and others [50]. They first focused the cubic dilation distribution in the active layer caused by the stress distribution [55]. Strain fields $u_{xx}$ and $u_{yy}$ are related to the stress $S_{xx}$ and $S_{yy}$ as follows:

$$u_{xx} = [(1+v)/E]\left[(1-v)S_{xx} - vS_{yy})\right] \tag{4.7}$$

$$u_{yy} = [(1+v)/E]\left[(1-v)S_{yy} - v\,S_{xx})\right] \tag{4.8}$$

Then, the cubic dilation $\Delta V/V$ is given by [52]

$$\frac{\Delta V}{V} = u_{xx} + u_{yy} = \left[\frac{(1+v)(1-2v)}{E}\right]\left(S_{xx} + S_{yy}\right) \tag{4.9}$$

They calculated the $\Delta V/V$ of the active layer as functions of the distance $X$ from the center of the stripe [50]. In both cases, $\lambda = 770$ and 800 nm, $\Delta V/V$ has a

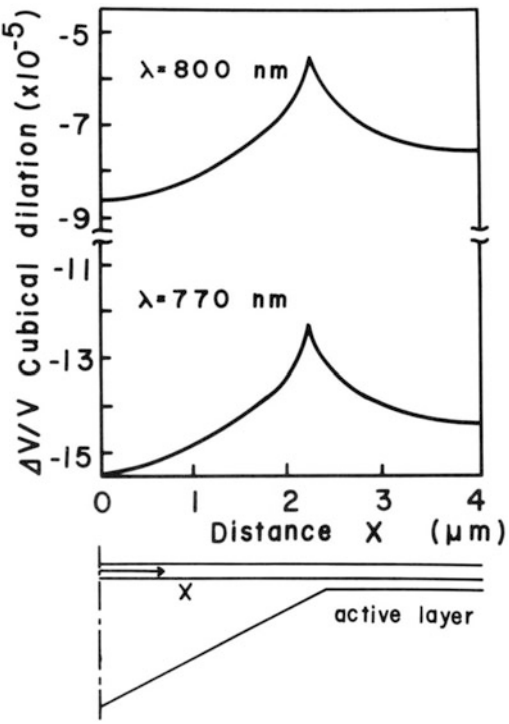



**Fig. 4.27** Plot of cubical dilation d$V/V$ ($l = 770$ and 800 nm) as a function of the distance $X$ from the stripe (Reprinted with permission from [50]. Copyright 1985, American Institute of Physics)

maximum near the edge of the stripe (see Fig. 4.27). When $\lambda = 770$ nm, it is more compressive with decrease in lasing wavelength.

Based on the TEM observations and the calculations, they have proposed the following mechanism for the visible AlGaAs lasers:

Step 1: During operation of the laser, interstitial atoms and vacancies are generated in the active layer by non-radiative recombination of minority carriers.

Step 2: Interstitial atoms in the active layer migrate toward the regions with maximum cubic dilation (i.e., form the center of the channel to the edges), lowering the stress and thus the total energy of the system.

Step 3: The interstitial atoms concentrated on the stripe edges form point defect clusters and dislocation microloops.

Step 4: These process finally causes generation of high-density non-radiative recombination centers on both sides of the stripe region which appear as dark band along the stripe.

Since there is no clear explanation for the point defect generation, the model proposed by de Cooman (see Sect. 4.5.2) may be also applied to this situation.

Anyway, these results led us to the conclusion that, in AlGaAs visible DH lasers (when the lasing wavelength is shorter than 780 nm), especially in CSP-type lasers,

the diode degrades unexpectedly rapidly due to strong enhancement of gradual degradation mode by the internal stress. Similar results have also been reported in degraded oxide-stripe geometry AlGaAs [56, 57].

### 4.6.2 *Strained High-Power InGaAs/AlGaAs MQW Lasers*

In this subsection, we describe degradation of strained InGaAs/AlGaAs MQW lasers reported by Frigeri et al. [58].

The laser diodes they used are InGaAs/AlGaAs double quantum well structures, emitting at 0.94 μm. In this structure, the Mg-doped p-AlGaAs layer is used as a p-cladding layer. This layer receives highly tensile strain by the MQW structure. After an accelerating testing at high temperature, some of the tested diodes degraded. They evaluated these degraded diodes by PL imaging and TEM.

Figure 4.28 shows a PL image of the degraded laser. A dark band-shaped region elongating in the <100> direction 45° inclined to the stripe is clearly observed. They further investigated the areas of dark band and non-dark band by cross-sectional TEM as shown in Fig. 4.29. In the heavily degraded region, dislocation networks consisting of interstitial-type dislocation dipoles extending to both upper and lower area of the MQW were observed (Fig. 4.29a). On the other hand, a high density of microdefects: dislocation loops and point defect clusters similar to those observed in the gradually degraded AlGaAs DH LEDs shown in Fig. 4.24 were observed in the p-cladding layer just above the MQW structure.

From these results, they proposed a model for the degradation mechanism:

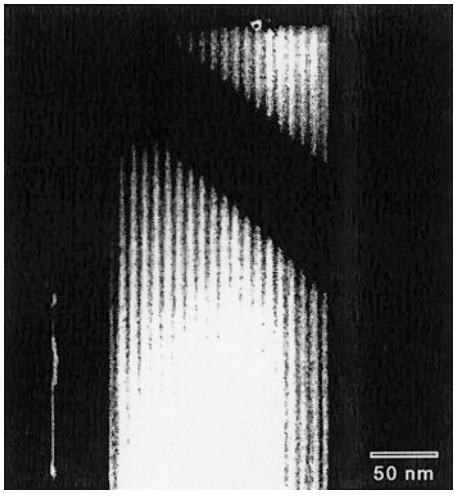



**Fig. 4.28** PL image of the active layer of the degraded InGaAs/AlGaAs MQW (double quantum well) laser. Dark band elongating along the <100> direction is observed (Reprinted with permission from [58]. Copyright 1997,Taylor & Francis)

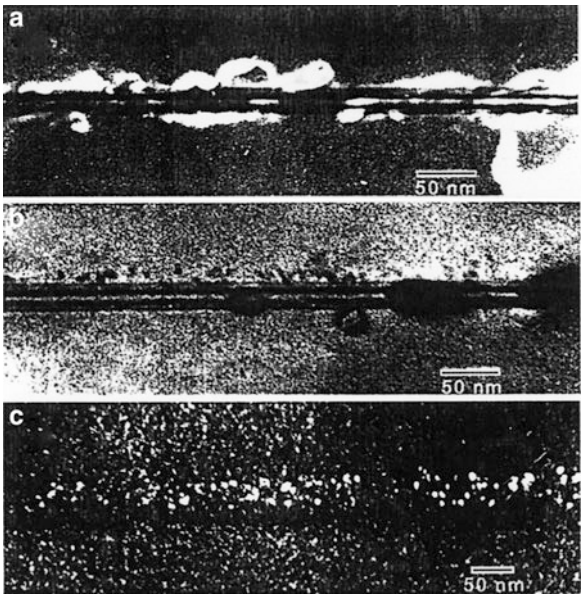



**Fig. 4.29** Cross-sectional TEM images of the degraded lasers shown in Fig. 4.28. (**a**) a 400 dark-field TEM image of the region corresponding to the dark band region shown in Fig. 4.28; (**b**) a 220 bright-field TEM image of the region close to the edge of the dark band region; (**c**) a 220 weak-beam dark-field image of the region outside of the dark band region (Reprinted with permission from [58]. Copyright 1997, Taylor & Francis)

(a) Generation of point defect clusters (possibly, clusters of supersaturated interstitial atoms) in the p-cladding layer just above the MQW structure.
(b) The point defect clusters grow in size and form interstitial-type micro-dislocation loops.
(c) The micro-dislocation loops further grow high density of interstitial-type dislocation dipoles.

Regarding the generation of point defect clusters prior to the loop formation, they also raised two candidates:

1. High strain field in the vicinity of the active layer
2. Diffusion of the p-type dopants (Mg atoms, in this case)

The factor (1) accelerates the diffusion of p-type dopant atoms, and the factor (2) is a key mechanism for the point defect reaction: kick-out mechanism producing vacancy-interstitial pairs by diffusing species (interstitial Mg atoms, in this case). Although the exact mechanism for the diffusionMQW structure has not been clarified yet, this may be explained by the facts that (1) the penetrated photoemission from the active region into the cladding layer or (2) overflow of minority

carriers in the active region towards the p-cladding layer enhances the non-radiative recombination process. In this case, there must be some point defect complexes for causing non-radiative recombination event in the region, for example, defect center such as [Mg-X]. It is also assumed that local strain field can accelerate the phenomenon.

Finally, based on the above discussion, one can envisage the possibility of the strong enhancement of the gradual degradation by the local strain field in the case of ridge waveguide lasers, vertical cavity surface-emitting lasers (VCSELs), etc.

## 4.7 Elimination of Gradual Degradation

In the previous sections, the results of analyses on gradually degraded optical devices (i.e., AlGaAs DH LEDs, AlGaAs DH lasers, InGaAsP/InP DH LEDs, InGaAsP/InP DH lasers, and strained InGaAs/AlGaAs MQW lasers) are described, and five models for the degradation mechanism are presented. In this section, possible methods for eliminating the gradual degradation are discussed based on these results.

### *4.7.1 Elimination of Deep-Level Defects*

As described in Sect. 4.4.1, in AlGaAs DH LEDs, increase in the concentration of deep-level defects A and B was found to be strongly associated with the gradual degradation. Therefore, one can expect that the gradual degradation is caused by the increase in the deep-level defects during operation. Although the detailed mechanism for the gradual degradation in this case is not clear at present, one can conclude that if such deep-level defects are present in as-grown crystals, gradual degradation can be enhanced during operation of optical devices fabricated from these crystals. Since deep-level defects A and B are considered to be associated with point defect complexes, deep levels related to single impurities may not be associated with the gradual degradation. From these considerations, it is concluded that in order to eliminate the gradual degradation, we must suppress the introduction of deep-level defects during crystal growth and/or device fabrication processes (i.e., ion implantation, dry etching, etc.). Table 4.4 shows the results of characterization of deep levels in various epitaxial III-V alloy semiconductors. From these results, we can predict the possibility of gradual degradation in optical devices fabricated from various materials for the active layer.

**Table 4.4** Deep levels in various epitaxial alloy semiconductors

| Material | Growth method | Trap level (type) (eV) | Possible origin | Ref. |
|---|---|---|---|---|
| AlGaAs | LPE | 0.43 (H) | $Ga_{As}$-related | [8] |
| | | 0.87 (H) | $Ga_{As}$-related | [8] |
| | MOVPE | 0.25 (E) | Not clarified | [59, 60] |
| | | 0.62 (E) | Al-impurity complex | [59, 60] |
| | | 0.82 (E) | EL2 | [59, 60] |
| | MBE | 0.42 (E) | DX center | [59, 60] |
| | | 0.55 (E) | Native point defects | [61–63] |
| | | 0.65 (E) | Impurity complex | [61–63] |
| GaAsP | VPE | 0.16 (E) | DX center | [64] |
| | | 0.38 (E) | Recombination center | [65] |
| InAlAs | MBE | 0.27 (H) | Native point defects | [66] |
| | | 0.45 (H) | Impurity complex | [66] |
| | | 0.65 (H) | Native point defects | [66] |
| | | 0.95 (H) | Impurity complex | [66] |
| | | 0.35 (E) | DX center | [66] |
| | | 0.54 (E) | DX center | [67] |
| InGaAs | LPE | 0.69 (PL) | Due to lattice mismatch | [68] |
| | | 0.30–0.32 (E) | $V_{As}$-impurity complex | [69, 70] |
| InAlP | MOVPE | 0.29 (E) | DX center | [71] |
| | | 0.66 (E) | $V_{In}$, $V_{Al}$ complexes | [71] |
| InGaP | LPE | 0.39 (E) | $V_P$-related | [72] |
| | | 0.61 (E) | Due to lattice mismatch | [72] |
| | VPE | 0.32 (E) | DX center | [73] |
| | | 0.37 (E) | DX center | [73] |
| | MOVPE | 0.84 (H) | Due to lattice mismatch | [74] |
| InAlGaAs | MBE | 0.64 (E) | Not clarified | [75] |
| | | 0.16 (H) | Not clarified | [75] |
| InGaAsP | LPE | 0.28–0.31 (E) | $V_P$-complex | [76] |
| | | 0.38–0.40 (E) | $V_{As}$-complex | [72] |

### *4.7.2 Reduction of Residual Impurities*

When the gradual degradation occurs in AlGaAs DH LEDs, in addition to the increase in deep-level defects in the active region, high density of dislocation loops is also generated as by-products of the point defect reaction [8]. Similar dislocation loops are also generated in InGaAsP/InP buried heterostructure lasers [43, 45] and strained InGaAs/AlGaAs MQW lasers [58]. It is expected that these dislocation loops are generated by condensation of excess native point defects (i.e., vacancies or interstitials through non-radiative recombination process of minority carriers). In this process, the condensation of native point defects may occur at some nucleation centers. Residual impurities can be one of the candidates. Therefore, one must reduce the concentration of such impurities during crystal growth.

### 4.7.3 Selection of Appropriate Dopant Element and Control of Dopant Concentration

Since one of the major phenomena governing the gradual degradation of optical devices is diffusion of dopant atoms, one cannot perfectly avoid the degradation. However, we may eliminate the degradation rate by selecting appropriate dopant. In the case of diodes in which Zn-doped p-cladding layer is used, we observe the formation of dislocation loops just outside of the stripe region during the accelerating testing at high temperatures. This may be due to the fact that recombination-enhanced diffusion of Zn-interstitials easily occurs during the operation of lasers and LEDs. Therefore, in order to eliminate the degradation rate, it is important to select appropriate dopant atoms other than Zn (C, Be, Mg, etc.) and to control the dopant concentration as low as possible.

### 4.7.4 Elimination of Stress

As described in Sect. 4.6, the gradual degradation is unexpectedly accelerated by the presence of internal stress in AlGaAs DH lasers (see Sect. 4.6.1) and InGaAs/AlGaAs MQW lasers (see also Sect. 4.6.2). Moreover, it is suggested that there is a threshold value in the stress for causing the acceleration of gradual degradation in the former case: when the compressive stress in the stripe-edge region of the active layer exceeds $3 \times 10^8$ dyne/cm$^2$, then the gradual degradation starts to occur. Therefore, we must eliminate the internal/external stress around the active region. In order to eliminate the internal stress, it is essential to achieve appropriate device designing. In this case, it is useful to simulate the stress distribution around the active region of the optical device by finite element method [50].

### 4.7.5 Elimination of Diffusion of Metallic Atoms from the p-Electrode

In the optical devices such as InGaAsP/InP DH LEDs, where the gradual degradation due to the increase in the deep-level defects and/or generation of dislocation loops hardly occurs, one of the dominant factors governing the degradation could be the diffusion of metallic atoms from the *p*-electrode (i.e., gold in the case of optical devices with Au-based p-electrode) [39–42]. In this case, it is expected that degradation takes place over the long term of operation by the diffusion of the metallic atoms toward the active region, causing a gradual increase in the non-radiative recombination in the active region. To eliminate this degradation mode, one must basically use non-alloy ohmic contact with an appropriate barrier metal. At present, Ti/Pt/Au is thought to be the best electrode material for p-contact for optical devices.

## 4.8 Summary

This chapter focuses on classical research into rapid degradation and on advances in research into gradual degradation. It is anticipated that demand for semiconductor optical devices (including VCSELs) will increase even further in the future, making their long-term reliability increasingly important. To this end, it is vital that device manufacturers and equipment manufacturers cooperate in the development of new devices and systems, elucidate degradation mechanisms, and attempt to prevent field failure.

Meanwhile, although the efforts of numerous researchers have already yielded a certain degree of understanding of the physics of the phenomenon of degradation, many unanswered questions still remain. It is hoped that research into degradation mechanisms will progress further in the future—to include resolution of the phenomenon at the atomic level.

## References

1. O. Ueda, J. Electrochem. Soc. **135**, 11C (1988)
2. O. Ueda, *Reliability and Degradation of III-V Optical Devices* (Artech House Publishers, Norwood, Boston, MA and London, England, 1996)
3. M. Fukuda, *Reliability and Degradation of LEDs and Semiconductor Lasers* (Artech House Publishers, Boston, MA and London, England, 1991)
4. P.M. Petroff, R.L. Hartman, Appl. Phys. Lett. **23**, 469 (1973)
5. S. Kishino, H. Nakashima, R. Ito, O. Nakada, Appl. Phys. Lett. **27**, 207 (1975)
6. O. Ueda, H. Imai, T. Fujiwara, S. Yamakoshi, T. Sugawara, T. Yamaoka, J. Appl. Phys. **51**, 5316 (1980)
7. P.M. Petroff, Semicond. Insul. **5**, 307 (1983)
8. K. Kondo, O. Ueda, S. Isozumi, S. Yamakoshi, K. Akita, T. Kotani, IEEE Trans. Elect. Device **ED-30**, 321 (1983)
9. C.H. Henry, P.M. Petroff, R.A. Logan, F.R. Meritt, J. Appl. Phys. **50**, 3721 (1979)
10. O. Ueda, H. Imai, T. Kotani, K. Wakita, H. Saito, J. Appl. Phys. **50**, 6643 (1979)
11. O. Ueda, K. Wakao, A. Yamaguchi, S. Komiya, S. Isozumi, I. Umebu, J. Appl. Phys. **58**, 3996 (1985)
12. O. Ueda, S. Yamakoshi, T. Sanada, I. Umebu, T. Kotani, O. Hasegawa, J. Appl. Phys. **53**, 9170 (1982)
13. O. Ueda, S. Yamakoshi, I. Umebu, T. Sanada, T. Kotani, J. Appl. Phys. **54**, 6732 (1983)
14. O. Ueda, Nikkei Electron. **5-5**, 152 (2008) (in Japanese)
15. O. Ueda, S. Isozumi, T. Kotani, T. Yamaoka, J. Appl. Phys. **48**, 3950 (1977)
16. K. Ishida, T. Kamejima, J. Electron. Mater. **8**, 57 (1979)
17. O. Ueda, K. Wakao, A. Yamaguchi, S. Isozumi, S. Komiya, J. Appl. Phys. **57**, 1523 (1985)
18. H. Saito, T. Kawakami, IEEE J. Quantum Electron. **QE-13**, 564 (1977)
19. O. Ueda, I. Umebu, S. Yamakoshi, T. Kotani, J. Appl. Phys. **53**, 2991 (1982)
20. K. Ishida, T. Kamejima, Y. Matsumoto, K. Endo, Appl. Phys. Lett. **40**, 16 (1982)
21. P.M. Petroff, L.C. Kimerling, Appl. Phys. Lett. **29**, 461 (1976)
22. S. O'Hara, P.W. Hutchinson, P.S. Dobson, Appl. Phys. Lett. **30**, 368 (1977)
23. R.D. Gold, L.R. Weisberg, Solid State Electron. **7**, 811 (1964)
24. O. Ueda, S. Isozumi, S. Yamakoshi, T. Kotani, J. Appl. Phys. **50**, 765 (1979)

25. B. Monemar, G.R. Woolhouse, Appl. Phys. Lett. **29**, 65 (1976)
26. O. Ueda, S. Yamakoshi, T. Yamaoka, Japan. J. Appl. Phys. **19**, L251 (1980)
27. Y.M. Cheng, R.W. Herrick, P.M. Petroff, M.K. Hibbs-Brenner, R.A. Morgan, Int. Reliab. Phys. Symp. **34**, 211 (1996)
28. For example, A Thesis by David T. Mathes, University of Virginia, p. 98 (also Figs. 4-30 and 4-31 shown in pages 100 and 101, respectively) (2002)
29. S. Yamakoshi, T. Sugahara, O. Hasegawa, Y. Toyama, H. Takahashi, *IEDM Technical Digest,* December, 1978, p. 642
30. S. Yamakoshi, O. Hasegawa, H. Hamaguchi, M. Abe, T. Yamaoka, Appl. Phys. Lett. **31**, 627 (1977)
31. D.V. Lang, R.A. Logan, J. Electron. Mater. **4**, 1053 (1975)
32. D.J. Mazey, R.B. Barns, A. Howie, Philos. Mag. **7**, 1861 (1962)
33. G. Thomas, M.J. Goringe, *Transmission Electron Microscopy of Materials* (Wiley, New York, 1979)
34. I. Hayashi, in *Proceedings of 15th International Conference on Physics of Semiconductors*, J. Phys. Soc. Jpn. **A57**(Suppl) (1980)
35. G.M. Glom, J. Cryst. Growth **36**, 125 (1976)
36. O. Ueda, S. Yamakoshi, S. Komiya, K. Akita, T. Yamaoka, Appl. Phys. Lett. **36**, 300 (1980)
37. K. Wakita, H. Takaoka, M. Seki, M. Fukuda, Appl. Phys. Lett. **40**, 525 (1982)
38. P.M. Petroff, SPIE **2683**, 52 (1996)
39. S. Yamakoshi, M. Abe, S. Komiya, Y. Toyama, IEDM Tech. Dig., December 1979, p.122
40. S. Yamakoshi, M. Abe, O. Wada, S. Komiya, T. Sakurai, IEEE **QE-17**, 167 (1981)
41. A.K. Chin, C.L. Ziepfel, S. Mahajan, F. Ermanis, M.A. DiGiseppe, Appl. Phys. Lett. **41**, 555 (1982)
42. S. Mahajan, A.K. Chin, C.L. Ziepfel, D. Brasen, B.H. Chin, R.T. Tung, S. Nakahara, Mater. Lett. **2**, 184 (1984)
43. S.N.G. Chu, S. Nakahara, M.E. Twigg, L.A. Koszi, E.J. Flynn, A.K. Chin, B.P. Segner, W.D. Johnston Jr., J. Appl. Phys. **63**, 611 (1988)
44. O. Ueda, Jpn. J. Appl. Phys. **49**, 090001 (2010)
45. B.C. de Cooman, C.W.T. Bulle-Lieuwma, J.A. de Poorter, W. Nijman, J. Appl. Phys. **67**, 3919 (1990)
46. I. Mito, M. Kitamura, K. Kobayashi, S. Murata, M. Seki, Y. Odagiri, H. Nishimoto, M. Yamaguchi, K. Kobayashi, J. Lightwave Technol. **1**, 195 (1983)
47. U. Goesele, F. Morehead, J. Appl. Phys. **52**, 4617 (1981)
48. K. Endo, K. Kobayashi, H. Fujii, Y. Ueno, Appl. Phys. Lett. **64**, 146 (1994)
49. S. Tomiya, T. Hino, S. Goto, M. Takeya, M. Ikeda, IEEE Sel. Top. Quantum Electron. **10**, 1277 (2004)
50. M. Ikeda, O. Ueda, S. Komiya, I. Umebu, J. Appl. Phys. **58**, 2448 (1985)
51. G.H. Olsen, M. Ettenberg, J. Appl. Phys. **48**, 2543 (1977)
52. E.D. Pierron, D.L. Parker, J.B. Mcneely, J. Appl. Phys. **38**, 4669 (1967)
53. M. Ettenberg, R.J. Paff, J. Appl. Phys. **41**, 3926 (1970)
54. P.A. Kirkby, P.R. Selway, J. Appl. Phys. **50**, 4567 (1979)
55. L.D. Landau, E.M. Lifshitz, *Theory of Elasticity* (Pergamon, New York, 1970)
56. B. Wakefield, J. Appl. Phys. **50**, 7914 (1979)
57. M.J. Robertson, B. Wakefield, P.W. Hutchinson, J. Appl. Phys. **52**, 4462 (1981)
58. C. Frigeri, J.L. Weyer, M. Baeumler, S. Muller, W. Jantz, J. Luft, G. Herrmann, W. Spath, *Institute of Physics Conference Series No. 160*. IOP Publish, Bristol, England, p. 483 (1997)
59. R.H. Wu, D. Allsop, A.R. Peaker, Electron. Lett. **18**, 75 (1982)
60. P.K. Bhattacharya, S. Subramanian, M.J. Ludowise, J. Appl. Phys. **55**, 3664 (1984)
61. K. Hikosaka, T. Mimura, S. Hiyamizu, *Institute of Physics Conference Series No. 63*. IOP Publish, Bristol, England, p. 233 (1981)
62. S. Naritsuka, K. Yamanaka, M. Mihara, M. Ishii, Jpn. J. Appl. Phys. **23**, L112 (1984)
63. P.M. Mooney, R. Fischer, H. Morkoc, J. Appl. Phys. **57**, 1928 (1985)

64. E. Calleja, E. Munoz, B. Jimenez, A. Gomez, F. Garcia, F. Kellert, J. Appl. Phys. **57**, 5295 (1985)
65. M. Kaniewska, J. Kaniewski, J. Appl. Phys. **63**, 1086 (1988)
66. W.-P. Hong et al., J. Electron. Mater. **16**, 271 (1987)
67. K. Nakashima et al., Phys. Stat. Sol. **A103**, 511 (1987)
68. T. Yagi, Y. Fujiwara, T. Nishino, Y. Hamakawa, Jpn. J. Appl. Phys. **22**, L467 (1983)
69. Z. Chen et al., *Proceedings of Conference on Semi-Insulating III-V Materials* (IOP Publish, Malmo, 1988), p. 231
70. S. Laualiche et al., Appl. Phys. Lett. **51**, 1361 (1987)
71. M.O. Watanabe, Y. Ohba, J. Appl. Phys. **60**, 1032 (1986)
72. Zhu Q-S, K. Hiramatsu, N. Sawaki, I. Akasaki, Jpn. J. Appl. Phys **28**, L1326 (1989)
73. K. Kitahara, M. Hoshino, M. Ozeki, Jpn. J. Appl. Phys. **27**, L110 (1988)
74. J.F. Chen et al., J. Appl. Phys. **67**, 3711 (1990)
75. D. Biswas et al., J. Appl. Phys. **67**, 2450 (1990)
76. Y. Sasai, Y. Yamazoe, M. Okuyama, T. Nishino, Y. Hamakawa, Jpn. J. Appl. Phys. **18**, 1415 (1979)

# Chapter 5
# Catastrophic Optical Damage in High-Power, Broad-Area Laser Diodes

**Aland K. Chin and Rick K. Bertaska**

**Abstract** Catastrophic optical damage (COD) is semiconductor material within the optical cavity of laser diodes that is thermally damaged by the laser light. COD results in the failure of laser diodes. The phenomena of COD in high-power, broad-area laser diodes are described along with methods to eliminate it.

## 5.1 Introduction

The semiconductor laser diode (LD) was demonstrated in 1962 [1–4]. By 1966, catastrophic optical damage (COD), the first failure mode of LDs, was reported [5, 6]. COD is semiconductor material that is thermally damaged by the laser light; the damaged material is contained within the optical cavity. For COD to occur, a local region of the optical cavity must have a combination of sufficiently high optical absorption and sufficiently high optical intensity that the local temperature quickly increases to the melting temperature of the material. On hindsight, it is not surprising that COD was the first failure mode since the first LDs could only operate in the pulse mode and the LDs were driven to peak powers up to 130 W [6]. Researchers found irreversible changes in device performance associated with surface damage at the facets of the LDs when the critical optical power density was exceeded [5, 6].

Since COD was first observed, there has been interest in understanding this phenomenon as well as in methods to eliminate it. COD is the only known failure

A.K. Chin (✉)
Somerville Laser Technology, 15 Ward St., Somerville, MA 02143, USA
e-mail: aland@slt-lasers.com

R.K. Bertaska
New England Analytical, 1 Chestnut St, Nashua, NH 03060, USA

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_5,
© Springer Science+Business Media New York 2013

mode remaining for broad-area laser diodes (BA-LDs), and it is the opinion of the authors that COD will always remain since (1) it occurs when the optical power density locally exceeds a critical value that is material dependent and (2) there is always a demand for higher optical power density.

Beginning in 1987, there was strong interest in improving the reliability of LDs emitting in the 9XXnm spectral region. Single transverse-mode LDs, with ultrahigh reliability, were required to optically pump, single-mode, erbium-doped fiber amplifiers (EDFA) [7, 8] for telecommunication applications. It is desirable to optically pump EDFAs with LDs operating at ~980 nm since such amplifiers have lower noise figures relative to optical amplifiers pumped at other wavelengths [9].

Coincidentally, reliable, high-brightness, multi-spatial mode (broad-area) LDs were required as optical pumps for high-power, double-clad, fiber lasers [10] used in numerous commercial applications, e.g., free-space communication, printing, and material processing. Ytterbium-doped fiber lasers emitting at ~1,080 nm and pumped with BA-LDs emitting in the 9XXnm spectral region resulted in high-power conversion efficiency due to the small quantum defect between pump and laser. However, 9XXnm LDs suffer from COD at less than desirable optical power densities. Due to these two large commercial applications, there are significantly more studies of COD occurring in LDs emitting at 9XXnm relative to LDs emitting at other wavelengths. Although the values of various parameters associated with COD may change with wavelength, we believe the description of COD in this chapter, based mainly on 9XXnm LDs, is applicable to LDs at other wavelengths. We also believe that formation of COD in single-mode LDs is the same as in broad-area laser diodes (BA-LDs) except that, by definition, the optical cavity of a single-mode LD cannot support internally circulating modes as discussed in Sect. 5.3.1.

On a cursory level, the phenomenon of COD appears complex due to the different measurements and analytical images of COD in BA-LDs of different materials and in BA-LDs with improved design and fabrication. The first BA-LDs were homo-junctions in bulk GaAs crystals which had poor performance due to poor electrical and optical confinement both parallel and perpendicular to the plane of the p-n junction [1–4]. The facets (mirrors of the Fabry-Perot cavity) were formed by polishing or cleaving. Optical coatings used to protect the facet or to modify the facet reflectivity were often not present. Electrical contacts to both the p-side and n-side of the BA-LD were broad-area contacts that provided no electrical or optical confinement. Subsequently, epitaxial layers were incorporated to improve electrical and optical confinement perpendicular to the plane of the p-n junction. Methods to deposit epitaxial layers changed from liquid phase epitaxy (LPE) to molecular beam epitaxy (MBE), metal-organic chemical vapor deposition (MOCVD), and various combinations. Facet-formation procedures were improved, and optical coatings were instituted and subsequently improved. Electrical confinement and optical confinement in the plane of the p-n junction were incorporated.

It is beyond the scope of this chapter to describe, in detail, BA-LD structures and fabrication procedures. It is also beyond the scope of this chapter to present a review of all the extensive studies, over more than four decades of research and development, of the phenomenon of COD observed in the multitudinous types of laser diodes. In this chapter, the authors attempt to provide a simple, consistent model of the formation of COD in BA-LDs, supported by direct observations, to allow researchers to continue to improve LD performance and reliability. Modifications [11, 12] discussed in Section 5.3.1, based on a portion of this model, to the device design and fabrication procedure have already significantly improved laser diode performance and reliability.

## 5.2 Time Line

The following provides a chronological list of significant events related to the phenomena of catastrophic optical damage (COD) in laser diodes.

*1962* The semiconductor laser diode (LD) is demonstrated [1–4].
*1966* Failure of homo-junction GaAs LDs, pulsed to high optical output power, by catastrophic optical damage (COD) is reported [5, 6].
*1972* The presence of internally circulating modes in double-heterostructure (DH) GaAlAs/GaAs LDs is reported [13, 14].
*1974* An extensive study of COD in DH GaAlAs/GaAs LDs is presented [15].
*1977* Failure of DH GaAlAs/GaAs LDs by COD attributed to internally circulating modes (IC) is reported [16, 17].
*1979* An extensive description of the COD phenomenon in optically pumped, DH GaAlAs/GaAs epitaxial material is reported [18].
*1983* A study of COD in optically pumped InGaAsP/InP epitaxial material is reported [19].
*1985* Impurity-induced disordering (IID) of quantum well(s) to create a transparent window at the facets is invented [11].
*1989* The optical density at which COD occurs is found to decrease over time as a result of chemical reaction of the surface with the ambient [20].
*1991* The ultrahigh vacuum, cleave-and-coat fabrication procedure (E2 process) to passivate LD facets is invented [12].
*1995* COD in DH InGaAsP/InP LDs is reported [21].
*1998* COD in aluminum-free, separate-confinement, DH, double quantum-well LDs is reported [22].
*2009* COD due to the presence of IC modes in quantum-well BA-LDs is reported [23].
*2009* COD initiating within the optical cavity of 980-nm BA-LDs is found to be related to an unintentional fabrication error (solder void) [24].

## 5.3 Detailed Description of COD

### *5.3.1 Cavity Modes*

Much of the information on COD is obtained from postmortem images of failed LDs since CODs initiate in less than 18 ns [18] and propagate at speeds as high as 20 m/s [25]. Henry et al. [18] initiated and monitored the growth of CODs in GaAlAs/GaAs DH wafers using 18-ns optical pulses at a repetition rate of 250 Hz from an external cavity-dumped Ar-ion laser at 5,145 Å. While not stated explicitly, we believe that a single optical pulse was sufficient to initiate a COD since the thermal time constants of their GaAlAs/GaAs DH wafer are estimated to be much less than the 4 ms between external laser pulses.

As will be discussed further in Sect. 5.3.4, COD propagates in the direction of the stimulated emission [18]. The COD becomes a permanent image of the optical mode that caused the material damage when the intensity of the optical mode was sufficient to initiate and maintain a molten region within the optical cavity. For convenience, we briefly describe transverse and internally circulating (IC) modes within the optical cavity.

Figure 5.1a is a top view of the optical cavity of a BA-LD. The front and rear facets form the mirrors of a Fabry-Perot cavity. In the plane of the p-n junction, in the direction perpendicular to the optical axis, optical gain is confined to the cavity by restricting the electrical contact.

Transverse modes reflect between the front and rear facets. Transverse mode filaments [26] (linear regions of locally higher current and optical density) responsible for the peaks in the near-field intensity profile (labeled a–d) are indicated by the arrows. A portion of the energy within the optical cavity is transmitted though each facet depending on the reflectivity of each facet [27].

Internally circulating (IC) modes [13, 14, 16, 17, 23] are less well known, in part because they are entirely contained within the optical cavity by total internal reflection and their energy is entirely dissipated within the laser. Figure 5.1b is the same as Fig. 5.1a with the addition of an example of an IC mode. The IC modes totally reflect off of the front and rear facets with an angle of incidence equal to the critical angle (~16° for GaAs).

The presence of IC modes has been inferred mainly from a bimodal distribution of LD efficiency [14, 16, 17] within a batch of diodes; IC modes are operating in the portion of the distribution with the lower efficiency. Additionally, intentional design changes to eliminate IC modes appeared to be successful [14, 16, 17] until 2009 when Chin et al. [23] reported the contribution of IC modes in BA-LDs that failed by COD. As discussed further in Sect. 5.3.4, Chin et al. [23] inferred the presence of IC modes, after reviewing numerous, published postmortem images of BA-LDs that failed by COD, and observed that portions of many CODs propagate at an angle of ~16° to the optical axis.

The gain of IC modes relative to the transverse modes depends on a number of device parameters:

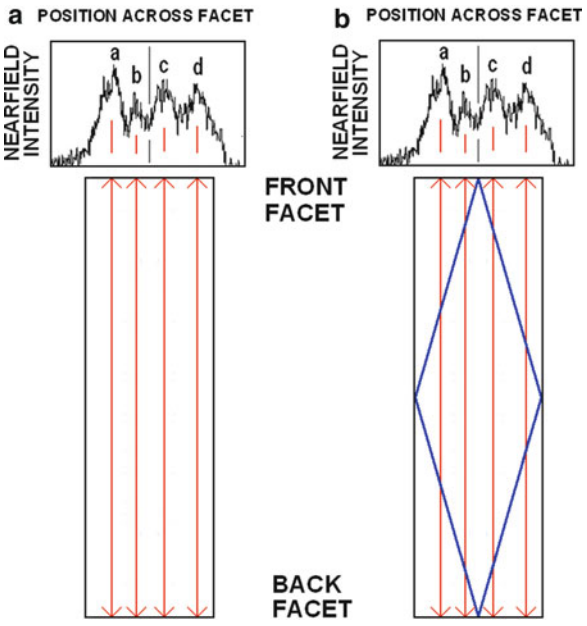



**Fig. 5.1** (**a**) Top view of Fabry-Perot cavity with transverse cavity mode filaments (*vertical arrows*) responsible for near-field intensity peaks labeled *a–d*. (**b**) Image in Fig. 5. 1 (a) with the addition of a temporary internally circulating (IC) mode that has no effect on the near-field intensity profile. The IC mode can be propagating clockwise or counterclockwise

- Ratio of the emission aperture divided by the cavity length [14]
- Front-facet and back-facet mirror loss [13]
- Side-wall reflectivity [14, 17]

The ratio ($R$) of the emission aperture width divided by the cavity length in BA-LDs reported by Ettenberg et al. [14] is in the range of 0.20–0.34. IC modes are suppressed in lasers with low values of $R$. Recent BA-LDs generally have emission aperture widths in the range of ~50 to ~100 μm and a cavity length in the range of ~1 to ~5 mm. For these devices, values of $R$ are in the range of 0.01–0.10, and thus IC modes should be suppressed by the device geometry.

The optical loss for transverse modes in a Fabry-Perot cavity depends on the front-facet and back-facet mirror reflectivity, whereas reflection of IC modes at the facets is due to total internal reflection [13, 14, 16, 17, 23]. IC modes appear for values of $\{F = (1/2L)\ln(1/R_1R_2)\}$ which are more than ~50 $cm^{-1}$, where $L$ is the cavity length and $R_1$ and $R_2$ are the front and back reflectivity, respectively [13]. Recent BA-LDs generally have $R_1$ ~0.05, $R_2$ ~0.95, and $L$ in the range of 1–5 mm. For these devices, $F$ is in the range of 3–15 $cm^{-1}$, and thus IC modes should not appear.

The reflection at the sidewalls of early sawn-cavity BA-LDs is believed to be due to micro-facets, ~6 μm in dimension [14]. With this model of IC modes, Henshall [17]

increased side-wall reflection losses to suppress IC modes by proton bombarding the sides. Proton bombardment affects both the electrical and optical properties of GaAs [28], but Henshall [17] did not state the specific effect leading to optical loss. In modern dielectric isolated BA-LDs, reflections at the sidewalls of the optical cavity are believed to be due to total internal reflection [23].

### 5.3.2 Initiation

During initial operation of a BA-LD, COD generally occurs at the front facet of the laser where the optical power density locally exceeds the COD limit ($P_{COD}$) of the facet [5, 6, 15, 18]. To ensure long-term, reliable operation, LDs are generally operated at a power level in the range of 30–50 % below the initial value of $P_{COD}$. However, for LDs with *unpassivated* facets, optical absorption at the facets continues to increase due to surface chemical and photochemical reactions, and thus $P_{COD}$ decreases with time [20, 29, 30]. Bou Sanayeh et al. [31] correlated the COD locations at the front facet of a BA-LD with locations where the near-field intensities and consequently the local temperatures are highest. The high local temperatures presumably accelerated the decrease in local $P_{COD}$ [20, 29, 30] of the unpassivated facet.

If the facets are *passivated*, $P_{COD}$ is stable in time during device operation, and a BA-LD should not fail by COD if it is operated significantly below $P_{COD}$. However, extensive studies of passivated BA-LDs do show the occurrence of random failures by COD [32, 33]. Fabrication errors, such as pinholes in the silicon thin film used to protect the facets in the E2 process [12] or material defects generated during the IID process [11], undoubtedly account for some, but not all, of the failures of LDs by COD during long-term operation. LDs that fail due to fabrication errors should be batch related, and this was not reported [32, 33].

The authors have proposed that the random failure by COD of BA-LDs with passivated facets is due to the interaction of transverse modes and ring-cavity modes [23]. The ring-cavity modes are the same as the internal circulating (IC) modes discussed in Sect. 5.3.1. To avoid confusion, ring-cavity modes will be referred to as IC modes in this chapter due to priority.

Figure 5.1b illustrates how we believe an IC mode initiates a COD. Support for this model is provided in ref. [23]. Transverse modes are generally present in the optical cavity of a BA-LD. The appearance of IC modes is clearly temporary; otherwise, the slope efficiency would be anomalously low [13, 14, 16, 17].

The pulse power at which a BA-LD fails by COD depends on the pulse duration. Figure 5.2 illustrates the general trends of $P_{COD\text{-}pulsed}$, the power at which the laser fails by COD under pulsed conditions, as a function of pulse duration [34]. For pulse durations more than ~1 μs, likely related to a thermal time constant, $P_{COD\text{-}pulsed}$ is approximately constant. However, $P_{COD\text{-}pulsed}$ increases with shorter pulses, increasing approximately an order of magnitude, as the pulse duration decreases from ~1 μs to ~1 ns.

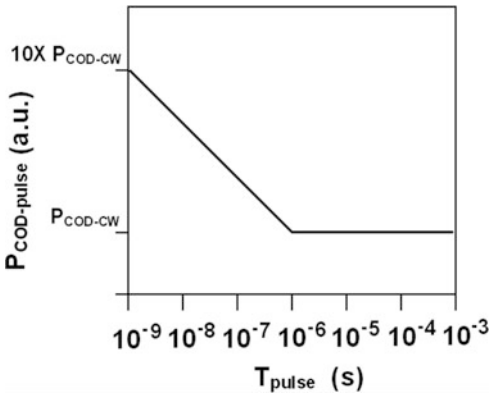



**Fig. 5.2** $P_{COD\text{-}pulsed}$, the power at which COD occurs in a broad-area laser diode during pulse operation, as a function of pulse width

The intensity of transverse filaments has spatial-temporal fluctuations at the relaxation oscillation frequency ($\Omega_{rel}$) and a filament frequency ($\Omega_{fil}$) [35]. The changes in intensity as well as $\Omega_{rel}$ and $\Omega_{fil}$ depend on the output power. For BA-LDs that operate at more than ten times the threshold current, $\Omega_{rel}$ and $\Omega_{fil}$ have values of ~1.0 GHz (1 ns period) and ~5 GHz (0.2 ns period), respectively. At these frequencies, the intensity fluctuations of transverse filaments would have to exceed ten times the CW value of $P_{COD}$ to cause COD. Such large intensity fluctuations of transverse filaments are highly unlikely, and thus failure of passivated BA-LDs operated below $P_{COD}$ is a rare event, leading to the observed high device reliability.

The appearance of IC modes depends on the gain of the IC modes relative to the gain of the transverse modes. We have proposed [23] that during certain transverse-mode fluctuations, IC modes appear. At the front facet, the intensity of the IC mode adds to that of the transverse modes at *only* the location where it reflects off the front facet. If the combined optical intensity at this location is sufficiently high for a sufficiently long time, COD initiates.

Most combinations of transverse mode and IC mode will lack either the intensity or duration to cause COD. However, as the operating current is increased, there will certainly be a value of operating current where COD will occur. Additionally, at a lower value of operating current, after sufficient time, a fluctuation in transverse mode/IC mode, sufficient to cause COD, should also occur. This is clearly a random process and likely accounts for the observed random failures in life tests of passivated BA-LDs [32, 33].

### *5.3.3 Defect Structure*

The melted-resolidified (MR) material resulting from COD is highly non-radiative, resulting in dark lines in images that spatially map the radiative efficiency or minority-carrier lifetimes [15, 18]. The material composition and material quality of COD depend on the materials that were intermixed in the molten state.

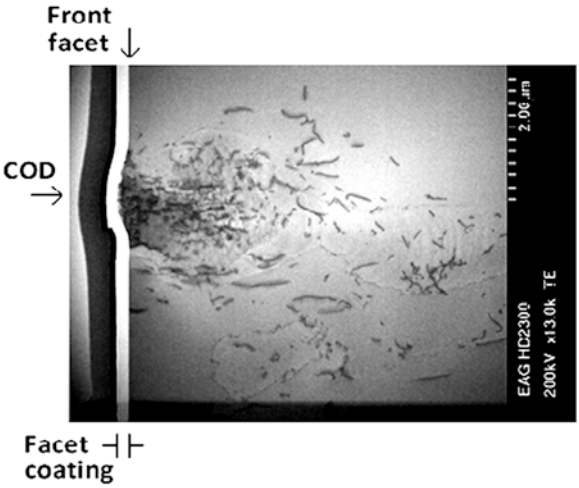



**Fig. 5.3** TEM image of a COD in the plane of the p-n junction. A melted-resolidified region containing numerous material defects is observed

Prior to COD, the epitaxial layers of a BA-LD are defect-free and single crystal. Subsequent to COD, the composition of the material, resolidified under nonequilibrium conditions, is no longer uniform, is unlikely to be lattice matched to the surrounding crystal, and is unlikely to be free of material defects.

While the defect structure of dark-line defects (DLDs) was pertinent to devising methods to eliminate DLDs, the defect structure of COD serves mainly to distinguish it from DLDs which are networks of dislocations oriented along the <110> and <100> directions [36, 37].

As an example, we discuss the defect structure of COD in GaAlAs/GaAs epitaxial structures. CODs in GaAlAs/GaAs BA-LDs contain the least complicated defect structure since the lattice constant of AlAs is only 0.2 % larger than that of GaAs [38]. For failed LDs made of different materials, defects not observed in GaAlAs/GaAs LDs result from lattice mismatch and alloy or compound formation. In InGaAsP/InP 1.3-μm laser diodes, small (20–100 nm) metallic InGa precipitates adjacent to voids have been observed [21]. In aluminum-free, InGaAs/InGaP, 980-nm laser diodes, the COD material is substantially polycrystalline [22].

In GaAlAs/GaAs DH wafers, material analysis [18] of the COD shows crystalline material with four typical types of defects:

- Small dislocation loops (diameter <200 Å) and dipoles separated by defect-free regions
- Large dislocation loops (size between 2,000 and 5,000 Å) elongated along the direction of the COD
- Zones of material with segregated material composition
- Non-radiative regions containing point defects and clusters of point defects

Figures 5.3 and 5.4 are TEM images of COD in the plane of the p-n junction and the plane perpendicular to the p-n junction, respectively, of failed 808-nm GaAlAs/GaAs BA-LDs. In both images, the fine dark lines are dislocation loops or portions

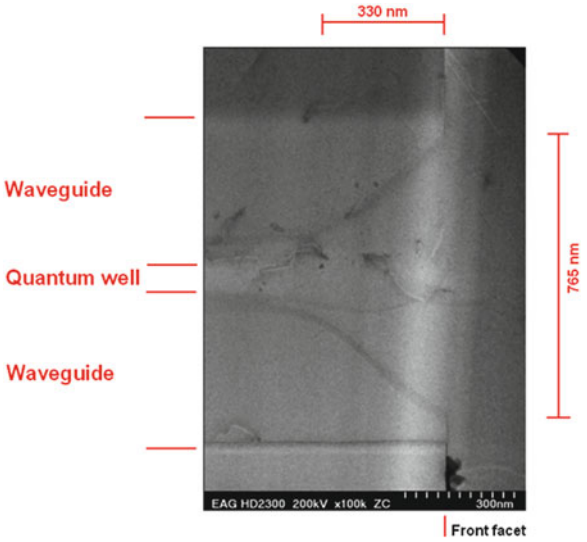



**Fig. 5.4** TEM image of a COD at the front facet in the plane perpendicular to the p-n junction. The COD extends significantly into the waveguide layers at the facet but tapers quickly to approximately the width of the quantum well within ~330 nm

of dislocation loops. Regions of material with segregated material composition are barely visible. There are no dislocation networks oriented along either the <100> or <110> directions as in DLDs.

The MR region in Fig. 5.4 is unexpected and has not been reported previously. At the facet, the MR region extends substantially (~350 nm) into the adjacent waveguides and necks down to the quantum well in a distance of ~330 nm, from the front facet, along the optical axis. The image is unexpected because the waveguides are composed of material that is transparent to the emission from the quantum well. Figure 5.4 indicates that there can be sufficient optical absorption at the facet, depending on the facet preparation procedure, in the region of the transparent waveguides to cause COD.

A review of the phase diagram of the materials in the active layer and surrounding waveguide layers is required to understand the defect structure of the COD. The pseudo-binary phase diagram of GaAs-AlAs [39] is shown in Fig. 5.5. The pseudo-binary phase diagrams of other III-V materials are similar [39]. The liquidus and solidus lines intersect only at the GaAs and AlAs endpoints. The material phase between the liquidus and solidus lines consists of both liquid and solid. A liquid whose composition is not at the endpoints will solidify into a solid of uniform composition if cooled under equilibrium conditions. However, a liquid, if cooled suddenly, will separate into material of various compositions. Lattice mismatch between the various materials due to the nonequilibrium cooling results in the various defects observed in the COD.

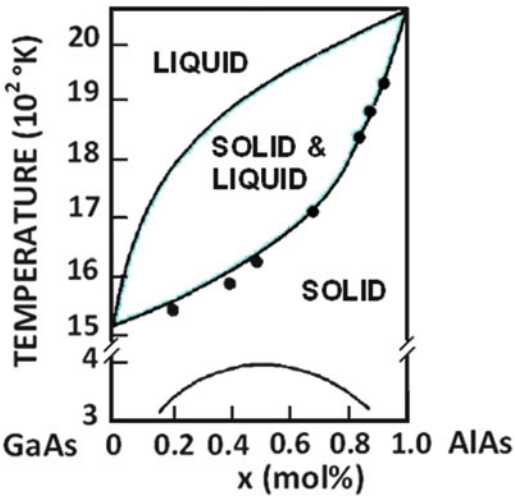



**Fig. 5.5** Pseudo-binary phase diagram of GaAs-AlAs

### *5.3.4 Propagation*

Growth of COD is unrelated to crystallography [18]. As described in Sect. 5.3.3, the COD consists of a locally melted and resolidified region. COD grows toward the incident stimulated emission, consistent with the propagation of a molten zone which moves toward the stimulated emission, the source of energy, as it melts [18, 21].

For COD involving only a transverse-mode filament (TMF), the COD is expected to initiate at the front facet at a location of a TMF, as shown in Fig. 5.6a. The COD propagates roughly linear along the optical axis as depicted in Fig. 5.6b. The surface of the molten region is highly reflective. The surface of the molten region toward the back facet forms a Fabry-Perot cavity, whose length is decreasing as the COD propagates, with the back facet. When there is insufficient optical power in the optical cavity, labeled "A" in Fig. 5.6b, to maintain a molten region, the COD stops propagating.

However, after reviewing recently published postmortem images [23–25, 40–44] of COD in failed BA-LDs, only two images, reported by Sin et al. [40], were found to be similar to Fig. 5.6b. To obtain an image of a COD with a linear track, the IC modes can be intentionally suppressed [25]. The ability to suppress IC modes is also a confirmation of IC modes. A custom 3.6-mm cavity-length, 100-μm aperture, 980-nm laser diode with an aperture fabricated in the n-contact to enable viewing of the optical cavity during device operation was used. The current to the laser diode was ramped at the rate of 1 A/h until device failure by COD at an optical power of ~10 W (~11.6 A). Using fast electronics, the current was intentionally stopped after a 20 % drop in optical output, and the COD had propagated ~20 μm into the optical cavity as shown in Fig. 5.6c, a magnified top view of the optical

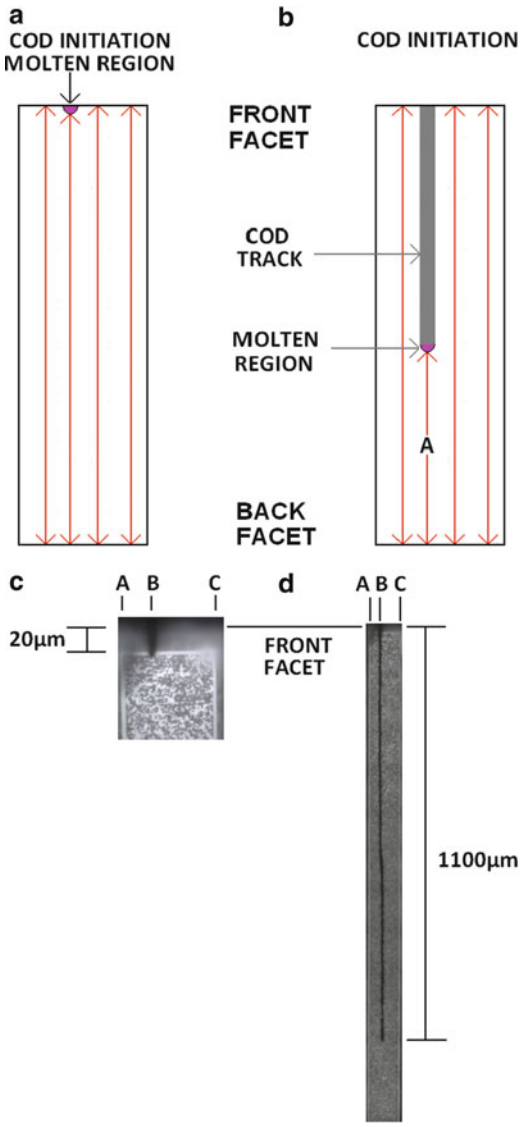



**Fig. 5.6** (**a**) Illustration of a BA-LD optical cavity with a COD initiating at the location of a transverse-mode filament. Filaments are indicated by *arrows* between the front and back facets. (**b**) Illustration of a linear COD track due to a transverse-mode filament. (**c**) Electroluminescence (EL) image of the optical cavity of a BA-LD where a COD was intentionally stopped shortly after it was initiated. (**d**) EL image of the BA-LD in (**c**) after the COD propagated ~1,100 μm into the optical cavity

cavity near the front facet. Locations "A" and "C" are edges of the emission aperture, and location "B" indicates where the COD initiated on the front facet. After obtaining the electroluminescence image in Fig. 5.6c, the device was operated at a current of 3A during which the COD propagated an additional distance of ~1,100 μm along the optical axis as shown in Fig. 5.6d. The absorption from a small COD, as in Fig. 5.6c, is sufficient to suppress IC modes but not the transverse modes due to the generally higher optical loss for IC modes.

Most postmortem images of a COD [23–25, 40–44], regardless of emission wavelength, involve both a transverse mode and an internally circulating (IC) mode as illustrated in Fig. 5.7. The images are complex since they generally contain three types of CODs.

In Fig. 5.7a, a random IC mode is shown reflecting off of the front facet. The intensity of the combined IC and transverse mode is sufficient to initiate a COD at the location where the IC mode reflects off of the front facet. The COD propagates toward the laser radiation and, thus, often has a main trunk along the optical axis, labeled "D" in Fig. 5.7b due to a transverse-mode filament.

Branches of the COD, such as the one labeled "F," are due to IC modes. The presence of an IC mode appears as linear features [23] at an angle of 16° to the optical axis, the critical angle for total internal reflection. In some instances, the linear features are dense enough to be obscured, but the edges of the conglomerates are roughly at an angle of 16° to the optical axis.

There are also instances when the branch reverts back to a direction along the optical axis as in the feature labeled "E." Relative to transverse modes, CODs due to IC modes are shorter in length. Unlike the transverse modes, the COD represents a significant optical loss for the IC mode since it is along its optical path. As the length of the COD increases, the loss for an IC mode increases significantly. However, there may be sufficient optical power in a TMF to further propagate it as in "E."

An example of actual COD branches, observed in all failed diodes from the same batch as the one in Fig. 5.6c, d but operated without suppressing IC modes, is shown in Fig. 5.7c. While images of COD, with features shown in Fig. 5.7b, can be viewed in refs. [23–25] and [40–44], the image provided by Hodges et al. [43] contains all the COD features described above.

The speed of growth ($v_{COD}$) of COD is dependent on material properties as well as optical power density. There are only a few reported values of $v_{COD}$ [15, 18, 19, 25]. Early measurements [15, 18] on GaAlAs/GaAs DH wafers and LDs report $v_{COD}$ in the range of 2–4 m/s, whereas $v_{COD}$ is in the range of 2–4 $\times$ $10^{-4}$ m/s for InGaAsP/InP DH wafers [19]. For 808-nm quantum-well lasers [25], $v_{COD}$ is as high as 20 m/s. These growth speeds are inconsistent with point-defect motion and serve to distinguish CODs from DLDs which grow by point-defect motion and thus have growth speeds that are 5–6 orders of magnitude smaller [45, 46].

The remarkable data in Figs. 5.8, 5.9, and 5.10 demonstrate some of the unique characteristics of failure by COD. Earlier values of $v_{COD}$ [15, 18] were obtained from a sequence of snapshot images where the length of the COD increases either by sequential pulsed operation of a laser diode or pulsed optical excitation of a laser

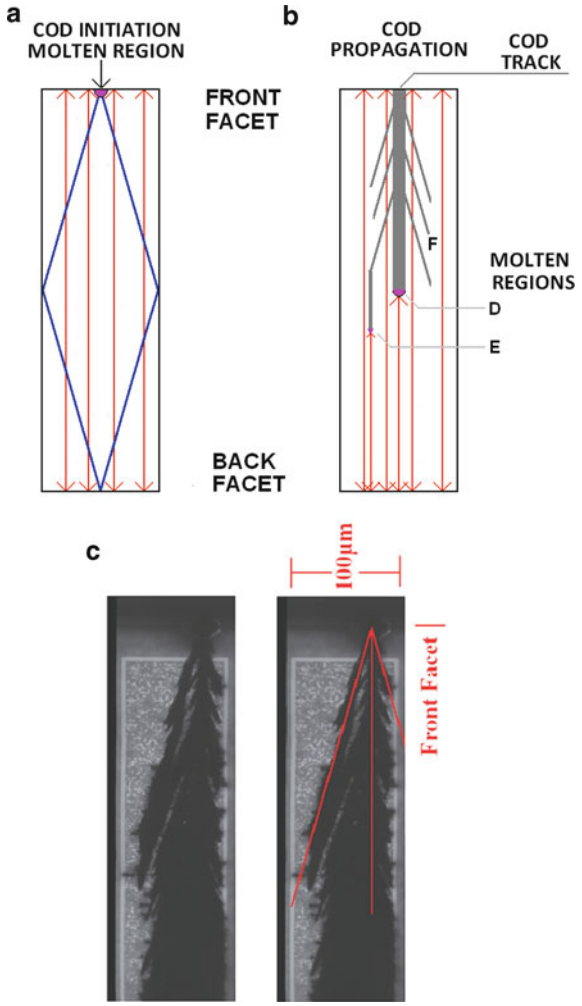



**Fig. 5.7** (**a**) Illustration of a COD initiating due to absorption, at the front facet, of optical power from an internally circulating (IC) mode and transverse modes. *Vertical arrows* represent the transverse-mode filaments causing intensity peaks in the near field. *Lines* at an angle of incidence of 16° to the front facet illustrate an IC mode. (**b**) Illustration of a COD propagating by a combination of transverse and IC modes. Feature "D" is mainly due to a transverse-mode filament. Features "F" are mainly due to IC modes. Feature "E" initiates from a transverse mode, branches due to an IC mode, and changes back to feature along the optical axis due to another transverse mode. (**c**) *Top view* of the 100-µm emission region at the front facet of a 980-nm BA-LD that failed by COD formation. The drawn lines are at an angle of ±16° to the front facet

wafer using an external laser. To capture acceleration as the COD propagates, a high-speed CCD camera was used to image the optical cavity through a window in the n-contact of a 980-nm LD [25]. The propagation of the COD is never intentionally stopped.

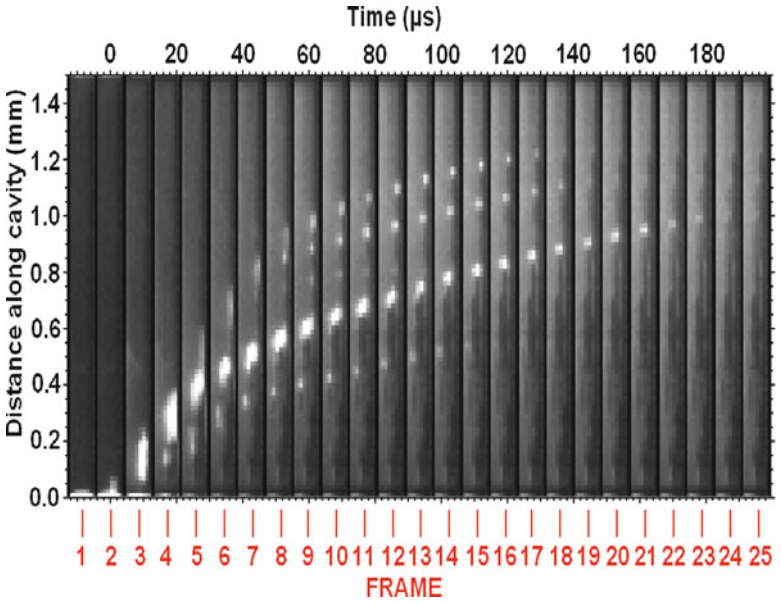



**Fig. 5.8** Twenty-five sequential images of the optical cavity, viewed through a window in the n-contact. Each image is 8.5 μs in duration. The *bright spots* are due to scattering of the laser light from the molten semiconductor region of the COD. The COD initiates at the front facet and propagates toward the back. As it propagates, it separates into at least five branches

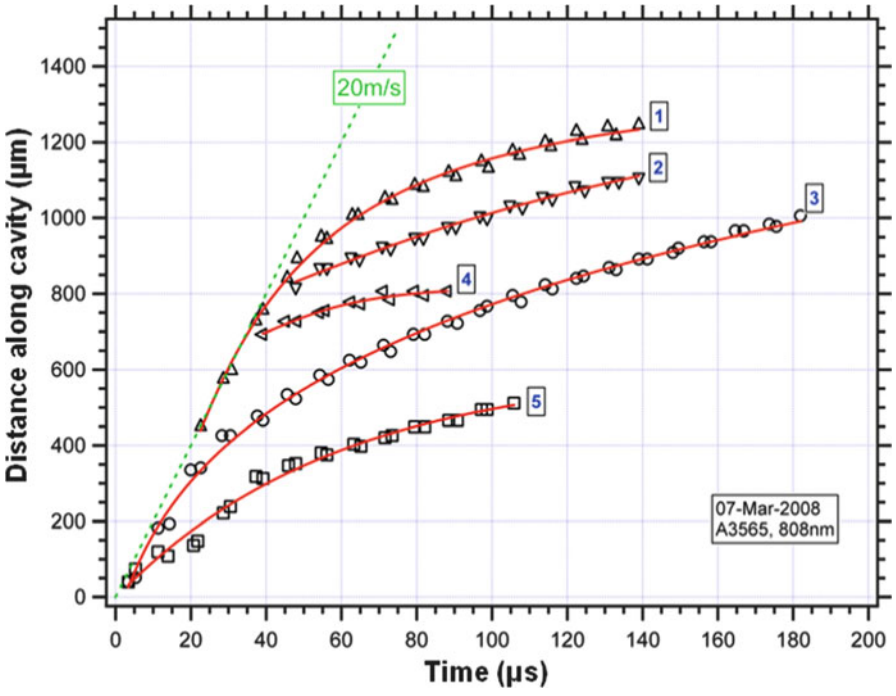



**Fig. 5.9** Distance as a function of time for COD propagation. The device is a 100-μm aperture, 1.5-mm cavity length, 980-nm laser diode

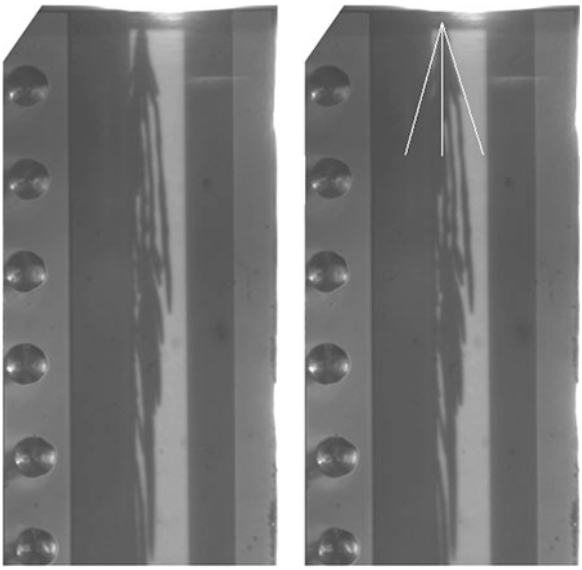


**Fig. 5.10** Postmortem image of the COD in Figs. 5.8 and 5.9. The drawn *lines* are at $\pm 16^\circ$ angles of incidence to the front facet as a visual aid

It is known that COD is a very fast phenomenon [15, 18]. To obtain a fast frame rate without resorting to an expensive streak camera, the image of the optical cavity was limited to 8 pixels wide. Each frame, lasting 8.5 μs, captures the image of the entire optical cavity. Figure 5.8 shows 25 sequential frames of an LD undergoing failure by COD where the first frame is just prior to COD initiation. The molten regions involved in the COD failure scatter the laser light and appear as bright spots in the optical-cavity images.

Figure 5.9 is a plot of the positions of these bright spots within each frame. The track of this COD has five branches. Branching is not expected if only transverse mode filaments are involved. Since COD grows toward the incident radiation, the branching must be due to the combination of transverse and IC modes as described above. As the COD propagates into the optical cavity, the speed, which is proportional to the optical intensity, decreases. The speed of all five tracks approaches a value of 20 m/s at the initiation time of the COD. This value is considerably larger than previously reported [15, 18] values, in the range of 2–4 m/s, in DH GaAlAs/GaAs LDs and wafers where the COD propagates due to only a transverse mode. The initial high speed in Fig. 5.9 is presumably due to the much higher intensity when IC modes combine with transverse modes.

The presence of the IC modes is clearly revealed in the postmortem EL image in Fig. 5.10. The COD forms a $16^\circ$ angle of incidence to the front facet.

### 5.3.5 COD Within the Optical Cavity

COD has been found to initiate within the optical cavity [24, 42, 47] of a BA-LD and not at the front facet. In all three studies, portions of bulk COD propagate at a 16° angle to the optical axis indicating that IC modes were involved. Observation of bulk COD indicates a local increase in the optical absorption within the optical cavity that exceeds that of the front facet. Bulk COD is difficult to study since its presence can only be detected with destructive analytical techniques or LD structures with a window in the n-contact. Unless bulk COD initiates very close to the front facet, they result in a uniform decrease in the near-field intensity profile.

Only Chin et al. [24] have correlated the location of a large ($\geq$5 μm diameter) void in the $Au_{80}Sn_{20}$ solder used to attach 980-nm BA-LDs to the heat sink with the initiation point of a bulk COD. Solder voids lead to a region, within the optical cavity, of elevated temperature which causes the material to locally expand. At the location of the solder void, the quantum well is thus under tension, leading to indium out-diffusion from the quantum well [48] and a local increase in optical absorption and eventual failure by COD.

It should be noted that Sin et al. [42] and Crump et al. [47] also observed bulk CODs in 980-nm BA-LDs. BA-LDs without a high diffusivity component in the active layer are not expected to fail by COD occurring within the optical cavity even when a solder void is present.

## 5.4 Methods to Eliminate COD

The earliest researchers [5, 6, 15, 18] have demonstrated that COD is due to local melting of the semiconductor when there is sufficient absorption of the laser beam. Further investigations indicated that CODs form mainly at the front facet of the LD cavity due to the combination of localized optical absorption and high optical intensity. Initially, the high optical absorption at the facet is caused by surface states from the facet-formation process. Subsequently, the optical absorption at the facet further increases due to chemical interaction with the ambient [12, 20].

Based on the above simple model of COD formation, two elegant facet-formation procedures [11, 12] were invented in an attempt to eliminate this failure mode. Both procedures are highly proprietary, and little is known as to the time of invention and early device results. The long-term success of both procedures in significantly improving LD performance and reliability [32, 33] validates the COD formation model on which the procedures were based.

The first procedure was invented by researchers at IBM [12] and is known as the E2 process. In the E2 process, the facets are formed by first cleaving the LD wafer in ultrahigh vacuum ($<10^{-9}$ Torr) to prevent surface-state absorption from any initial contamination, e.g., oxygen or water vapor, from the ambient [12, 20].

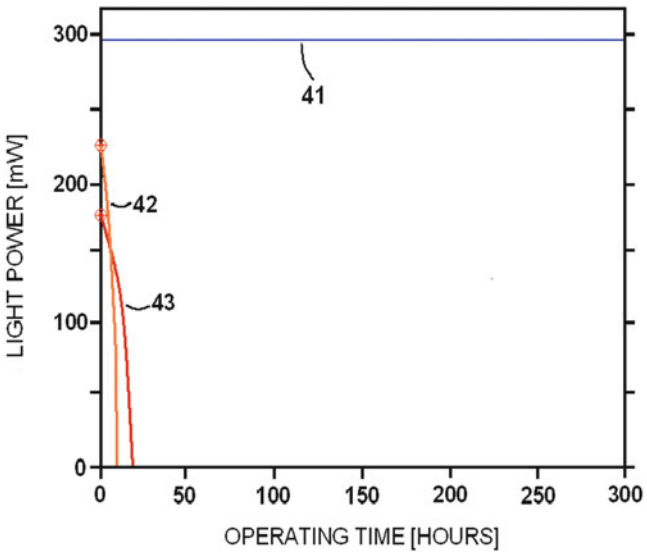



**Fig. 5.11** Aging data for 980-nm, single-mode laser diodes. The facets of diode 41 were passivated with the E2 process. The facets of diode 42 and 43 were not passivated

To prevent subsequent increases in surface-state absorption, the facets were then coated with a thin silicon film, ~20 nm [12], that is

- Transparent to the laser radiation
- Stable against interaction with the LD semiconductor materials [49]
- Impervious to the ambient [49]

Figure 5.11, reproduced from ref. [20], presents the aging data for three single-mode LDs to illustrate the effectiveness of the E2 process in preventing COD. Device 41 is passivated with the E2 process, whereas devices 42 and 43 are fabricated without the E2 process. Device 41 has a higher value of $P_{COD}$ consistent with an uncontaminated facet with a lower optical absorption. The optical output of device 41 is also stable over the 300-h aging time consistent with the protection from the ambient by the silicon film. Relative to device 41, devices 42 and 43 have lower values of $P_{COD}$ and fail by COD within 25 h.

The second procedure is the formation of a transparent, stable, defect-free window at each of the facets by silicon impurity-induced disordering (IID) [11] of the quantum well. Figure 5.12, a TEM image of an IID window in a 980-nm InGaAs/GaAlAs/GaAs laser diode, illustrates the high quality of the material achieved with this process. By diffusing silicon through the InGaAs quantum-well (QW) layer, the QW material intermixes with the material of the adjacent, wide-bandgap GaAlAs waveguide layers, forming a material (InGaAlAs) that has a wider bandgap than the QW. The IID layer significantly reduces the prospect of COD because it is defect-free, transparent to the emission from the QW, and eliminates optical absorption due to contamination from the ambient.

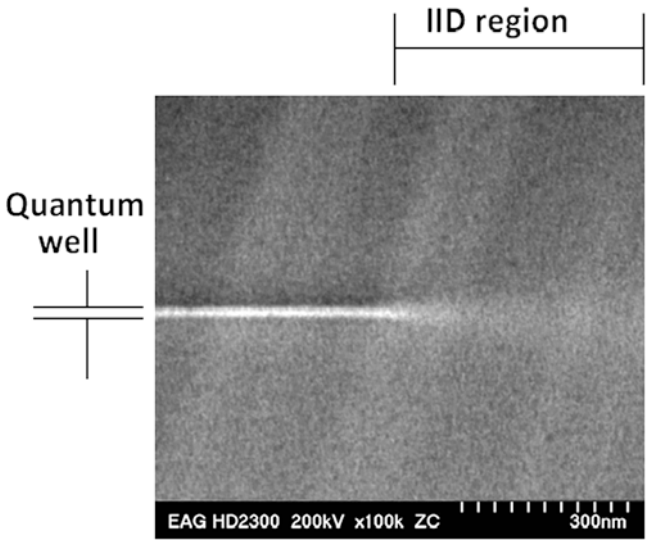



**Fig. 5.12** TEM side view of the interface between an InGaAs quantum well and the region that where the quantum well is disordered by IID

## 5.5 Summary

Catastrophic optical damage (COD) of semiconductor laser diodes is material, within the optical cavity, thermally damaged by the optical intensity of the laser light. The laser light is locally absorbed. The local temperature of the semiconductor material reaches the melting temperature, and COD initiates. If there is sufficient gain to maintain lasing after COD initiation, then the molten region propagates in the direction of the incident laser light, forming a COD track; the material resolidifies on the opposite side since no laser light is present to maintain the molten state. The resolidified material within the COD track is highly nonradiative due to the presence of material defects resulting from the rapid nonequilibrium solidification process. The COD track is often along the <110> direction, but unlike dark-line defects (DLDs), the direction is unrelated to crystallography. Laser diodes from III-V compounds are generally fabricated from (100)-oriented wafers to allow the Fabry-Perot mirrors (facets) to be formed by the (110) cleavage planes. This means that the optical axis of the cavity is along the <110> direction. The CODs that propagate due to transverse modes within such an optical cavity are naturally along the <110> direction.

COD initiates at a local region of high optical absorption and high optical intensity. For properly designed and manufactured optical cavities, the locations of high optical absorption are the front and rear facets. The facet-formation process, generally cleaving, results in broken chemical bonds which leads to optical absorption at the surface. The optical absorption at the surface is further increased by

chemical reaction of the exposed surface to the ambient, most likely oxygen and water vapor. Although the optical absorption at the front and rear facets are near identical, the reflectivity of each facet is intentionally modified by the deposition of optical coatings to have most of the light emitted from the front facet. COD is generally observed at the front facet due to the combination of the high optical intensity and the optical absorption at the surface. COD that occasionally occurs at the rear facet or within the optical cavity is likely a result of a fabrication error. The only report of a cause of a COD initiating within the optical cavity indicates that the COD is initiated by thermal stress from a large void in the solder.

The cause of COD failure in LDs is optical absorption at the front facet leading to local melting of the semiconductor. As long as there is gain within the optical cavity, a liquid region will be maintained on the surface of the COD facing the laser light, whereas the remaining portions of the COD, no longer heated by the laser light, will resolidify. The COD track is thus a permanent image of a portion of the laser mode that caused the COD.

Since the width of the optical cavity of early LDs was comparable to the length, both transverse modes and internally circulating modes were supported. Transverse optical modes leave linear COD tracks along the optical cavity which is generally along the <110> direction and can be easily mistaken for <110> DLDs. However, the defect structure of COD and <110> DLDs is completely different. The former consists of a melted and resolidified region containing defects, whereas the latter consists of a dislocation network. For COD due to internally circulating modes, no postmortem images of early LDs were reported. Since the critical angle for GaAs is ~16°, the COD track should be at an incident angle of 16° to the surface where the COD initiated. Such COD tracks due to both transverse modes and internally circulating modes (ring-cavity modes) were noticed, by the authors in 2009, in numerous published COD images as well as their own images.

The proposed cause of COD suggested several approaches to researchers to improve laser diode performance and reliability that have been successfully verified by novel devices that actually have significantly higher brightness and reliability.

The power density at which LDs fail by COD ($P_{COD}$) and the speed at which a COD ($v_{COD}$) propagates should depend on the material system. Different materials have different material properties, e.g., thermal conductivity, melting temperature, and heat of fusion, that affect the values of $P_{COD}$ and $v_{COD}$. While the COD phenomena in GaAlAs/GaAs LDs and InGaAsP/InP LDs are similar, $P_{COD}$ in the quaternary LDs (50–100 MW/cm$^2$) is an order of magnitude higher than in the ternary LDs (6–12 MW/cm$^2$). The value of $v_{COD}$ for GaAlAs/GaAs LDs (2–4 m/s) is four orders of magnitude higher than in InGaAsP/InP LDs (2–8 $\times$ $10^{-4}$ m/s). As a result, some LD manufacturers have successfully replaced GaAlAs/GaAs with aluminum-free and aluminum-free active layer material systems to improve LD brightness and reliability.

Since the optical absorption at the surface is affected by chemical reaction of the semiconductor with the ambient, it is not surprising that $P_{COD}$ decreases with increasing time and temperature. While the COD failure mode has not been eliminated, two elegant methods have been invented to stabilize the laser diode

facet against chemical reactions, i.e., impurity-induced disordering (IID) of the quantum well to form a window laser and ultrahigh vacuum cleave and coat, also known as the E2 process.

In the E2 process, the LD facets are cleaved in ultrahigh vacuum ($<1 \times 10^{-9}$ Torr) to prevent contamination of the newly formed surface by the ambient. Without breaking vacuum, the cleaved surfaces are coated by a thin film of amorphous silicon that does not interact with the LD semiconductor and is impervious to the ambient. Laser diodes passivated with the E2 process have significantly larger values of $P_{COD}$, due to the lack of initial interaction with the ambient, and the silicon film prevents $P_{COD}$ from decreasing by ambient interaction over time.

An equally effective method to increase and maintain $P_{COD}$ is to form a layer that is transparent to the lasing radiation at the facet (window structure) by IID of the quantum well. The quantum-well layer is adjacent to waveguide layers that have higher energy bandgaps. The windows at each facet are formed by intermixing the atoms of the quantum well, in a region ~25 μm adjacent to each facet, with the material of the waveguide layers. The quantum well ceases to exist in the window layer, replaced with defect-free material that has a higher bandgap than the quantum well, and thus increases $P_{COD}$. The window layer also acts as a physical barrier between the quantum well and the ambient, thus passivating the facet.

BA-LDs fabricated with the E2 process or IID window structures are highly reliable, but some units still fail by COD during long-term operation when operated at a power level significantly below $P_{COD}$. A likely explanation for some of the failures is small fabrication errors, e.g., pinholes in the silicon film for E2 passivated LDs or material defects unintentionally generated by the IID process. The fabrication errors eventually create sufficient optical absorption during device operation to initiate the COD.

Another explanation for failure by COD of broad-area LDs fabricated with the E2 process or IID window structures is suggested by the following. BA-LDs exhibit mainly transverse modes but can support internal circulating (IC) modes. Transverse and IC filaments in multimode LDs have spatial-temporal fluctuations. Both modes tend to reflect at specific locations on the front facet, but most combinations of transverse mode and IC mode will lack either the intensity or duration to cause COD since the LD is operated significantly below $P_{COD}$. However, over sufficiently long times, there will be a time when there is sufficient intensity and duration where interference of the two optical fields can initiate a COD. This is clearly a random process and likely accounts for the observed random failures in life test of BA-LDs fabricated with either the E2 or IID process.

## References

1. R.N. Hall, G.E. Fenner, J.D. Kingsley, T.J. Soltys, R.O. Carlson, Coherent light emission from GaAs p-n junctions. Phys. Rev. Lett. **9**, 366–368 (1962)
2. M.I. Nathan, W.P. Dumke, G. Burns, F.H. Hill, G.J. Lasher, Stimulated emission of radiation from GaAs p-n junctions. Appl. Phys. Lett. **1**, 62–64 (1962)

3. T.M. Quist, R.H. Rediker, R.J. Keyes, W.E. Krag, B. Lax, A.L. McWhorter, H.J. Zeiger, Semiconductor maser of GaAs. Appl. Phys. Lett. **1**, 91–92 (1962)
4. N. Holonyak Jr., S.F. Bevacqua, Coherent (visible) light emission from $Ga(As_{1-x}P_x)$ junctions. Appl. Phys. Lett. **1**, 82–83 (1962)
5. D.P. Cooper, C.H. Gooch, R.J. Sherwell, Internal self-damage of gallium arsenide lasers. IEEE J. Quantum Electron. **QE-2**, 329–330 (1966)
6. C.D. Dobson, F.S. Keeble, The surface damage of high output gallium arsenide lasers, in *Gallium Arsenide: 1966 Symposium Proceedings* (Institute of Physics and Physical Society, London, 1967), pp. 68–71
7. R.J. Mears, L. Reekie, I.M. Jauncie, D.N. Payne, Low-noise erbium-doped fibre amplifier operating at 1.54 μm. Electron. Lett. **23**, 1026–1027 (1987)
8. E. Desurvire, J.R. Simpson, P.C. Becker, High-gain erbium-doped traveling-wave fiber amplifier. Opt. Lett. **12**, 888–890 (1987)
9. P.C. Becker, N.A. Olsson, J.R. Simpson, *Erbium-Doped Fiber Amplifiers: Fundamentals and Technology* (Academic, San Diego, 1997), p. 70. ISBN 0-12-084590-3
10. F. Hakimi, H. Po, E. Snitzer, Optical fiber laser, U.S. Patent 4,680,767 (1987)
11. N. Holonyak Jr., Semiconductor device fabrication with disordering elements introduced into active region, U.S. Patent 4,511,408 (1985)
12. M. Glasser, E.E. Latta, Method for mirror passivation of semiconductor laser diodes, U.S. Patent 5,063,173 (1991)
13. M. Ettenberg, H. Kressel, Dependence of current density and efficiency on Fabry-Perot cavity parameters: single heterojunction (AlGa)As-GaAs laser diodes. J. Appl. Phys. **43**, 1204–1210 (1972)
14. M. Ettenberg, H.F. Lockwood, H.S. Sommers, Radiation trapping in laser diodes. J. Appl. Phys. **43**, 5047–5051 (1972)
15. B.W. Hakki, F.R. Nash, Catastrophic failure in GaAs double-heterostructure injection lasers. J. Appl. Phys. **45**, 3907–3912 (1974)
16. G.D. Henshall, Gallium aluminum arsenide heterostructure lasers: factors affecting catastrophic degradation. Solid State Electron. **20**, 595–602 (1977)
17. G.D. Henshall, The suppression of internally circulating modes in (GaAl)As/GaAs heterostructure lasers and their effect on catastrophic degradation and efficiency. Appl. Phys. Lett. **31**, 205–207 (1977)
18. C.H. Henry, P.M. Petroff, R.A. Logan, F.R. Merritt, Catastrophic damage of $Al_xGa_{1-x}As$ double-heterostructure laser material. J. Appl. Phys. **50**, 3721–3732 (1979)
19. H. Temkin, S. Mahajan, M.A. DiGiuseppe, A.G. Dentai, Optically induced catastrophic degradation in InGaAsP/InP layers. Appl. Phys. Lett. **40**, 562–566 (1982)
20. A. Moser, E.E. Latta, D.J. Webb, Thermodynamics approach to catastrophic optical mirror damage of AlGaAs single quantum well lasers. Appl. Phys. Lett. **55**, 1152–1154 (1989)
21. C.W. Snyder, J.W. Lee, R. Hull, R.A. Logan, Catastrophic degradation lines at the facet of InGaAsP/InP lasers investigated by transmission electron microscopy. Appl. Phys. Lett. **67**, 488–490 (1995)
22. K.H. Park, J.K. Lee, D.H. Jang, H.S. Cho, C.S. Park, K.E. Pyun, J.Y. Jeong, S. Nahm, J. Jeong, Characterization of catastrophic optical damage in Al-free InGaAs/InGaP 0.98 μm high-power lasers. Appl. Phys. Lett. **73**, 2567–2569 (1998)
23. A.K. Chin, R.K. Bertaska, M.A. Jaspan, A.M. Flusberg, S.D. Swartz, M.T. Knapczyk, R. Petr, I. Smilanski, J.H. Jacob, A comprehensive model of catastrophic optical-damage in broad-area laser-diodes. Proc. SPIE **7198**, 719846-1 (2009)
24. A.K. Chin, R.K. Bertaska, H. Eppich, M.A. Jaspan, J.H. Jacob, Bulk catastrophic optical-damage of 980 nm, $In_xGa_{1-x}As/GaAs$, high-power, broad-area laser-diodes due to a void in $Au_{80}Sn_{20}$ solder, in *Fall Meeting of the MRS, Symposium B*, Boston, MA, 2009
25. J.H. Jacob, R. Petr, M.A. Jaspan, S.D. Swartz, M.T. Knapczyk, A.M. Flusberg, A.K. Chin, I. Smilanski. Fault protection of broad-area laser diodes, Proc. SPIE **7198**, 719841-1 (2009)

26. F.R. Nash, Mode guidance parallel to the junction plane of double-heterostructure GaAs lasers. J. Appl. Phys. **44**, 4696–4707 (1973)
27. H.C. Casey Jr., M.B. Panish, *Heterostructure Lasers, Part B: Materials and Operating Characteristics* (Academic, New York, 1978), p. 176 (pt. B). ISBN 0-12-163102-8
28. J.C. Dyment, J.C. North, L.A. D'Asaro, Optical and electrical properties of proton-bombarded p-type GaAs. J. Appl. Phys. **44**, 207–214 (1973)
29. A. Moser, E.E. Latta, Arrhenius parameters for the rate process leading to catastrophic damage of AlGaAs-GaAs laser facets. J. Appl. Phys. **71**, 4848–4863 (1992)
30. J. Hashimoto, I. Yoshida, M. Murata, T. Katsuyama, Aging time dependence of catastrophic optical damage (COD) failure of a 0.98-μm GaInAs-GaInP strained quantum-well laser. IEEE J. Quantum Electron. **33**, 66–70 (1997)
31. M. Bou Sanayeh, P. Brick, W. Schmid, B. Mayer, M. Möller, M. Reufer, K. Streubel, J.W. Tomm, G. Bacher, Temperature-power dependence of catastrophic optical damage in AlGaInP laser diodes. Appl. Phys. Lett. **91**, 041115-1–041115-3 (2007)
32. N. Lichtenstein, M. Krejci, Y. Manz, J. Boucart, B. Valk, J. Müller, C. Button, S. Weiß, S. Pawlik, B. Sverdlov, Recent developments for BAR and BASE: setting the trends. Proc. SPIE **6876**, 68760C-1-687560C-12 (2006)
33. V. Rossin, M. Peters, E. Zucker, B. Acklin, Highly reliable high-power broad-area laser diodes. Proc. SPIE **6104**, 610407-1-610407-10 (2006)
34. A.K. Chin, unpublished results
35. J.R. Marciante, J.P. Agrawal, Beam filamentation and its control in high-power semiconductor lasers. Proc. SPIE **3283**, p. 302 (1998)
36. B.C. DeLoach, B.W. Hakki, R.L. Hartman, L.A. D'Asaro, Degradation of CW GaAs double-heterojunction lasers at 300 K. Proc. IEEE **61**, 1042–1044 (1973)
37. P. Petroff, R.L. Hartman, Defect structure introduced during operation of heterojunction GaAs lasers. Appl. Phys. Lett. **23**, 469–471 (1973)
38. H.C. Casey Jr., M.B. Panish, *Heterostructure Lasers, Part B: Materials and Operating Characteristics* (Academic, New York, 1978), p. 8 (pt. B). ISBN 0-12-163102-8
39. V.T. Bublik, V.N. Leikin, Calculation of the pseudobinary alloy semiconductor phase diagrams. Phys. Stat. Sol. (a) **46**, 365–372 (1978)
40. Y. Sin, N. Presser, B. Foran, S.C. Moss, Investigation of catastrophic degradation in high power multi-mode InGaAs strained quantum well single emitters. Proc. SPIE **6876**, 68760R-1-68760R-12 (2008)
41. Y. Sin, N. Ives, N. Presser, S.C. Moss, Root cause investigation of catastrophic degradation in high power multi-mode InGaAs-AlGaAs strained quantum well lasers. Proc. SPIE **7583**, 758307-1-758307-12 (2010)
42. Y. Sin, N. Presser, B. Foran, N. Ives, S.C. Moss, Catastrophic facet and bulk degradation in high power multi-mode InGaAs strained quantum well single emitters. Proc. SPIE **7198**, 719818-1-719818-12 (2009)
43. A. Hodges, J. Wang, M. DeFranza, X. Liu, B. Vivian, C. Johnson, P. Crump, P. Leisher, M. DeVito, R. Martinsen, J. Bell, A CTE matched, hard solder, passively cooled laser diode package combined with nXLT facet passivation enables high power high reliability operation. Proc. SPIE **6552**, 65521E-1-65521E-9 (2007)
44. M. Bou Sanayeh, A. Jaeger, W. Schmid, S. Tautz, P. Brick, K. Steubel, G. Bacher, Investigation of dark line defects induced by catastrophic optical damage in broad-area AlGaInP laser diodes. Appl. Phys. Lett. **89**, 101111–101114 (2006)
45. O. Ueda, *Reliability and Degradation of III-V Optical Devices* (Artech House, Norwood, 1996), p. 261. ISBN 0-89006-652-3
46. B.D. Swartz, Dark line defect growth in optically pumped $Al_xGa_{1-x}As$ laser material. J. Appl. Phys. **58**, 677–682 (1985)
47. P. Crump, G. Blume, K. Paschke, R. Staske, A. Pietrzak, U. Zeimer, S. Einfeldt, A. Ginolas, F. Bugge, K. Häusler, P. Ressel, H. Wenzel, G. Erbert, 20 W continuous wave reliable

operation of 980 nm broad-area single emitter diode lasers with an aperture of 96 μm. Proc. SPIE **7198**, 719814-1-719814-9 (2009)
48. S.N.G. Chu, N. Chand, W.B. Joyce, P. Parayanthal, D.P. Wilt, Generic degradation mechanism for 980 nm $In_xGa_{1-x}As/GaAs$ strained quantum-well lasers. Appl. Phys. Lett. **78**, 3166–3168 (2001)
49. A.K. Chin, I. Camlibel, L. Marchut, S. Singh, L.G. Van Uitert, G.J. Zydzik, Evaluation of silicon films as a diffusion mask and encapsulant for InP and GaAs. J. Appl. Phys. **58**, 3630–3634 (1985)

# Chapter 6
# Reliability and Degradation of Vertical-Cavity Surface-Emitting Lasers

**Robert W. Herrick**

**Abstract** Vertical-cavity surface-emitting lasers (or VCSELs) are one of the largest-selling types of semiconductor lasers made today and are widely used in fiber-optic data communications equipment and a large fraction of optical mice made today. The lasers have been shipped with a field failure rate of $<10$ ppm/year over the past decade, thanks to extensive precautions that have been taken to qualify the designs and to screen out weak devices and weak batches. In this chapter, we will go over in detail what precautions are taken to assure high reliability for the most demanding applications, all the way from the design process to high-volume shipment. Then we discuss common failure mechanisms seen in VCSELs and how failure analysis is performed to determine the cause of failure and prevent future recurrences.

## 6.1 Introduction

The vertical-cavity surface-emitting laser (VCSEL) is a type of semiconductor laser that emits light out the top or bottom of the device, rather than in the plane of the device as previous semiconductor lasers had and most continue to. The VCSEL is credited to Kenichi Iga [1], with the modern implementation being demonstrated CW by Jewell [2] and shortly thereafter Coldren's group at UCSB [3]. By the mid-1990s, it was starting to be commercialized, and the first VCSEL-based fiber-optic transceivers became commercially available [4].

The primary advantages of VCSELs include easy wafer-level testing, no need for facet cleaving or coating, a symmetric output beam without the astigmatism that plagues competing in-plane lasers, and better reliability due to the fact that there are no exposed, electrically pumped facets. The costs have thus been lower than were obtained for any competing lasers.

R.W. Herrick (✉)
C8 Medisensors, 6375 San Ignacio Ave., San Jose, CA 95119, USA
e-mail: rherrick@c8medisensors.com

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_6,
© Springer Science+Business Media New York 2013

It is estimated that over 700 million VCSELs have been sold. Data communication applications are said to account for a majority, with VCSELs for optical mice and cell phone thumb trackpads accounting for a growing portion of sales [5]. It is anticipated the volumes of VCSELs used will continue to increase. With their low capacitance, low drive currents, and efficient fiber coupling, VCSELs are an excellent short-range data communication source. Millions of 10-Gbps transceivers are sold every year, with products of $4\times25$ Gbps under development for 100-Gbps links and 40-Gbps single channel links demonstrated in research laboratories [6].

While VCSELs had much better reliability than the problematic CD lasers used in GLM transceivers that went before them, they have not been without their problems at times. The primary reliability difficulties in VCSELs have come from the climb dislocations that have occurred on very small subpopulations of parts. (These are also known as "dark-line defects," or $\langle 100 \rangle$ DLDs, where the number refers to the Miller index of the direction they prefer to travel along, diagonal to the natural cleave planes of the device.) While these might only be present on a fraction of a percentage of parts, when these are agglomerated in a large data communication center with thousands of links, the failure rates for the system can still involve multiple unscheduled failures per year.

VCSELs present an excellent opportunity to study many of the pitfalls possible in laser reliability. Some of the broadest data sets in semiconductor laser reliability have been obtained in the qualification and ongoing reliability testing of VCSELs. We will start by introducing VCSEL operation, then discuss how VCSELs are developed, qualified, and fielded into production. We will also discuss failure analysis for returned VCSELs. Finally, we will briefly discuss the future of VCSEL reliability and the demanding new applications under development.

Our approach is based on the availability of extensive failure analysis characterization techniques. While it may at times be possible to "guess" what is wrong with the laser without access to failure analysis, only the luckiest of engineers will be able to find and eliminate all important failure mechanisms through a "black box" approach, without the ability to "look inside the box" using analytical methods.

Our recommendations are geared toward data communication VCSELs used in high-speed links, where failure rates of a few FITs (i.e., $<10$ ppm/year of operation) are needed by the systems designers. For other less demanding applications, such as computer mice, many of the recommendations are optional and may not be required.

### *6.1.1 Advantages of VCSELs and VCSEL Applications*

In-plane lasers (including most other semiconductor lasers in this book) are still more common than VCSELs. As shown below in Fig. 6.1, they lase along the length of the chip, from one facet to the other, before light comes out one end. In-plane lasers (also known as "edge emitters") tend to have a larger electrically pumped area, which translates into higher threshold and operating currents than VCSELs

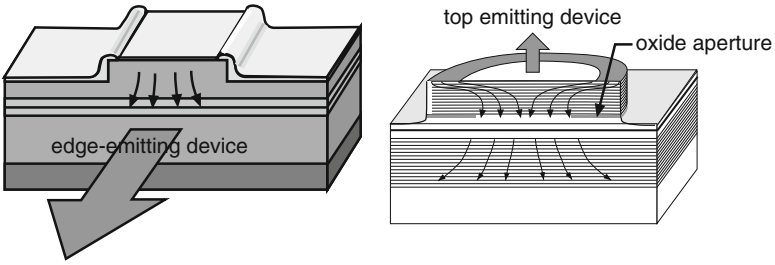



**Fig. 6.1** In-plane lasers (*left*) have emission through the cleaved sides of the die; VCSELs (*right*) have mirrors above and below the active region and emit through the top

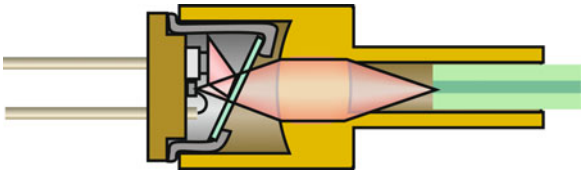


**Fig. 6.2** A cross section of a VCSEL header can, as might be used for high-speed data communications. The part is controlled through the leads at the left. The light is emitted from the VCSEL in the center of the can. Part of the light is reflected off the angled coated glass and onto a monitor photodiode. The remainder of the light is focused with a lens barrel onto a fiber-optic patch cable

(which, of course, is a disadvantage). But on the plus side, in-plane lasers are often capable of producing more single-mode power, have lower thermal resistance, and are available in a much wider range of wavelengths.

VCSELs are popular for a few different reasons. The vertical emission allows easy wafer-level probing, whereas in-plane lasers require more costly cleaving, coating, and bar-level testing before singulation into individual die can take place. The simpler processing and optics of VCSELs can drive lower costs in many applications, such as the VCSELs used in optical mice, which are sold for under US$0.25 each. VCSELs have a roughly circular beam, without the astigmatism that is present in in-plane lasers. This further simplifies lens design and cuts optics costs. A very simple "transmitter optical subassembly" (or "TOSA") design can be used, as shown in Fig. 6.2. VCSELs have lower drive current than in-plane lasers, which reduces power consumption; they also are usually designed to change drive currents much less over a wide range of temperatures [7].

A single-mode VCSEL – as might be used in an optical mouse – can be fabricated far more cheaply than would be the case for an in-plane laser. The vertical emission eliminates the needs for rotated submounts in most applications, which also eliminates piece parts and equipment complexity. While many fiber-optic lasers require expensive optical isolators to prevent longitudinal mode hops or instability, the VCSEL is fixed to one longitudinal mode and doesn't require optical isolators for any popular applications it is used in. Finally, from a reliability

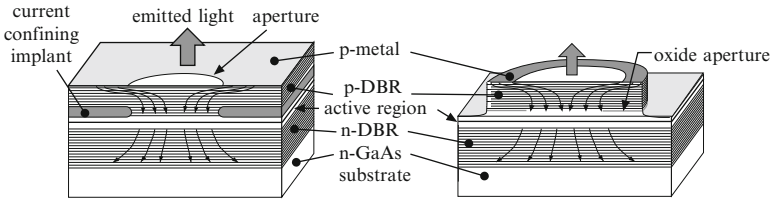



**Fig. 6.3** Proton-implanted VCSEL has gain only in the nonimplanted regions (*left*); the oxide VCSEL also has optical waveguiding in the pillar (*right*)

standpoint, VCSELs are able to avoid facet degradation, since the active area is covered with ~3–4 μm of material and is laterally isolated from exposed edges by at least 10 μm in most designs. This is in contrast with edge emitters, where the active region goes all the way to the facet – except in expensive buried facet lasers – thereby creating gradual degradation at the facets, which is usually the main gradual failure mechanism for edge emitters [8]. For more information on VCSELs and their advantages, refer to the following book [9] or review papers [6, 10, 11].

### 6.1.2 How Vertical-Cavity Surface-Emitting Lasers Work

Most VCSELs produced today use a few GaAs quantum wells (usually 3 wells that are 7–8 nm thick each). The vast majority emit light at 850 nm. The mirrors are made from "distributed Bragg reflector" (DBR) stacks of $Al_xGa_{1-x}As/Al_yGa_{1-y}As$, where the low-index layer with the high aluminum mole fraction will have a value of "$x$" typically from 0.8 to 1.0. The high index of refraction layer has a low aluminum mole fraction $y$ typically 0.15–0.20 for 850-nm VCSELs. The small (typically 6–8% per interface) index discontinuity per interface is repeated in a stack of a few dozen $\lambda/4$ pairs, which collectively provides very high reflectivity, of 99.0–99.8% in the upper DBR and 99.95% in the lower DBR. It is traditional to use a GaAs substrate with $1–4\times10^{18}$ $cm^{-3}$ silicon ($n$-type) doping and to grow the bottom DBR n-type and the upper DBR p-type. Finally, lateral current control can be provided by either proton implantation or, more often for high-speed lasers, by an oxide aperture. The oxide adds predictability to the modal behavior during high-speed modulation but adds processing complexity and, as we shall see later, can create potential low-level reliability problems that need to be addressed. Diagrams of a proton-implanted VCSEL and an oxide VCSEL are shown in Fig. 6.3.

VCSELs can be designed to be relatively temperature independent, which is another advantage. They also have a characteristic rollover in their L-I curve. It has been our experience that VCSELs will age very quickly if operated past this rollover point. A number of "rules of thumb" have been established, such as "never operate at an average power greater than 1/3rd of the room-temperature rollover power" for 1–4-Gbps VCSELs (or over ½ the peak power for 10–17-Gbps VCSELs).

## 6.2 Steps to Producing a Reliable VCSEL

We will now go through the steps of bringing a VCSEL to market, in rough chronological order. Of course the very first step is staffing the team. At least a few members of the team require past VCSEL experience, either at a university or a past company. We have found through experience that the reliability engineer needs to have a strong laser background. It is much easier to teach a laser expert to be a good reliability engineer than it is to teach a statistician or reliability expert about lasers. In fact, attempts to do the latter have almost always been unsuccessful in our experience; treating the laser as a "black box" almost never works, as there simply aren't large enough data sets to use such techniques successfully.

### *6.2.1 Design: Understand Use Environment and Needs*

In the design phase, the wafer epitaxial structure is designed. Key variables include grading doping to minimize mirror resistance, minimizing mirror and cavity optical losses, and controlling lateral mode structure. Problems often result from chaotic "mode hopping" from one lateral mode to another, which results in unstable output power. As mentioned earlier, oxide VCSELs control this to some extent; additional modal control structures have also been employed to address this [12].

One design choice that needs to be made is the offset in wavelength between the peak of the mirror resonance and the gain peak. A large offset with the gain 10–15 nm longer than the resonator peak is optimal for low drive current at high-temperature operation [7, 13]. However, an unfortunate trade-off is often made to improve −40°C operation, which requires a reduced offset. This operation is something that is only performed in laboratory qualification testing and never occurs in real-world operation, yet the design changes to accommodate this often add 15% to the required bias current over what would be needed for an optimized design, thereby reducing the lifetime of all parts significantly.

An important trade-off exists between drive current and reliability. Almost all performance parameters improve with increased current: power, speed, jitter, mask margin, rise and fall times, etc. However, reliability degrades significantly with increased drive current. Most popular reliability models call for it to degrade as the square of the drive current, with additional degradation due to the higher thermal impedance of the VCSEL. The VCSEL designer should also talk to the module or product designers about the maximum drive current. Many early 1-Gbps IC laser driver designs had very high peak bias levels that could be reached, where the device would degrade very rapidly. This caused field failures to be generated when the customers had out-of-control HAST ("shake and bake") systems to screen their boxes that took transceivers well above their maximum allowable datasheet temperatures. Allowable bias currents need to be clamped to prevent rapid wearout degradation.

In a similar manner, many performance characteristics improve as device diameter is reduced, but again, reliability gets worse. So much of the challenge in designing reliable high-speed lasers is trying to make the biggest laser that can still give adequate high-speed performance.

The lateral design – that design reflected in the photomasks used in wafer fabrication – is also important. In general, confining the current to a smaller area lowers drive current and capacitance and inductance, thereby improving high-speed performance. However, it also increases thermal resistance (and thereby junction temperature), which is already high in VCSELs (typically 1.5–3°C/mW). Small devices also have high current density, which also drives rapid degradation. So designers need to balance the advantages of small VCSELs for high speed with the advantages of the large VCSELs for reliability.

One important design decision is the materials system to be used. As mentioned earlier, GaAs active regions, with AlGaAs DBRs, have proven popular where wavelength is not critical, because the 850-nm wavelength works well with silicon detectors and all compositions of AlGaAs are reasonably well strain matched to the GaAs substrate – therefore, complicated strain-balanced quaternary growths requiring careful composition control can be avoided.

There are three potential types of material systems. The critical feature is the electrical properties of dangling atomic bonds, as might happen either at point defects or mechanical defects such as dislocations. The three types, in order from most to least desirable, are:

1. The most desirable material systems have active regions without midgap recombination centers from dangling bonds. These do *not* show self-propagating climb dislocations. These include InGaAsP on InP substrates, for example, most 1.3- or 1.55-μm lasers [14, 15]. If physical damage occurs to these die, performance may be degraded if the lasing area is damaged, but it won't get any worse over time. Unfortunately, this material system is very hard to work with for VCSELs and is therefore not popular as of this writing.
2. The next most desirable materials system has active regions with midgap recombination centers from dangling bonds but sufficient compressive strain to pin dislocation growth. This includes $In_xGa_{1-x}As$ lasers on GaAs substrates emitting from 920 to 980 nm ($x = 0.05-0.20$) [16]. If these are damaged physically or by ESD, damage will not continue to get worse over time. Unfortunately, this wavelength is not popular because it does not work well with silicon photodetectors. It is possible to add both indium for compressive strain and aluminum for increased bandgap and make an 850-nm compressively strained quantum well or even one operating at 808 nm [17, 18]. However, the window of acceptable growth conditions is far narrower [19], and thus it has not yet been adopted for 850-nm VCSELs at this time. 850-nm VCSELs with strained quantum wells were first demonstrated by Ko et al. [20], but no attempt was made to quantify the reliability of these first devices, as the primary purpose was enhanced performance, rather than resistance to dislocations and better reliability. Note that these devices can still show

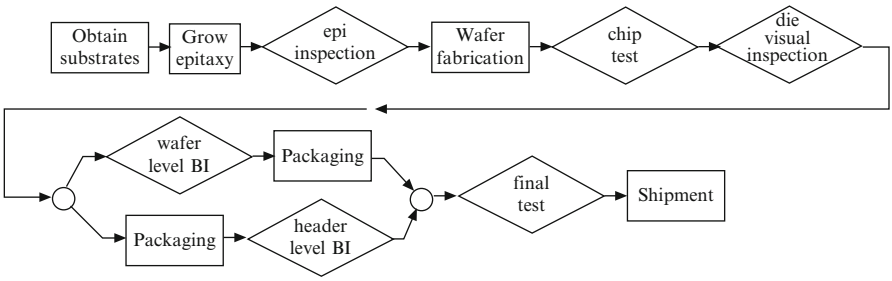



**Fig. 6.4** Simplified process flow for VCSELs. *Diamonds* show steps that involve measurement and can reject failed wafers or die. Flows for both wafer-level and die-level burn-in are shown

dislocation glide (REDG) if subjected to mechanical stress, even though they are not susceptible to the much more severe dislocation climb (REDC) failure mechanism [21].

3. The least desirable material system for reliability has active regions with midgap recombination centers from dangling bonds that have self-propagating climb dislocations. The dominant 850-nm VCSELs with GaAs quantum wells fit in this category. While in practice these have been made by many suppliers with failure rates as low as tens of parts per million per year, doing so requires a high degree of vigilance and wafer screening. Unfortunately, virtually all VCSELs fit into this category and thus require the extensive safeguards listed in this chapter. If materials system #1 or #2 with nonpropagating dislocations were used, many of the precautions discussed here could be relaxed while still obtaining adequate reliability. As mentioned earlier, this materials system has maintained dominance because of its ease of fabrication and favorable performance, even though reliability is a greater challenge.

## *6.2.2 Manufacturing*

A schematic view of the wafer fabrication process is shown in Fig. 6.4.

### 6.2.2.1 Epitaxy

To produce VCSELs, manufacturers start with pure "substrates," usually of GaAs a few degrees off from the 001 plane. They then place them in an "epitaxial reactor" and grow ~8–10 μm of material in the manner specified by the designer and the epi recipe. In other types of devices described in this book, the quality of epitaxial growth is a critical factor in the device reliability. But with VCSELs, it is far more difficult and an even bigger fraction of the value added. If the epitaxy has a good design, and it is properly grown, the rest of the process is likely to be relatively

problem-free since VCSELs lack many of the other low-yield wafer processing steps present in in-plane lasers. The epitaxial reactors that are used to grow the wafers typically will grow several wafers at a time. It is usually a safe assumption that the other wafers from this lot will be nearly identical, at least so long as they travel along an identical rotational track (e.g., if 1 substrate is at the center and 6 are along the outside track, the center part might be different than the other 6). Often, one part is "sacrificed" for destructive evaluation and cleaved into evaluation pieces to collect thickness and doping data on the rest of the epitaxial run.

It is beyond the scope of this chapter to address epitaxial growth, except to address practical considerations that affect reliability. Semiconductor laser reliability is typically more sensitive to epitaxial quality than microwave transistor (e.g., HEMT) reliability is. This is because of the very high minority carrier recombination rates in the active region. Even relatively low concentrations of traps will cause rapid degradation of the lasers during normal use. Fortunately, this is easy to detect with a good wafer qualification program. Unfortunately, such wafer qualification programs take time to give feedback (usually 1–2 months when wafer processing time and aging time are included), and problems that are detected often take another 1–2 months to isolate to root cause and fix. For this reason, it is foolhardy to try to reduce inventory of any present type of semiconductor lasers to less than a 3-month buffer stock. Every company will receive pressure from internal supply chain people to reduce inventories to 1 month or less. The laser production division must resist these pressures and educate the supply chain planners on the ways lasers differ from more mature components like metal stampings. While reactor uptime has improved substantially in the last 20 years, planners should count on any individual reactor being down for at least 1 month per year and multiple reactors going down if any shared gasses are contaminated.

Once the lasers have been grown, epitaxial evaluation is performed. Very tight control of layer thickness is needed with VCSELs ($<1\%$ deviation from target vs. $+/-10\%$ being acceptable for many in-plane lasers). Most firms will reject on-tenth to one-third of the wafers after this evaluation. Wafers are most commonly rejected for too high or too low of a growth rate, which then gives the improper resonator peak and offset from the gain peak. The wafers that pass then move on to wafer fabrication, where the goal is to control current and limit current to just a small portion of the die, to (in the case of oxide VCSELs) put a lateral optical waveguide into the device, and finally to passivate and metalize the device.

During the development process, growth temperatures, doping level profiles, and III-V gas flow ratios must be optimized. Furthermore, the device designer's epitaxial structure must be optimized. For example, two well-known design flaws are oxide confinement layers with abrupt grading (which result in failure in a matter of hours) or excessive n-doping in the cladding layer between the active region and the n-DBR, something we will discuss at length in a later section.

Wearout reliability is the length of time that average lasers should last during operation. It will vary during the course of a typical year, since it can be quite sensitive to minor perturbations. Leaks will develop in the reactor, thereby creating traps from unwanted oxygen incorporation. Various unwanted contaminant species

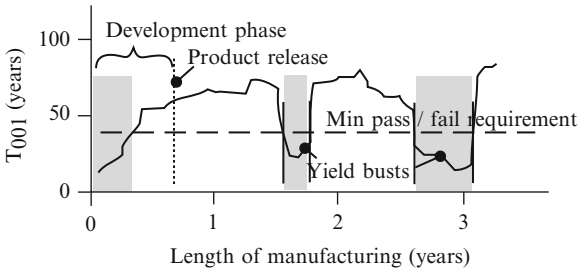



**Fig. 6.5** Theoretical graph of wearout lifetime (time to 0.1 % failure) versus production time shows how epitaxial growth quality can vary substantially over the course of a year but does not generally vary widely over short periods of time

will show up as various source gasses or organometallic liquids are swapped out. As the bubblers are used up, contaminants will become more concentrated toward the end of bubbler life. Carbon background will rise and fall, creating undesired changes in background level that modify the design from that desired by the epi designer and can even create out-of-spec resistance. This creates variability in the reliability of the parts that changes on a month-to-month basis. We have shown this schematically in Fig. 6.5. One of the biggest consequences of this variability is that it makes it difficult to know with certainty what the wearout reliability of the lasers will be from 1 month to the next. Yield cannot be restored by just increasing wafer starts. The important thing is to design the lasers with some margin, to have a strict wafer qualification, and to sample frequently in ongoing reliability testing. Having adequate inventory is also important – at least 3 months is recommended.

#### 6.2.2.2 Wafer Fabrication

After an epitaxial wafer has passed its tests, it moves into wafer fabrication, where most of the steps have relatively high yield and are not usually particularly difficult to manage from a reliability standpoint. One exception is the oxidation process on oxide VCSELs, which can be difficult both to develop and to maintain with high yield. Another problem that has been observed many times in VCSELs as well as other laser devices has been copper contamination, which we will be talking more about later. A third potential problem area is dielectric deposition on top of the device, for both stress and composition. Oxide VCSELs are known to be vulnerable to thermal shock, and special procedures need to be implemented to slow down thermal ramp rates. During wafer fabrication, the die size is defined and typically will be 200–350 μm on a side. With 3″ wafers, usually about 40–50,000 die will be available per wafer, depending on the exact die size and the size of the exclusion zone around the edge.

In wafer fabrication, “wafer lots” travel together through each of the fabrication steps. Typically, such lots will have 4–8 wafers in them.

#### 6.2.2.3 Testing

Once wafer fabrication is complete, the device heads on to wafer probe in most flows. Bad devices are covered with ink or designated as "bad" with their coordinates in a database. During probe testing, tip shape and pressure are important, since scratching can initiate cracks that will lead to field failures. A nice (but not essential) way to protect against this risk is to have thick gold plated onto the wafer. The thick gold can allow substantial probe tip indentation without risk of wafer damage, particularly if the plating process produces "spongy" gold with voids in it. If a simpler process without thick gold is used, the probe tip radius, material, spring force, tip angle, and scrub length must all be carefully optimized to minimize the material damage while still getting repeatable contacts.

#### 6.2.2.4 Wafer Qualification

Before device packaging of the tens of thousands of VCSELs per wafer takes place, the wafer is put on hold until it passes its wafer qualification. This is true except for low-cost VCSELs such as those used in computer mice, where a sampling approach can be used, rather than qualifying every wafer individually. A sample of usually 15–75 parts is then built up into headers that fit onto a test board, and the parts are then aged at very high current and high temperature. If the acceleration factor is sufficient, the test only needs to run for a week or two, since it can accumulate approximately 40 years of normal aging in that period. The parts are examined, and if none exceed the failure threshold, the wafer "passes" the qualification and is removed from hold and moved into manufacturing. If any part fails the test, the wafer fails the qualification. If multiple parts fail, special failure analysis will be performed, and engineering meetings held to understand whether corrective action is required or containment of parts made shortly before or after this wafer is necessary.

#### 6.2.2.5 Packaging and Burn-In

Once a wafer has been qualified, all die will be mounted in packages using automated "pick-and-place" equipment that uses a vacuum pickup nozzle. Potential issues here include the presence of contaminants on the pickup nozzle that could focus breaking forces on a die. This can result in cracks or stacking faults, which we will discuss in a future section. Then, after the other components have also been mounted, the silver epoxy is cured in an oven, and finally wirebonding takes place. Fortunately, damage during packaging has been relatively rare. The header assembly is often called a "TOSA," which stands for "transmitter optical subassembly."

Burn-in, if it was not completed at the wafer level, is completed by having the headers stuffed into burn-in boards and pretested before then being aged a few days

at high currents and moderately high temperatures. We will discuss burn-in in much greater detail later in this chapter. After burn-in, a repeat test is done, and if the device has changed by more than a 2–10% (depending on allowable process drift and test repeatability), the device is discarded.

The packaged TOSAs are now shipped for inclusion in the larger product. Serialization, to at least the wafer level, is usually maintained in the final product assembly, to allow tracking in case of future field failures.

#### 6.2.2.6 FMEA and PCDS: Tools to Anticipate Possible Future Problems

In a mature manufacturing environment that has scaled up over time, a great deal of experience has been collected with a wide variety of production problems, especially if employee retention is good. However, in a startup environment, or for a totally new product, it is difficult to predict what sort of manufacturing defects might be encountered. In order to anticipate such defects, a process called "failure modes and effects analysis" (or FMEA) has been developed. Typically, FMEA involves bringing together experts from a number of different areas for at least a few days into an off-site location. They start by going over all past failures, both in the field and in reliability testing, and going over the process in detail and then brainstorm what could go wrong with each step in the process. If possible, changes are made to eliminate the possibility of error. If not possible, extra detection can be put into place to check and see if the error has occurred, at least if it is one that could have significant impact to the customer.

A second (less common) process is "product compliance to datasheet" (PCDS). Extensive "corner case" characterization will take place when the product is new. For example, the product may be tested down to temperatures as low as $-40^{\circ}$C or have elaborate power supply noise rejection ratio (PSRR) measurements performed at a variety of temperatures and bias conditions. Many of these parameters will supposedly be "guaranteed by design" and not require periodic verification testing. However, for many of the more critical parameters, periodic (monthly or quarterly) reverification at the system level is recommended to allow detection of problems in corner cases.

## 6.3 VCSEL Burn-In and Screening Development and Optimization

To keep field failure rates low (below 10 ppm/year), it is necessary to burn in every VCSEL. In general, the burn-in would be expected to have some "radius of capture" for defects; defects within 5 or 10 μm of the emitting area would be caught with a burn-in of reasonable length, but those further out would require a very long (and expensive) burn-in to catch. To try to detect damage further out, a visual screen is added; reverse-leakage testing is also an option in looking for damage. These areas

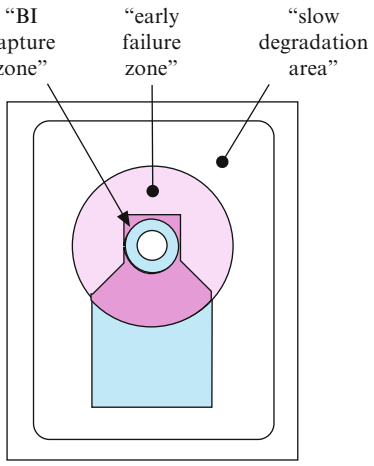



**Fig. 6.6** Drawing schematically showing that only damage near the emitting area can consistently be caught by burn-in, and a larger radial area which is vulnerable (which could be as large as the entire die) will need to be visually screened for damage to prevent field failures

are depicted conceptually in Fig. 6.6, keeping in mind that the actual values depend on chip design. Many manufacturers will reject even small visual flaws that are close to the emitting area and will reject only gross flaws further away from that area.

### *6.3.1 Burn-In Development*

The classical method of doing burn-in development [22] is fairly painful with VCSELs, due to the low maverick failure rate. Typically, you set up a matrix with several hundred parts, age them in your burn-in oven, and check frequently for failures until you have exceeded your likely burn-in duration by several times. You can then create a Weibull plot of the early failures and then calculate escape rates for various proposed burn-in lengths, trading off cost of field failures versus increased burn-in cost. In practice, even an experiment with 1,000 parts will often yield just one or two real failures, if that. Note that very tight pass/fail limits that are near the repeatability of the system may result in many "false rejects," but if those parts are aged further, the overrejected parts usually recover to near the performance of all the other devices in the batch. So burn-in development must be done on tens of thousands of parts if time optimization is to be performed. A more popular approach is to pick something that is thought to be more than long enough and then validate it with wafer qualification. Provided your wafer qualification is stringent and includes thousands of parts, and no premature failures are seen on the parts that pass burn-in and wafer qualification, the burn-in will hopefully be adequate, even if it may be more costly than is necessary. Future cost reductions can involve taking 10,000+ parts and burning them in with at least two steps (reduced burn-in + extra time to come up to the length of the original burn-in). Provided no additional failures are observed in the extended burn-in, you may be able to justify reducing burn-in length, which in turn reduces capital costs and cycle time.

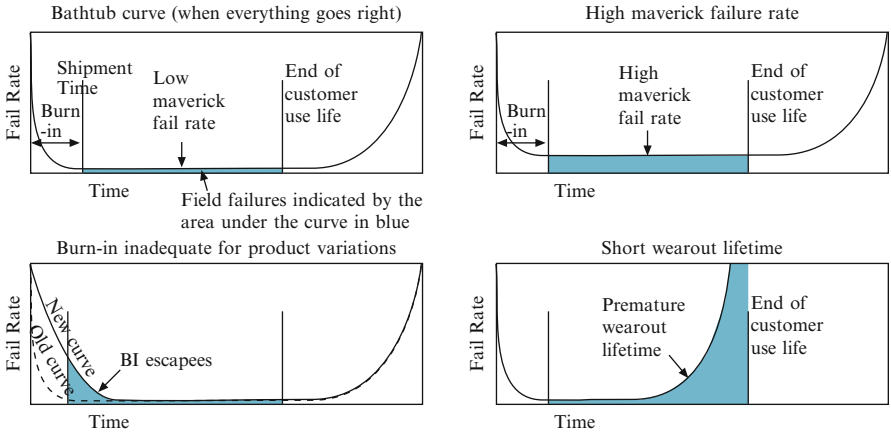



**Fig. 6.7** Four possible outcomes, expressed as "bathtub curves" (See text for discussion)

Optimal burn-in time and temperature are key questions. Since growth of climb dislocations appears to be optically pumped [18, 23], it is important to compromise between low temperatures which maximize spontaneous emission and high temperatures that "soften the lattice" to assist in defect growth. Oven temperatures of 70–100°C (and junction temperatures of 110–150°C) are popular compromises. Drive currents near the peak power at those temperatures are used; these are usually ~3× normal drive current for 1-, 2-, and 4-Gbps lasers and perhaps only 2× normal room-temperature bias for 10-Gbps + lasers. For example, Emcore reported using 12–13-mA burn-in currents on an 8-μm-diameter oxide VCSEL used for 10-Gbps operation and normally operating at 6-mA bias [24].

Parts can be thought of as having two separate populations: a weak subpopulation and a strong population. For effective burn-in to be possible, there needs to be a large separation in the strength between the two populations, where you can kill off all the weak devices, without substantially weakening the main population. There are laser types that do not have adequate separation between the populations, and those cannot go into production until the unscreenable defect mechanisms have been addressed through design, material, or processing changes or extra screening steps to separate the lasers in other ways [25]. Figure 6.7 shows the familiar bathtub curve, and Fig. 6.8 shows the less familiar integral of it. Referring to the upper left of Fig. 6.7, it shows the ideal outcome. The upper right shows the situation with normal burn-in behavior but a high maverick failure rate (e.g., from the "mirror oxide DLD" case described later in Sect. 6.6.1.1). Additional burn-in won't help. At the bottom left of Fig. 6.7, we see a case where burn-in behavior has changed from the original qualification behavior (e.g., from mechanical damage), and now the burn-in is no longer sufficient to catch all of the early failures. When field failures occur, this is the case that suppliers hope is responsible since it means that after the few unscreened infant failures have died off, few additional failures will be seen. At the bottom right, we see the situation where wearout lifetime has been significantly

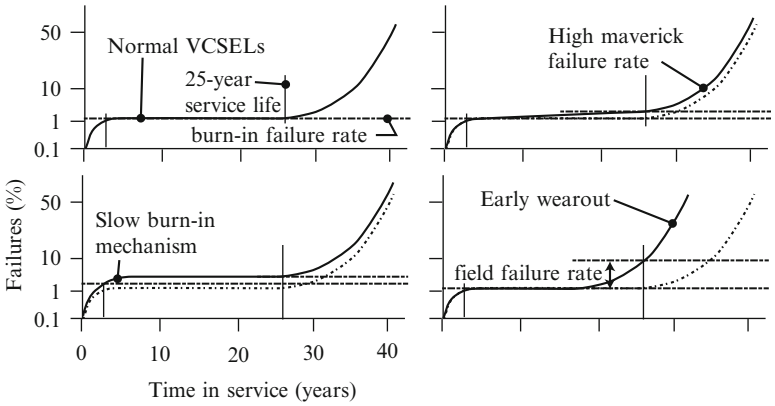



**Fig. 6.8** The same cases as the previous figure, in the form of cumulative failure graphs

reduced from the original qualification case. This is the case that all customers fear: that the failures observed are just the tip of the iceberg and that far more will be seen next month. If this is indeed the case, a recall is the best action to take, sooner rather than later. Fortunately, this has been very rare. Two examples of what might cause this would be epitaxial purity issues, as described in Fig. 6.5, or filamentation issues, as described in Fig. 6.25. The wearout case at the lower right is easily caught if every wafer has a rigorous wafer qualification test that does destructive sample testing before releasing a wafer to production. For an excellent treatment of burn-in, we recommend Finn Jensen's book on the subject [22].

One final note for optimists: no burn-in is possible without current going through the lasers. While the dream of just stuffing wafers into an oven for a couple of days with no drive electronics has long been held, the physics of dark-line defect (DLD) growth would require temperatures so high (>400°C) that they would cause the structure to decompose [26].

While Avalon Photonics (now Oclaro) and Honeywell (now Finisar) were early adopters of wafer-scale burn-in, it is now being more widely adopted [24]. Particularly for array products, or purchasing bare die from outside suppliers, there are substantial advantages to having known-good, burned-in die delivered and not needing to burn-in subassemblies. Avoiding the cost of burn-in racks and manual handling of thousands of subassemblies are also advantageous.

### *6.3.2 Screening Development*

As mentioned earlier, burn-in is necessary, but not sufficient to catch longer-term failures. So visual screening and electrical screening are almost universally used. The visual screening involves either operators or machine vision screening. Nomarski microscopes are far more likely to catch cracks or epi defects than

conventional bright-field microscopes. Machine vision is hard to use without overrejection if any contamination is present on this die. Dust particles are difficult to distinguish from epi defects. Dust is harmless and often very common on dies; epi defects will definitely cause device field failure if they are in an electrically active part of the device within at least 100 μm of the emitting area and have an available path between them and the emitting area. A second screen that is widely used is reverse bias current, at some applied voltage less than breakdown voltage (e.g., 10 V, with a limit of, say, 1 nA). Die with cracks, epi defects, or ESD damage will have substantially higher reverse leakage and be rejected by this screen.

## 6.4 VCSEL Qualification

Once the VCSEL design achieves its desired performance and initial yield targets, it goes through a new product qualification cycle, and special samples are built up for reliability testing. The qualification cycle usually takes 6–12 months to complete after the final design is obtained; it is reasonable for 6–24-month development work to precede the final qualification cycle. This section will discuss the design of the qualification matrix. For parts used in data communication links, Telcordia GR-468 is a guiding document [27]. However, other additional tests are needed to ensure adequate reliability. It is important to distinguish between "qualification tests," which are rather mild and designed so that no failures should be seen, and "wearout tests" which are extremely harsh and expected to result in the majority of the parts failing over a reasonable period (usually a few thousand hours).

### *6.4.1 Realistic Failure Criteria*

One problem with many laser qualifications is the lack of tough testing. R&D managers want to guarantee that they will meet development schedule goals and get the product to market quickly – hopefully avoiding lengthy "reliability improvement programs." Marketing people want to show "clean data" with no failures to customers. In the absence of strong reliability engineering or customer requirements, internal forces drive the company to do qualification testing only and skip "wearout testing" altogether. Fortunately, the VCSEL world had a large number of customers that required harsher test regimes (most of which remain proprietary standards) [28] as well as key suppliers that were willing to publish their wearout data [29]. These have helped the VCSEL world to develop more rigorous test regimes than those for most types of in-plane lasers.

One of the first criteria that need to be addressed is how much current increase can be tolerated before "failure" is obtained. This criterion depends on what sort of system the part is fielded in. For a part with a fixed bias, such as a first-generation parallel optic link, a 20% increase in threshold current is considered a failure.

For a part with a closed-loop system (where a monitor photodiode detects a decline in power and increases the current), a higher amount of degradation can be accepted. Depending on the manufacturer, a 20% increase in threshold (for open-loop applications) is popular; for closed-loop bias, a 2 dB (58.5% increase) in threshold or a 2-dB drop in output power for the nominal bias current is commonplace. From a practical matter, we observed that there was some margin there, as it took a 200–300% increase in threshold current before link failure occurred in the parts we tested, at least back on the early 1-Gbps links. Presumably less margin is available on 10-Gbps VCSELs today, since they are operating closer to their maximum. Performance criteria need to be validated using extensive testing under high-speed modulation as close as possible to the actual package/system configuration as possible.

Once performance criteria have been established, the pass/fail criteria need to be established: how many failures can be accepted over what period of time for each accelerated testing condition? This can be the source of a great deal of controversy internally before any data has been collected and can often safely be put off. In 90% of cases, the parts will either clearly pass or clearly fail their qualification, and in those cases, the need to spend hours debating what would be acceptable to customers can be avoided. However, the reliability engineer responsible should take a careful look at what sort of data might be expected from the proposed tests and where it might be expected to fail in each cell, assuming the part is just barely adequate to pass customer requirements. Check to see that there aren't lots of cells where failure might take 9 months or longer for the first failures to occur. While you certainly want at least a few such long-term cells, most should be focused on obtaining quick results. The core of our work took place at acceleration factors of 300–1,000×; in these tests, you can go through 40 years of aging in about 2–6 weeks and to worst-case customer use in a few months.

### 6.4.2 *Picking System-Oriented Metrics for VCSEL Wearout Lifetime Testing*

Early papers on VCSEL reliability cited mean time to failure (MTTF) at 25°C header temperature. Such data are not useful and indeed can be misleading. Let's discuss what metrics *are* useful and what a recommended "pass/fail" bar might be. MTTF is a commonly used statistic for box level systems and implicitly assumes a constant failure rate or a declining failure rate which is conservatively approximated with a constant failure rate. But where device wearout is an issue – as is the case for most semiconductor lasers – other metrics are more useful.

For specifying wearout lifetime, we prefer the time to 0.1% failure ($T_{001}$) or the time to 1% failure ($T_{01}$). With typical distribution (e.g., a lognormal distribution with sigma = 0.6), $T_{01} \sim$ ¼ of median lifetime $T_{50}$, although $T_{01}$ can be less for broader distributions. In terms of a temperature to cite, a lot depends on the thermal

engineering of the box, how much airflow there is, and whether the header is made of plastic or metal. But header temperatures of 60–70°C are commonplace for 1–10-Gbps transceivers in normal deployments. Therefore, many suppliers cite lifetime statistics at 70°C now, rather than 25°C.

For a great deal of datacom or telecom equipment design, two numbers are often used: a "normal" number where air conditioning is functioning as expected in the data center (25°C ambient) and an "emergency" number which can last up to 2 days at a 45°C ambient, where the equipment must continue to work and no damage is done to the equipment. As long as this excursion does not push the VCSEL too close to its rollover power, no long-term damage should be done. However, there are many larger pieces of switching equipment that are difficult to design without case temperatures reaching 80°C on a continuous basis; these are very challenging applications for reliability with high-speed (10-Gbps+)VCSELs.

### 6.4.3 Other Testing for Mechanical or Corrosion Issues

In addition to studies to examine laser degradation using high powers and currents, mechanical or chemically driven degradation can occur. These are commonly tested using temperature cycling or corrosion testing in "damp heat" chambers, usually at 85°C, with 85% relative humidity [30]. For example, any packaging with solder bumps or adhesives contacting the VCSEL pillar are highly vulnerable to cracking during temperature cycling and normal operation [21]. Some problems have been detected with conventional oxide VCSELs if packaged in nonhermetic packages and exposed to more severe climates [31, 32].

### 6.4.4 Testing to Develop a Wearout Acceleration Model

A first step in the process is to set up a "matrix" like that shown in Table 6.1 below. It should cover a few different oven temperatures and currents. It is important during the planning phase to also determine junction temperatures. Analysis is simpler if cells can be selected with similar junction temperatures but different currents, although this is not strictly necessary. While the testing at the most accelerated cells normally dominates test capacity – since it is used for wafer

**Table 6.1** Sample qualification matrix for 8-μm-diameter, 10- or 14-Gbps VCSEL

| | 85°C | 110°C | 130°C | 140°C | 150°C |
|---|---|---|---|---|---|
| 9 mA | x | | | x | x |
| 11 mA | | x | x | x | x |
| 13 mA | | x | x | | |

Elements with "x" are populated with at least 40 devices each. Note "diagonal" where higher currents are used at lower temperatures and lower currents at higher temperatures

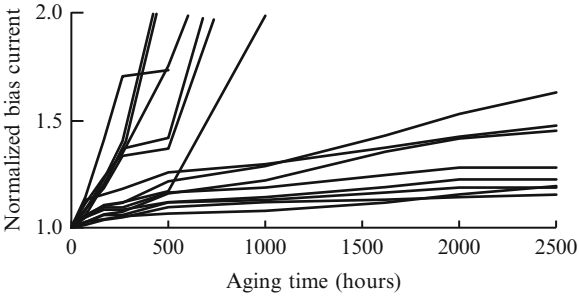



**Fig. 6.9** Aging data for 15 VCSELs aged at extreme stress

qualification – it is important to reserve 15–25% of capacity for long-term reliability testing, tests that go on for 9–24 months as well as a few cells that continue on for 3–5 years. These will go a long way to verifying that you don't have failure mechanisms like those shown later in Fig. 6.22 with lower activation energy than your dominant failure mechanisms.

As recommended before, one should have some idea of what the results might look like before running the experiment, to know how often to check for failures or how long the experimental cells might need to run. It has been traditional to set up the matrix with a 150°C maximum oven temperature and then the lower temperature cells at steps10–15°C below that. With common activation energies, it is desirable for each cell at the next lower temperature step to fail at 1.5 or 2× the time of the one above it. If the cells are spaced too far apart in temperature (as in our early publications, where a 25°C spacing was used [33]), you have to wait years for the results of the lower temperature cells to be available, which is far too slow for most development cycles. The currents used are usually in the range of 15–30 kA/cm$^2$ current density for 10-Gbps VCSELs and at or below their room-temperature rollover current (usually junction temperatures of 50–70°C above ambient for 1- and 2-Gbps lasers and 70–100°C above ambient for 10-Gbps lasers).

Data is then collected, looking like Fig. 6.9. The failure criterion set earlier (e.g., a 2-dB or 58.5% increase) is then used in this figure, and interpolation or extrapolation can be used to find the time to failure, as shown in Fig. 6.10. This "time-to-failure" data is then usually imported into the software package Reliasoft Weibull++ [34] and used to compute either a lognormal or Weibull distribution.

Some caution needs to be used not to try to predict failure time based on small amounts of degradation of much less than the failure criterion. In fact, for our studies, we have only counted parts that have either failed our criterion or come very close to failing, so only minimal extrapolation is needed. It is not uncommon for some analyses of 1.3 or 1.55-μm lasers to take parts that have degraded just a few percent and then use straight line extrapolation to find when the parts would cross the criterion assuming they continued to degrade at the same rate. However, this will give results that are meaningless and dangerous if applied to VCSELs, for reasons that should become clear in our in-depth discussion below in Sect. 6.5.3 on parametric analysis. Fortunately, we have not seen any analyses that made such flawed assumptions published for VCSELs.

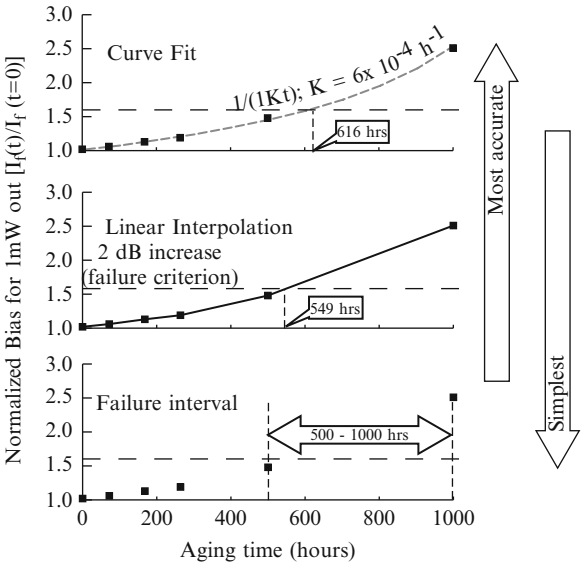



**Fig. 6.10** Three methods of defining failure time or failure interval: curve fit (at *top*), linear interpolation (*middle*), or failure interval (*bottom*)

### 6.4.4.1 Test Hardware

In VCSEL reliability testing, two types of test equipment could be used. The first is expensive "continuous monitoring" that has a dedicated photodetector monitoring each laser at all times. It also allows temperature and current control over a range that is usually specified in a custom built-to-order configuration, where the customer tries to anticipate their maximum future needs. One advantage of this type of equipment is that if parts are monitored often enough, even abrupt catastrophic failures can be tracked and characterized as to their exact time to failure and how quickly they went from "normal performance" to being total failures. However, the main disadvantage of continuous monitoring equipment has historically been high price per channel due to the need for many photodetectors and expensive test equipment. The hardware has typically cost $1,000 per channel and up, which is more than 10× the price of the lower cost option without continuous monitoring.

The lower cost option is shown in Fig. 6.11. A board with test sockets is fabricated. Also, a rack with edge card connectors to hold six or more boards is made, with a wiring harness being soldered onto the edge connectors. A separate matching rack outside the oven is then fabricated, with a driver board fitting in each mating slot. Traditionally, the driver board has one current setting and drives all lasers on the test board at the same current. Depending on the configuration, it may have a feedback loop that drives each laser individually, or it may have lasers in strings of multiple devices (e.g., eight devices in series). Alternately, very low-cost device boards may use a voltage source with a series resistor.

Earlier we mentioned that a 150°C oven temperature is the standard ambient temperature for use in wearout tests, with additional cells being done at lower

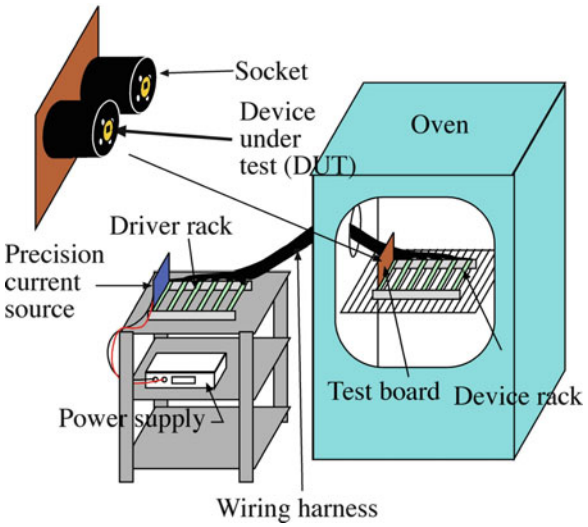



**Fig. 6.11** Low-cost test apparatus commonly used for accelerated VCSEL testing

temperatures to determine Arrhenius energy and verify that no lower activation energy mechanisms will eventually show up at lower temperatures. It is worth discussing hardware challenges of doing testing at 150°C and what needs to be done to go even higher in temperature if desired. First, it should be pointed out that almost all components that are commonly used in electronics are generally not suitable above ~105°C. Edge card connectors, FR4 board, standard lead-tin solder, and PVC-insulated wire all will give problems in a 150°C oven. These should be replaced with test sockets and edge card connectors with at least 150°C ratings (and ideally a 200°C rating): polyamide printed circuit boards, silver solder or high-temperature solder, and wire with Teflon insulation. Even if 200°C ratings are used, polyamide boards operated at 175°C or above will usually show "burning" after use and may need replacement a few times a year.

A few cautions are in order regarding procedures used to bias the VCSELs. The first is that many if not most manufacturers have experienced at least some problems with "EOS": electrical overstress from the reliability test equipment overdriving the parts well beyond the specified drive currents, if only for a fraction of a second. This is an extremely serious problem. Even if failure analysis is able to successfully identify the failure as definitely being due to EOS (e.g., metal contacts or wirebond totally melted), all other parts on the same board are under suspicion and may fail early. In many cases, the test is a total loss. While this may be recoverable if it happens early in the test, cells that are lost 6 months later may be unrecoverable and leave a hole in your test data, potentially delaying product release.

A number of precautions can be set in place to minimize the likelihood of EOS. The first is to set up the drive voltage from the power supply at the lowest practical level. Once this has been done, measure the peak current that is provided with the driver circuits wide open, if using closed-loop current sources at the highest oven temperature. This can often be measured by using a square wave to the board's power

supply input and monitoring the output on an oscilloscope; it will quickly peak, then rapidly drop as the circuit tries to bring it into compliance and finally level off at the desired current. This peak current should be at a level that the device can take repeatedly without damage, for a second or more. The second precaution is to place a surge protector strip between the outlet (or "mains") power and the power supply. The third precaution is to never hot plug and unplug boards; they should always be removed or inserted only when the power supply is completely off. The fourth precaution is to slowly ramp power supply voltage up and down (over a period of seconds), to prevent current inrush or inductive voltage spikes. Finally, the fifth precaution is that all driver boards are serialized and records are kept, and any boards implicated in suspect tests are removed from use until they can be retested and cleared. While more elaborate protective circuitry is available on the expensive "continuous monitoring" systems described earlier [35], the precautions listed above will work with any internally built system (as most are).

It is good practice to monitor that the drive current is at the desired set point for at least some of the devices, on a periodic basis, by monitoring the flow of current through a small in-series resistor provided for that purpose. This will validate that there haven't been equipment problems that might have resulted in under- or overstressing of the lasers. In addition, periodic auditing of the components with a calibrated milliammeter is worthwhile, in case the resistor monitoring method is inaccurate. The resistors should always be double checked on new boards as well before they are put into use.

The parts are tested at predetermined intervals for their L-I-V curves. This is done after they've been allowed to cool to room temperature. Test intervals are frequent at first and then increase as the test goes on, with the intervals increasing in a roughly logarithmic fashion, with each step being 1.5–3× the time of the step before. It is strongly recommended that at least a few early failures be "suspended" (i.e., removed from the test and archived for later analysis) at their failure point, to allow future failure analysis. This obviously requires extra analysis and technician labor but means that the devices will reflect how they looked nearer to the failure time, rather than having added defect structures that weren't relevant to their original cause of failure. If there is some debate as to what the failure point should be, this generally would not be any larger than a doubling of the threshold current, or a 50% drop in output power at nominal bias. Once the test intervals reach 1,000–2,000 h, they subsequently will be tested at that interval for the remainder of the test. The L-I-V data from the tester goes through a software routine that extracts key parameters, such as threshold current or power at a normal bias condition. Systems with low capital cost but high labor cost can involve an operator having to remove parts from the test board and place them into a test system one at a time. Better test systems allow an entire board with 20–100 VCSELs to be inserted into a system and data to be automatically collected within a minute or so, before the next board is inserted.

For thoroughness, some fraction of the aged parts should be tested under AC conditions at the datasheet maximum temperature, to make sure acceptable AC performance still exists. The original reliability problems with using CD lasers in

data links resulted from unanticipated changes in AC (high-speed) behavior that was not expected based on changes to the DC (steady state) parameters. While these problems seem unlikely to occur in future VCSELs based on the different physics of failures, and the results of the work that's been done so far, AC performance after aging to the failure threshold should be verified on at least a small sample of parts [36].

#### 6.4.4.2 How Much Acceleration Is Too Much

Part of the debate on how to set up an optimal reliability test matrix is "how much acceleration is too much," or "when is acceleration so high that it gives unrealistic results?" We propose three tests to answer this question:

1. Useful acceleration will allow you to "rank order" the reliability of a set of parts, from best to worst, as can be determined with slower aging closer to use conditions. Ideally, your highly accelerated results which say "parts from test lot A take twice as long to fail as those from test lot B" will be replicated with your much slower long-term test on the same batches of parts. If you can't replicate at least the rank order, the highly accelerated test may be failing due to a different failure mechanism that dominates at very high stress.
2. Useful acceleration should have a continuous and predictable increase in degradation rate as stress is increased (e.g., following an activation energy with a straight line across the Arrhenius plot). The presence of an abrupt "inflection point," on the Arrhenius plot, where the degradation rate increases much more rapidly with increasing stress is a good sign that some change in failure mechanism has taken place. A common example the way plastic-encapsulated integrated circuits cannot be tested above 150°C, since the encapsulant reaches its "glass transition temperature," reflows, and breaks wirebonds.
3. The cause of failure should be the same in the rapidly degraded parts as in more gradually degraded parts. The key signatures of failure should be the same in both populations, and the physics of failure should match. Otherwise, it's likely a new and different failure mechanism is present.

This having been said, far more acceleration can be usefully done than most people believed in the early days of VCSEL testing. A good reliability engineer will either get to the point where (a) he obtains a wafer qualification test that can age the parts to a 40-year equivalent in less than a week or (b) he has proven with extensive documentation that this is not possible using all three tests above. Nervous R&D managers and marketing people will want to set up very mild test cells that show that their lasers are "too good to fail," but in the end, rapid wafer qualification testing is absolutely essential to the manufacturing process.

We were able to get a wafer qualification process with $>1{,}000\times$ acceleration factor for VCSELs. The work we did on 2-Gbps oxide VCSELs showed that at currents past the L-I rollover point, the degradation rate went up more rapidly than the model called for [28]. Furthermore, in some of our earlier work using VCSELs

degraded to failure in 1–12 h [37, 38] which was done at much higher current, mirror damage appears to have occurred in ways that haven't subsequently shown up in parts degraded over thousands of hours [39, 40]. The conclusion is that aging to failure over just a few hours may be an overstress for most VCSELs, but aging to failure over a 4–8-day period is normally possible with good correlation and recommended as a routine wafer qualification test.

### *6.4.5 Maverick Failure Mechanisms: An Introduction*

Franklin Nash, in his excellent book [41] and reliability class [42], would point out that field reliability (the sort of failures you might see in a few years of normal use) is almost never measurable in laboratory testing using any of the techniques we've discussed so far. In fact, if you can detect such failures that might occur during use life in reasonably sized experiments, your parts may not be ready for field deployment (at least for parts in high-reliability applications). Please see Nash's book [41] for a further discussion with example calculations. The factors that dominate field failure are maverick failure mechanisms: failures due to "weak parts" that were not manufactured as designed. The way to prevent such failures was not to do huge reliability experiments. Such experiments cannot verify that something won't change next week that would render such an experiment irrelevant. Rather, the way to prevent failures was to have a change control system such as ISO 9000 as well as an FMEA process that brainstorms about what could go wrong and how such problems could be prevented or at least detected before they reach the customer. A strong failure analysis program, and corrective action for field and test failures, is also critical to success.

#### 6.4.5.1 The Difference Between Maverick and Wearout Failures

Let's explain the difference between maverick and wearout failures with some examples. The difference is critical – many of the mistakes in reliability engineering come from assuming that a problem is due to wearout failure when it's actually a maverick problem, or vice versa.

"Wearout" is a type of failure that occurs to many devices (and all semiconductor lasers) due to the process of "consuming" something that is needed to make them work. Tungsten filament light bulbs and tires are examples of items that have well-characterized "wearout" lifetime limitations. By contrast, "maverick" failure problems are ones that only apply to a small subpopulation of devices, which were not made as the designer intended. Two examples would be tire tread delamination [43] or the wirebonding problem shown below in Fig. 6.12.

Another common way of distinguishing maverick versus wearout problems in VCSELs is by looking at whether the failures are sudden (usually maverick failures) or gradual (usually wearout failures). This is shown below in Fig. 6.13.

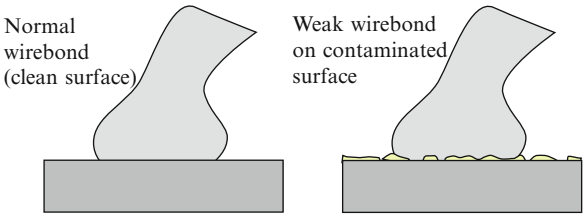



**Fig. 6.12** Drawing of a normal wirebond and a wirebond on a contaminated surface, to show one common way that a maverick failure mechanism might be generated

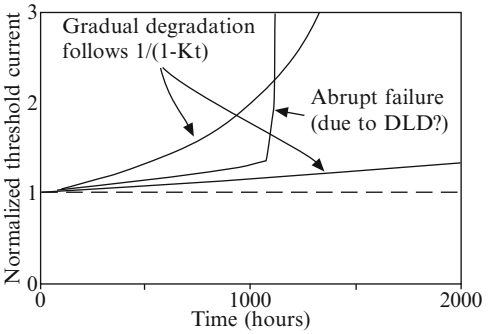



**Fig. 6.13** Abrupt failure (with a "kink" in the device parameters after some time) is generally associated with maverick failure

*Example 1. Assuming a wearout problem is a maverick problem*: The epitaxial growth recipe was changed to improve yield. During burn-in, an abnormally high fraction of the population failed. Further testing showed that extending burn-in time results in additional failures and that the majority of the parts will fail if the burn-in interval is extended to 10× its normal length. This is true for all wafers since the recipe change was made. The general manager, in an attempt to get production moving again, proposes "Let's add an additional device screening test, screen out the bad parts with short lifetime, and ship the parts that pass the test." The reliability engineer explains to him that all of the parts are bad and will fail prematurely, and that this is not a problem that is just isolated to certain parts of the wafer or certain devices. The epi recipe must be changed back, and the production line repopulated with good VCSEL epi wafers.

*Example 2. Assuming a maverick problem is a wearout problem*: Sporadic failure is observed on one device out of every third wafer qualification batch (of 50 parts) on a new product going through its qualification cycle. The quality manager proposes, "Let's just have a rigorous wafer qualification test; wafers that pass it can move forward with shipment, while wafers that fail will be scrapped." The reliability engineer explains that this problem is thought to be a 1% problem, that is, randomly distributed among wafers, and that it is not practical to determine whether some wafers have a higher incidence of this problem than others, since the population sizes would have to be very large, and even then, "passing" wafers might have just passed by being lucky during sampling. The reliability engineer advises pursuing deep failure analysis to find the root cause of the problem and design changes to eliminate the problem. A 1% field failure problem would be a very serious problem for this type of product.

### 6.4.6 Maverick Failure Reliability Testing

While it is notoriously hard to capture maverick failures in laboratory testing, it is not impossible, and a few steps will aid you in catching such failures and taking effective action to address them.

1. Have different stress types. As mentioned earlier, for accelerating wearout, 150°C testing is the most popular. This accelerates point defect motion in the electrically pumped area by "softening" the lattice and providing high thermal acceleration. However, optical pumping is greatly reduced at these high temperatures – most VCSELs aren't even lasing at 150°C, and spontaneous emission is orders of magnitude lower than it would be at, say, 100°C. Some maverick mechanism needs such optical pumping to be activated. If you have a diverse program that includes at least some cells at lower temperature, you've got a much better chance of catching failure mechanisms that have lower activation energy but different physics of failure.
2. Test a large number and wide variety of parts. There is no substitute for testing volume. The large VCSEL manufacturers show 2,000 or more devices at one condition in their reliability documentation. This at least gives you some hope of catching a 1% problem.
3. Test frequently. This is taken care of to some extent if a wafer qualification is performed on every wafer. But periodic testing at other stresses, as mentioned above in #1, should be repeated from time to time. It should be done on a quarterly or at least annual basis.
4. Look carefully at the data, and arrange to be notified if any parts that passed burn-in fail before the expected range of times, based on the 99% confidence limits for the overall distribution.
5. Don't dismiss the early failures as being "just a test artifact." Sufficient staffing and expertise are critical in being ready to do this sort of follow-up work. Perform FA and deep FA as needed to figure out if this is a different failure mode. If it is, determine why it happened and how to prevent it from happening again. If you don't, you may be seeing more of these in the future as customer returns!

### 6.4.7 Grouping Failures from Similar Mechanisms

The early failures from many of the test cells need to be studied, particularly with L-I-V testing and subthreshold electroluminescence (EL), as discussed later in the failure analysis section of this chapter. Do they look similar to one another? Or do they show substantial differences? If multiple failure mechanisms are present, the proper way to analyze them is not to just have all parts lumped together but rather to use a multiple population distribution. This is easiest if some signature or quick FA technique can be used to sort the failures into various "categories" or "buckets," and

then analyze each bucket independently. Please see Chap. 4 in Finn Jensen's excellent book "Electronic Component Reliability" for a further discussion [44]. Along with Nash's book [41], I have found these two the most helpful and relevant to semiconductor laser reliability.

## 6.5 VCSEL Reliability Analysis

Once the reliability data has been collected, it is time to analyze the data to draw conclusions about what the strengths and weaknesses are for the VCSELs tested.

### *6.5.1 The Stress-Strength Model*

It is helpful to think of laser degradation in terms of the strength of the parts and the stress that's being applied. As the aging weakens the parts, they will reach the point where they no longer have adequate strength and at that point they are logged as a failure [45]. The width of the strength distribution will determine how wide the width of the failure time distribution is. This is depicted below in Fig. 6.14. It should be self-evident that to maximize lifetime, you want the highest strength to start with, the lowest rate of strength decline, and the lowest operating stress you can have and still get adequate performance. More complex versions of this figure include a separate "low strength" lobe of parts that fail burn-in or fail during normal use life.

### *6.5.2 Sources of Variability in Wearout Lifetime*

Variability in wearout reliability comes from a number of different sources and contributes to the observed distribution. Reliability data is typically analyzed with a lognormal distribution, and the larger the sigma, the wider the variation in failure times. When we have studied laser arrays or neighboring parts, we typically find low values of sigma, typically 0.15–0.2. If you study parts from across an entire wafer, depending on the uniformity, our observed values of sigma varied from 0.3 to 0.4. For oxide VCSELs, aperture size variations also would cause a significant range in degradation in fixed current experiments, while in actual use, smaller devices would be operated at lower currents and thus similar current densities. Finally, the variations in epitaxial quality, as discussed earlier, would contribute variation. The overall result is shown below in Fig. 6.15. Overall sigma for the full population of lasers produced over years is typically 0.5–0.7.

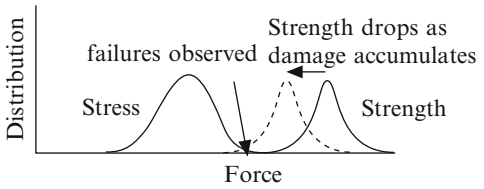



**Fig. 6.14** Schematic drawing of the distribution of aging stress and strength to resist degradation. When the strength falls below the applied stress, the device will fail

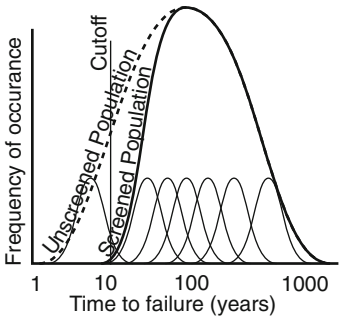



**Fig. 6.15** The overall distribution of failure time is made up of a combination of individual laser lots, with overlapping times to failure, that make up a wider reliability distribution

### *6.5.3 Parametric Analysis of VCSEL Degradation*

If you carefully study how VCSEL parameters vary with time in accelerated life testing, you'll see two trends superimposed. First, an initial transient can be seen, with sublinear slope, that levels off with time. Due to the complexity of the different types of point defects, traps, and free hydrogen, the initial transient can result in improvement *or* a degradation as these migrate through the structure. However, over the long term, the device approaches some stabilization, usually reaching 90% of the final change after the equivalent of a few years of worst-case use service. Second, superimposed on this is a superlinear term that can increase quite rapidly as the device approaches the failure criterion. This term captures the wearout degradation of the device. If you take any given lot, even for highly reliable VCSELs, you will see a fairly large difference between the best and worst devices: it is not unusual to see the longest-lived devices take ten times longer to fail than the early failures, even in modest population sizes (i.e., a lognormal sigma of 0.6 is common). For the best-behaved lasers, the sublinear (i.e., diffusion) behavior dominates, and little evidence of additional degradation is evident even after 20–30% of the VCSELs have reached failure on the life test. However, for the earliest devices that fail, the superlinear (or "degradation") term dominates, and the sublinear diffusion term can be ignored for all but the most precise of estimates. In terms of "why" this term increases in a superlinear manner (i.e., showing upward curvature like an exponential growth term), we can only speculate: heating due to optical absorption and optical pumping are likely to be the two most important factors.

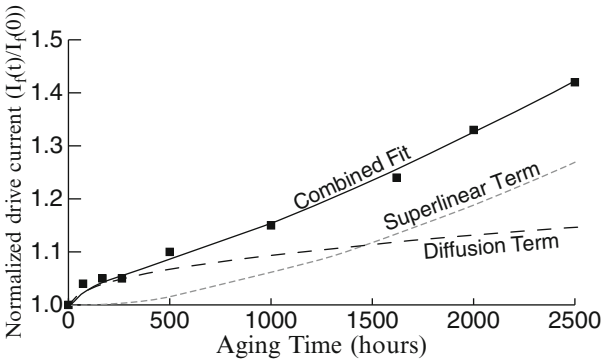



**Fig. 6.16** Data from the median device out of a group of 40 devices is shown in *black squares* and superimposed on the best fit line, which is a combination of two terms, as shown below in Eq. 6.3.

A sample device from a good-quality test run, which failed in approximately the median of its group, is shown in Fig. 6.16. One thing should be very clear, especially in studying devices that degrade more rapidly than this: any attempt to project time to failure using low stress cells will result in very dangerous, overoptimistic results that could give gross overestimates of expected lifetime. This is because you will be studying only the diffusion term and will not have the benefit of seeing the superlinear term within reasonable test durations.

Equation 6.1 captures the simple case. It is a formulation used by Waters [18] and has only one fitted parameter, $K$, which typically varies from $5\times10^{-3}$ $h^{-1}$ for lasers at very high stress to $1\times10^{-5}$ $h^{-1}$ for lasers under very low stress:

$$I(t) = \frac{I(t=0)}{(1-Kt)} \tag{6.1}$$

Where degradation is slower, where diffusion is large, or where a more precise fit is desired, Eq. 6.2 can be used. It has three fitted parameters. The middle term covers the diffusion that levels off, with a "Fick's law" formulation that takes the square root of the time, and has a fitted "$c_{diff}$" term that increases for higher stress or less stable lasers. For the data set covered in this section, values of $c_{diff}$ ranged from a high of $1\times10^{-2}$ to a low of $2\times10^{-3}$ $h^{-1/2}$. The right "power law" term covers the superlinear term as aging accelerates laser degradation and brings the VCSEL to the end of life. Two fitted parameters are used: a coefficient "$c_{deg}$" and a power law exponent "$n$" that increases as the "turn-on" becomes more abrupt. $c_{deg}$ is high for rapidly degrading lasers or low values of $n$ (e.g., $5\times10^{-5}$); for high values of $n$, and better stability in the laser, $c_{deg}$ was as low as $3\times10^{-10}$ for this data set. $n$ varied between 1.4 and 2.5 in this data set. For VCSELs with very abrupt degradation turn-ons, $t^n$ in the right term will need to be replaced with $(t\text{-}t_{incu})^n$, where "$t_{incu}$" is a fourth fitted parameter to capture the incubation time before the degradation cuts on.

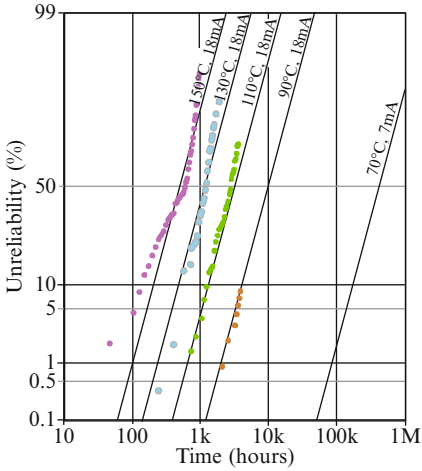



**Fig. 6.17** Weibull plot of lasers from the same seven wafers at four different temperatures (Courtesy of Chris Helms)

This fourth fitted parameter wasn't needed for this type of laser but has been needed elsewhere:

$$I(t) = I(t=0)\left[1 + c_{diff}\sqrt{t} + c_{\deg} t^{n}\right] \tag{6.2}$$

Once data is available, it is unfortunately required to spend some effort "reducing it" to something that approximates a table of "times to failure" or "failure intervals." Using intervals is easiest and will work well provided the parts are checked frequently or have a higher sigma value (i.e., a wide enough spread in failure times that no more than 10% of the parts fail in most of the intervals). If failures are clustered heavily in just a few bins, it may be desirable to use linear interpolation or curve fitting, as shown earlier in Fig. 6.10. Note that while some automated tools are available for fitting parametric data (e.g., "Degradation Analysis folios" in Weibull++ version 7), we have not seen these work well with VCSEL data, so very careful scrutiny is needed before accepting any fitted parameters.

### *6.5.4 Determining an Acceleration Model*

The key to accelerated testing for VCSELs is using the right combination of thermal and current acceleration to speed up the aging that would normally take years of normal operation and get the results in a matter of months. The Weibull or lognormal analysis needs to be repeated for each cell (i.e., each temperature and current test lot) in the qualification matrix. A sample Weibull plot is shown in Fig. 6.17. The values for 1% failure, 10% failure, and 50% failure can then be filled in. Once this has been done, a serious analysis for activation energy and current acceleration factor can begin in earnest. Packages such as Reliasoft Alta [46] can be

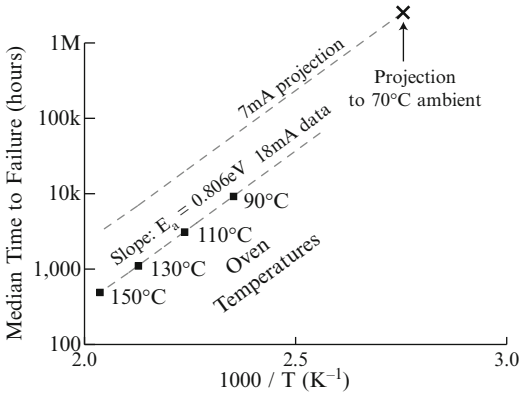



**Fig. 6.18** Arrhenius plot, using the data shown in the previous figure

used to simultaneously solve for the best fit to both the thermal and current models, but we have found it very difficult to trust "least squares fitting" since it is often thrown off by outliers, so we prefer to do much of the analysis by hand. Many different Arrhenius and thermal acceleration plots are created, such as those shown in Fig. 6.18. Early on, when one is often rushing to perform the analysis, data from only a few of the most accelerated cells may have a sufficient number of failures available for analysis, especially if too little acceleration was used in the high-temperature and high-current cells. One "temporary trick" that can be used is to set the failure criterion artificially low (e.g., a 10% increase in laser threshold), with the hope that these earlier times to failure will all scale equally (e.g., all might be close to 1/3rd the time to the actual failure criterion). At this point, it may be easy to get discouraged, as truly neat, conclusive data that fit the expected behavior are rare – the good fit shown in Fig. 6.18 is the exception, not the rule. Temperatures in the Arrhenius analysis should use the *junction* temperature, *not* the ambient temperature. This therefore requires knowledge of what the thermal resistance is for the VCSEL. The thermal resistance is normally characterized by measuring spectral shifts as a function of shifting stage temperature and injected current [7], but confusion among modes can make accurate measurement difficult. Thermal resistance varies approximately as the reciprocal of the oxide aperture diameter. One also has to measure how hot the test vehicle (usually TO cans) is above the ambient temperature to accurately understand the test results. Errors in any of these can result in projections being significantly off from the correct values, so it's worth spending time to make sure these are accurately known.

The model for degradation of VCSELs is nearly always thought of as being the product of two elements: a current acceleration factor and a thermal acceleration factor. The current acceleration factor is typically taken to be the ratio of the life test accelerated current divided by the normal operating current to some exponent "*n*." This exponent "*n*" was found by Waters and Bertaska to be "1" for edge emitting lasers in their excellent study [47]. However, for VCSELs, the exponent has been found to be higher and is conservatively assumed to be equal to "2" in most models. The thermal acceleration is characterized by an "activation energy" or "$E_{a.}$"

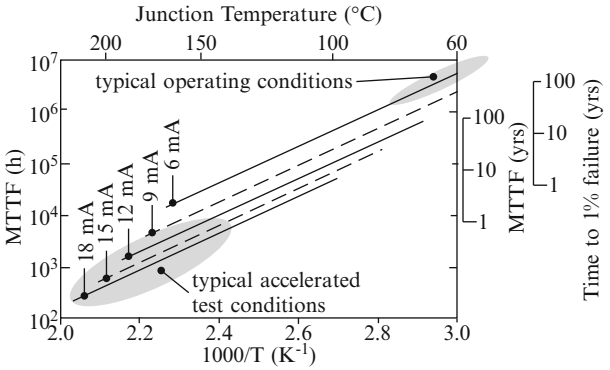



**Fig. 6.19** Projection of MTTF and T01 versus temperature and drive current for idealized VCSEL reliability model on 12-μm oxide VCSEL. One needs to be cautious with such a model to account for worst-case lasers, which could be significantly weaker than an "average laser"

It is shown below in Eq. 6.3. The model should then be applied to form graphs to support engineering design efforts and set acceptable temperature and bias limits. A "big picture" graph is shown in Fig. 6.19; it shows the many orders of magnitude in degradation rate between where lasers are tested at and where they are used and hopefully makes clear how someone can qualify a laser for decades of use based on only a year or so of testing.

$$\mathrm{AF} = \left(\frac{I_{acc}}{I_{use}}\right)^2 \left[\exp\left(\frac{E_a}{k_B}\left({}^1\!/\!_{T_{j,use}} - {}^1\!/\!_{T_{j,acc}}\right)\right)\right] \tag{6.3}$$

Equation 6.3 expresses the "acceleration factor" as the product of the "current acceleration factor" and the exponential term that is the "thermal acceleration factor."

### 6.5.5 *Using Spreadsheets for Reliability Calculations*

The calculations needed for VCSEL-based applications are often similar to those used for other semiconductor laser applications.

It is often necessary to create graphs of percentage failure versus time, for various bias currents or operating temperatures. In such a case, a model is created with the expected time to failure for a given accelerated condition (usually the primary wafer qualification condition). Ideally this would be the "worst-case passing wafer," although at a minimum this needs to be what can be routinely expected by average or below-average passing wafers. This is then brought back to a realistic use condition, usually 70°C and a reasonable bias current that is high enough to work for most VCSEL. The time $t$ (in hours), the median life value "$\mu$"

(i.e., the natural log "ln" of median time to failure $T_{50}$ in hours, which typically ranges from 12 to 15.5), and sigma ("$\sigma$") (usually 0.3–0.8) are then plugged into the spreadsheet function "lognormdist($t$,$\mu$,$\sigma$)," which yields the fraction of failed devices of the cumulative distribution function $F(t)$. If a Weibull function is desired, the spreadsheet function used instead is "Weibull($t$,$\beta$,$\eta$,true)," where the shape factor $\beta$ usually ranges from 2 to 3, the characteristic life $\eta$ usually ranges from $5\times10^5$ to $5\times10^7$ h, and the "true" flag tells the function to give the cumulative failure distribution. The inverse function is also frequently used in making projections: finding the time required until 1 or 0.1% of failures will be observed. This can be calculated with the "loginv($F$,$\mu$,$\sigma$)" function, where $F$ is the fraction of failed devices desired. The result returned is the time in hours until that cumulative failure rate will be reached. For those using the Weibull function, the time would be calculated with the formula $t = \eta[-\ln(1-F)]^{(1/\beta)}$. The aforementioned "lognormdist," "weibull," and "loginv" functions are built into most spreadsheets, including Microsoft Excel, OpenOffice, Google Docs, and Apple iWork "Numbers." In addition, the functions can be found in Reliasoft's Weibull++ "general spreadsheet" as well as Matlab and other computational software. At times, it is useful to build multipopulation models, with a weak population (e.g., burn-in failures or maverick devices) as well as the dominant population we have been analyzing here. In such instances, good values to use for the lognormal function would be $\sigma = 2-3$, or $\beta = 0.6-0.8$. $\mu$ or $\eta$ is adjusted to reflect either simulation needs or data collected on burn-in failures or field failures. The fraction of the weak population typically will be less than 5%.

*Example 1*. Calculate the cumulative percentage having failed after 15 years, with a lognormal distribution having a median life $T_{50}$ of 600,000 h and $\sigma = 0.58$. Solution: time = 15 years$\times$8,760 h/year = 131,400 h. $\mu$ = ln (600,000) = 13.3. $F(t)$ = lognormdist($t$,$\mu$,$\sigma$) = lognormdist(1.314e5, 13.3, 0.58) = 0.0045 = 0.45%.

*Example 2*. Calculate the number of years to 0.1% failure for a part with $\mu = 14.5$ and $\sigma = 0.82$. Solution: $G(p)$ = loginv (0.001, 14.5, 0.82) = $1.57\times10^5$ h/8760 h/year = 18.0 years.

### 6.5.6 *Confidence Limits in Projecting Accelerated Lifetimes*

After calculating the best fit to the data collected, normally the engineer is asked to put confidence limits around them. It is worth discussing how this is normally done and then discussing the limitations.

First, based on the limits of the statistical sample size, there is uncertainty in both the true median value (which will affect $\mu$ or $\eta$ for lognormal or Weibull distributions, respectively) as well as the shape factor ($\sigma$ or $\beta$). These combine to make the uncertainty smaller in the middle of the distribution (the spread at the 50% failure level) and progressively larger as you're trying to project the early failure rates that you care most about. Normally, the calculation of the confidence limits is

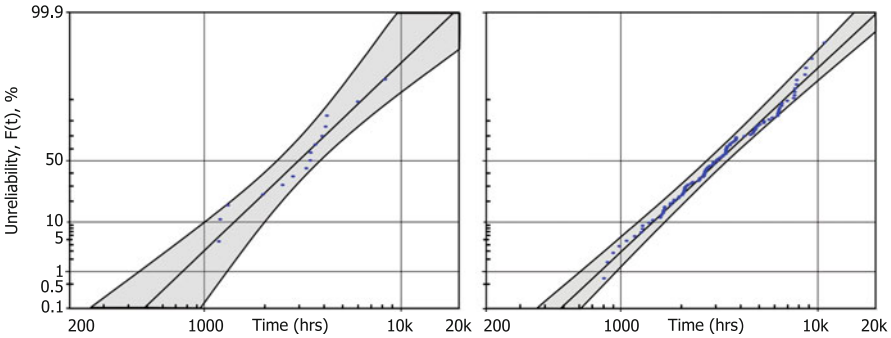



**Fig. 6.20** Lognormal plots with sample of only 15 (*left*) has wider confidence limits than same distribution with 100 parts (*right*)

done by simply entering the failure data into a program such as Reliasoft's Weibull ++ and allowing it to make the calculations using Fisher confidence matrices. Note that there are two options: the more commonly used "type I" confidence limits, calculated on what range of times could be possible for any given cumulative failure rate, or "type II" confidence limits, which calculate the uncertainty of percentage failures seen for any given time. When overlaid on the same graph, the two are somewhat similar but different enough to not be interchangeable. The user should be careful to choose the one relevant to the question they're trying to answer. One minor caution would be one that's obvious to anyone schooled in statistics: a 90% two-sided confidence limit gives the same boundaries as two separate single-sided 95% confidence limits. So be careful in choosing the options in the program. Ninety percent two sided or ninty-five percent one sided is the industry standard. Of course, confidence limits get smaller with larger samples, as can be seen in Fig. 6.20. They also vary depending on covariance (how closely the data falls on the line). The best discussion of methods for calculating confidence limits can be found at the website owned by Reliasoft: www.weibull.com; this also includes formulas, detailed discussion of various methods of calculation, and examples.

Looking at additional sources of error, these confidence limits are only statistically determined (based on sampling theory) and do not include many important sources of uncertainty. For example, the range of wafer quality mentioned in Sect. 6.5.2 (Fig. 6.15) is *not* captured in these confidence limits. More importantly, the uncertainty on the activation energy and current acceleration factor is not accounted for in such calculations. To do so, one would use a product such as Reliasoft's ALTA; the computation of such confidence limits from a matrix of different acceleration conditions can be quite complex. Depending on sample sizes, and how far from use condition your test cells are, it is common for the confidence limits at the use condition to be +/− two orders of magnitude (yes, 0.01× to 100× the best fit line!). In practice, no one publishes the very unsatisfying confidence limits, since customers wouldn't be very happy to know that their product could fail

in 0.5–5,000 years. But a good reliability engineer will account for these additional uncertainties by leaving some additional design margin present when setting pass/fail limits, particularly for newer products.

### 6.5.7 Calculating the Ceiling on the Random Failure Rate and FIT Rates

Finally, the analyst will be called on to calculate the statistical upper limit on the hazard rate from use-level tests. Unlike many of the wearout tests described earlier with extreme acceleration, qualification tests are run at modest conditions (often the upper bounds of datasheet conditions), and any failure results in the part not passing qualification. A cumulative time "$T$" is calculated from the product of the length of the test, the number of parts run, and the acceleration factor of the test over the use condition desired (usually the worst-case use condition, leading to the smallest acceleration factor). Assuming an exponential distribution with a constant (or declining) failure rate, the upper bound on the hazard rate $\lambda$ is then given by the formula [44]:

$$\lambda = \frac{\chi^2/2}{T} \tag{6.4}$$

This uses the "chi-squared" distribution, where chi-squared is a function of the confidence level desired (usually both 60 and 90% confidence will be calculated) and the number of failures observed (usually 0, although occasionally 1 or 2 for larger tests or systems). For the zero failure case, $\chi^2/2 = 2.305$ for 90% confidence and 0.916 for 60% confidence. For multiple failures, chi-squared can be looked up in tables or calculated in spreadsheets using the "chiinv((1-CL/100),$v$)" function where CL is the confidence level in percent and $v$ is the number of degrees of freedom and is calculated using $v = 2r + 2$, where $r$ is the number of failures on the test.

*Example*. Two failures were observed on a 2,000-h-long test with 100 devices, run at 5 times higher stress than normal worst-case use stress. Calculate the upper bound on the fit rate at the 90% confidence level. Solution: $T$ = number of devices×length of test×acceleration factor = 100 devices×2,000 test hours×5 equiv use hours/test hour = $1\times10^6$ equivalent device hours. Find chi-squared using spreadsheet: chiinv((1−90/100), (2×2 + 2)) = chiinv(0.1, 6) = 10.64. $\lambda = (10.64/2)/10^6 = 5.32\times10^{-6}$ h$^{-1}$. This would ordinarily be expressed in "FITs," or failures per $10^9$ h – perhaps more usefully thought of as "parts per million failure rate, per kilohour of operation." So in this case $\lambda$ = 5,322 FITs – not a very good result for datacom applications.

Most customers are looking for the manufacturer to have completed testing long enough and on a large enough population of parts to show that the ceiling is below

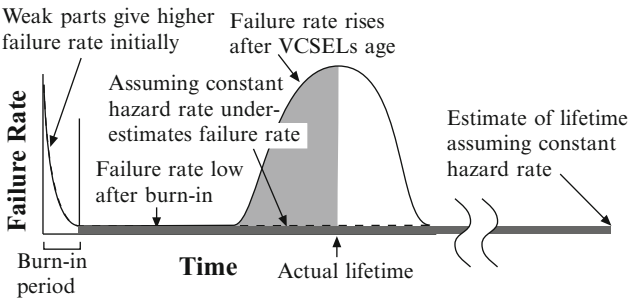



**Fig. 6.21** Plot of failure rate versus time shows the correct distribution with the curve as well as the incorrect very long projection assuming a constant failure rate from using only the early data

250 FITs before they're willing to take the risk of field deployment. This is most certainly *not* to say that 250 FITs is an acceptable rate of failure in the field for datacom applications. Even 5 FITs has been sufficient to cause customer crises when being shipped into banking centers, and with the proper precautions, it is possible to get down to fractions of a FIT field failure rate.

### *6.5.8 Overoptimistic Projections of Reliability*

There are a number of ways that it is possible to come to a projected VCSEL reliability result that is overly optimistic. The most common methods, in order of popularity, are shown below. Customers of VCSELs should be aware of these methods and assured that these methods are *not* being used in making projections for products that they are purchasing.

1. Assume too low of an operating temperature (see Sect. 6.4.2); the user should have an idea of the actual header operating temperature that is likely to be in use.
2. Use extrapolation to a projected time to failure, based on small amounts of degradation (e.g., extrapolating time to 50% degradation based on time to 5% degradation), rather than using an actual time to failure (see Sect. 6.4.4.1).
3. Use exponential statistics when applied to a wearout population (see Fig. 6.21) [41]. For example, if we were to use actuarial data on people 40 years old and younger, and assume failure rate is constant with age, we would come to the conclusion that median age at death is approximately 250 years old, with a 50/50 chance of living another 250 years. Again, this is primarily a risk encountered when avoiding high-acceleration life tests.
4. Ignore maverick failure modes (see Fig. 6.22 below) [42]. We discussed maverick failure mechanisms earlier in the Sect. 6.4.5 and the fact that they dominate field failure rates but usually are not easily captured by laboratory testing of reliability. Note that most VCSEL wearout activation energies are 0.6–0.9 eV (the steep

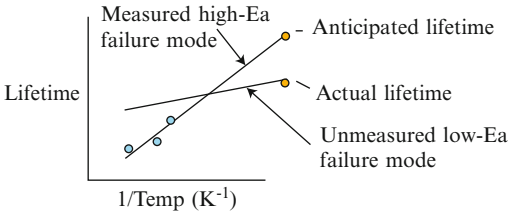



**Fig. 6.22** Arrhenius plot uses highly accelerated data points at the *lower left* to project the use lifetime at the *upper right*. However, if low activation energy failure mechanisms exist, they may dominate the lifetime at use temperatures, as shown with the *line* with the *lower slope*

"measured" line shown); most maverick failure mechanisms are 0.15–0.4 eV (the flatter upper line in Fig. 6.22).

5. Take a single highly accelerated result, then use assumed activation energies, distributions, etc., and make projections based on it, without providing evidence of low degradation rates using longer-term lower-acceleration tests of your own.

## 6.6 What Can Go Wrong: Common VCSEL Failure Mechanisms

In this section, we'll cover the most common failure mechanisms that have been observed in VCSELs used for data communications applications. It will be in approximate order of the number of failures seen. It should be kept in mind that the failures still are in the few parts per million of all the parts shipped, but it took a great deal of work to keep the failures at a low level.

### *6.6.1 Maverick Failure Mechanisms*

#### 6.6.1.1 Mirror Oxide Climb Dislocation

The most serious VCSEL return problem has been due to climb dislocation failures. As mentioned earlier, climb dislocations are also known as "⟨100⟩ dark-line defects" or "DLDs" and grow very quickly once they get close to the active region [23. This occurred most commonly due to cracking at the edge of the "mirror oxide" [48–51]. In oxide VCSELs, the mirrors are made out of stacks of low aluminum AlGaAs and high aluminum AlGaAs. In the high aluminum stack, the mole fraction $x$ for the $Al_xGa_{1-x}As$ is typically 0.92. An even higher mole fraction layer with $x = 0.98$ is used to confine the current and also as the optical "lensing" or "waveguiding" element. Ideally, the oxide layer would oxidize, and the mirror layers would not. However, in real life, the oxide layer typically oxidizes perhaps 15 μm, and the

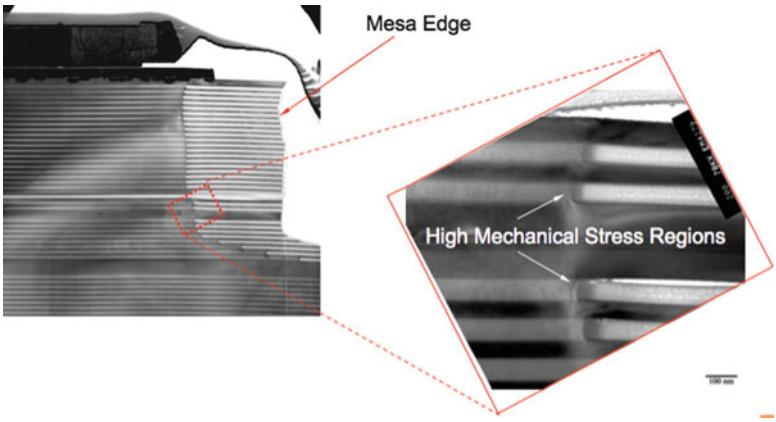



**Fig. 6.23** Cross-sectional TEM of edge of oxide VCSEL, with the area at the edge of the mirror oxide highlighted. This is the vulnerable area, where dark-line defect (DLD) networks can originate

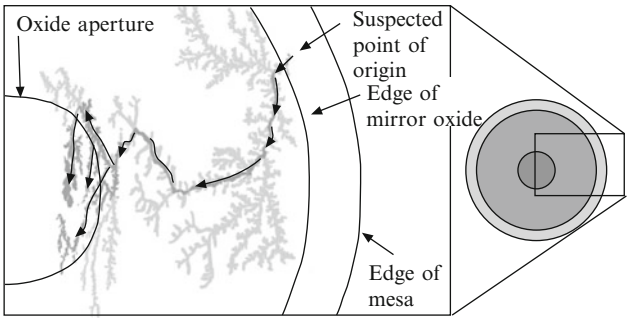



**Fig. 6.24** Trace of plan-view TEMs of a field failure that originated at the edge of the mirror oxide

mirrors typically oxidize about 4 μm. Unfortunately, the mirror layers are thick, and there are many of them, so the cumulative stress is very high. Occasionally, it is high enough that delamination occurs, and "punch out dislocations" go through the active region at a spot near the edge of the mirror oxide [50].

The structure is shown in Fig. 6.23. The climb dislocation then travels toward the emitting area, apparently by optical pumping. This process can take years to occur. An example of a field failure that was carefully analyzed is shown in Fig. 6.24 [23, 49]. There is some field return data as well as laboratory evidence that says that the problem varies substantially on a wafer-by-wafer basis, but no one has ever discovered conclusively what accounts for the variability in the failure rate. While it is known that grading the oxide layers helps and that the oxidation conditions make some difference, many of the companies have eliminated the problem by taking steps to electrically inactivate the edges of the VCSEL, which prevents motion of the climb dislocations.

Failure Analysis for Climb Dislocations: The lowest cost technique is plan-view TIVA (see the failure analysis section later in this chapter). The existence of climb dislocations leading to the mesa edge (for mesa VCSELs) or oxidation holes (for planar VCSELs) makes it highly likely that the dislocation occurred due to the mechanism described in this preceding section (i.e., DLDs originating from the mirror oxide). Plan-view TEM is more expensive but can be used if desired to confirm the point of origin and look for punch-through dislocations going through the active region that are the point of origin for this type of defect.

#### 6.6.1.2 Electrostatic Discharge (ESD)

VCSELs have a reputation for being extremely sensitive to ESD or poor handling, which is only partially deserved. Due to the expense of deep failure analysis, and the difficulty involved, many devices that get placed in this category may have little or no evidence to support ESD as a root cause. Later, we will discuss in the failure analysis section ("Identifying ESD," on the next page) what evidence we *do* find convincing.

##### Precautions

ESD control is a subject that would take multiple volumes to address, so we can't do it justice in this chapter, but we will make a few suggestions to get the reader started. Every VCSEL manufacturer as well as their customers should have an employee who is the "ESD control program manager." This person should be genuinely interested in the subject – we have had many unsuccessful experiences trying to "assign" personnel this title, only to have them ignore ESD in favor of their other work responsibilities. A substantial investment in reading [52] and education [53, 54] needs to be made at the start, as well as instruments for ESD auditing [55].

In the past decade, a new paradigm has been identified, of avoiding contact to metal without first bleeding off the charge with static dissipative materials in a controlled manner (μSec or mSec, rather than nSec for full discharge) [56, 57]. Special sockets, tweezers, and fixtures can be found that eliminate these risks. Such items are necessary if using photodetectors that are directly wirebonded to the fixture but not always essential for VCSELs. Low tribocharging shoes, flooring, and gloves are also available and worth pursuing in any ESD improvement program.

##### ESD "Charge Device Model" Case Study

One example of a nonobvious ESD threat occurred from a type of ESD called "charge device model" (CDM) ESD. This type of ESD occurs when a static charge builds up on a part, and it then rapidly discharges (usually from touching metal). The pulse width is only 1 nSec It was found that shipping packaged "TOSA" assemblies

with plastic barrels in bags allowed such movement, tribocharging, and subsequent discharging when the leads touched. Shipping in trays is therefore recommended after the parts have been packaged in plastic packages. Note that VCSELs are not particularly sensitive to CDM compared with most electronic devices. The devices we tested required 1,200 V (and 1.2 nC) to fail. The relatively high CDM failure threshold relative to electronics appears to be due to greater spreading of the power dissipation [58] than occurs in electronics where it is concentrated in a thin channel. Note that handling typical VCSEL plastic assemblies with non-ESD-rated gloves is capable of building up levels of charge that exceed these levels.

### Human Body Model (HBM) or Machine Model (MM) Damage

While VCSELs are not particularly sensitive to short pulse-width damage from CDM, longer pulses will cause failure at much lower voltage levels. Unfortunately, if this damage occurs after the VCSEL burn-in, it can cause field failure months after deployment. A number of papers on the subject have been published [23, 59, 60]. Reported values of HBM damage can occur in the 100–200-V range – something that requires special precautions during assembly to prevent.

### Identifying ESD in Failed Devices

If one is trying to identify whether ESD caused failure, one needs both quick and also more detailed and positive methods of identifying ESD. Some returns will have melting on the surface or melted wirebonds, so the diagnosis of EOS is simple. The next easy method of identifying ESD is reverse bias I-V. A good VCSEL should have a very sharp reverse breakdown. If a "soft knee" is identified, with a lot of reverse leakage visible even in a linear plot with a microamp scale on the *y*-axis, it is highly likely ESD has occurred. However, one needs to be very careful to set up the sweeper or parameter analyzer to put a series resistor in place and control current to 10 μA. Avalanche breakdown is so rapid during the reverse sweep that the closed-loop control circuitry in many parameter analyzers cannot clamp the current quickly enough and devices with minor damage will "crater" and show melt spots after going through such a sweep. That is, the I-V sweep will be a destructive test rather than a nondestructive one. A resistor of 100 kOhms in series with the fixture would make it unconditionally safe for such reverse sweeps. A third test that can be done is electroluminescence at low current and looking for small black spots along the edge of the oxide aperture, at least if the device can be examined immediately after ESD and before it has operated very long. (However, once the device has failed, these damage spots are usually obscured by climb dislocation growth.) Reverse bias photoemission can be another indicator, although this has not been extensively used with devices that have been further aged, so we are not clear on the consistent usefulness of this technique. For examples of all three methods, please see Krueger's paper [59].

More positive proof of ESD requires TEM. Both plan view and cross section have been used, and with sufficient time and money, positive proof of ESD breakdown or thermal damage can often be seen [61]. In the absence of such evidence, or strong indicators from the tests in the previous paragraph, one should refrain from making strong claims that the device failed due to ESD.

#### 6.6.1.3 Electrical Overstress (EOS)

ESD occurs due to static charge that builds up on the device. This type of charge creates a pulse that lasts 50 uSec at the longest. But testers with switching transients can create much longer pulse-width overstress events that are referred to as "EOS." Such events can occur with much lower levels of voltage stress; as little as 5 V is sufficient. Such events have resulted in VCSEL field failures. During the development process, VCSELs should be characterized with a pulse source to see how large of a pulse can be applied before damage occurs – either immediate damage or latent damage after accelerated aging. This should be repeated for a number of pulse widths. A failure "signature" should also be established, so the FA engineer knows what to look for if EOS parts are seen in future failures. All test equipment needs to be carefully audited for "spikes" using oscilloscopes to make sure no switching transients are occurring. To the extent test equipment can have its maximum voltage set to lower levels, it can help to eliminate the risk of EOS. In addition, it appears that design plays a role. Many "research" VCSELs are published with the high-index DBR layers using an $Al_xGa_{1-x}As$ composition of $x = 0.12$, which we have calculated may result in >10% of the spontaneous emission being absorbed in the mirror layers; see Chap. 4 of my PhD dissertation for an in-depth discussion of this issue [62]. Higher-bandgap DBR layers with $x = 0.16$ are used in production designs for this reason, among others. For failure analysis, thermal damage exists in larger areas. Often "Newton rings," with rainbow-colored patterns, will surround the area of damage or discoloration of the top metal. SEM inspection of the top of the die will also show evidence of thermal damage in most cases where serious EOS has occurred.

#### 6.6.1.4 Stacking Faults, Particle Smashing

All GaAs active region semiconductor lasers are vulnerable to impact damage. Contamination such as dust particles that get on photomasks (for contact aligners) or onto pickup nozzles (for pick-and-place systems) can cause cracks or stacking faults in the parts below. However, this has traditionally been a relatively rare cause of failure. The cracks can easily be seen with an infrared microscope that uses a wavelength cutoff filter that only looks at wavelengths with energy below the bandgap [63].

#### 6.6.1.5 Scribe or Sawing Damage

Scribe damage propagating in from the edges of the die has always been a risk with GaAs-based lasers [18]. Usually, some method needs to be used of wet etching away a trench to scribe through or otherwise electrically inactivating the material at the die edge. Either way, the trench or isolated area must be wide enough that scriber cracks or saw damage will never propagate into the electrically active region. Backside TIVA is the best method for identifying the damage source for such parts (please see the FA section of this chapter).

#### 6.6.1.6 Epitaxial Defects

Epitaxial defects are another classical source of failures for GaAs active region lasers. In fact, in the early days of semiconductor lasers, they were the dominant failure mechanism. Nowadays, they are relatively rare. They could occur either due to the built-in threading dislocations in the substrate or due to particulates falling off of the ceiling of the epitaxial reactor and onto the substrate before epi growth and thereby inducing a threading dislocation. They should be screened out through multiple stages of screening. First, after epitaxial growth, every wafer should go through an epi defect count in a fixed area. If the concentration of defects exceeds a certain level, the wafer should be rejected, and the source substrates should be checked, along with the cleanliness of the reactor roof. In the latter case, going too long between reactor cleaning allows very thick layers of epi material to build up on the roof and flake off. Second, a visual check needs to be performed on every die before it is shipped to check for epi defects and cracks – either a manual or machine vision inspection. If one is present on the die, the die should be rejected. Note that for mesa-style devices, or ones with most of the die electrically inactivated, epi defects are of much less concern.

Failure Analysis: So long as epitaxial defects are not covered by thick gold, they can easily be seen if a Nomarski microscope is used. Backside TIVA will give conclusive evidence if dislocations start from epi defects.

### *6.6.2 Wearout Failure Mechanisms*

Because maverick failures are difficult to screen out in low levels, they have been responsible for the large majority of the field failures seen by customers. By contrast, the wearout failures listed below are almost always easy to detect provided a rigorous wafer qualification program has been put in place. However, if the problems below arise, they have shut down many manufacturers for 1–4 months before supply was restored. Fortunately, in many cases a significant buffer stock of material was available, which limited the business damage.

#### 6.6.2.1 Copper Contamination

Copper contamination has caused many premature failures in parts where no wafer qualification procedure is used. Copper is highly mobile and can move from its initial location in the p-metal or the package where it is harmless to a location near the active region where it degrades laser performance. This is because copper acts as a trap, causing a large fraction of the carriers to recombine nonradiatively, which appears as a "premature wearout" problem. There are two common ways copper contamination can occur:

1. *Equipment errors in maintenance or operation.* Electron beam (E-beam) or sputter deposition machines can be evaporating or sputtering from outside the target or from the shutter (copper is a popular material to use for deposition systems due to its high thermal conductivity).
2. *Contamination of the source metal, especially gold.* Much of the gold on the market is purified ("reclaimed") from jewelry; popular 14-karat jewelry is 14–28% copper. Many suppliers will not guarantee copper levels below 1–2%, which is far too high for use in electronics.

Three precautions can protect you from copper contamination. First, qualify every wafer (or at least every wafer fab run) with a wearout reliability test. Second, use a graphite crucible to hold the gold (if using an e-beam machine) to minimize the intermixing between the source metal and the machine. Third, obtain the gold from virgin sources rather than reclaimed sources, and always obtain a certificate of compliance listing the measured levels of contaminants in the metal. Copper should be less than 10 ppm to prevent premature degradation.

*Failure Analysis*: Wearout reliability is greatly reduced in copper-contaminated lots. High reverse bias leakage is present with copper contamination. Copper exceeding 0.1% in any layer of the p-metal stack is suggestive of copper contamination being the cause of failure – this could be verified by SIMS. Baking the parts for several hours will often allow the contamination to migrate away from the active region and recover to the original forward L-I and reverse I-V characteristics. Then, operating parts again in forward bias under accelerated aging will allow the copper to return back to the active region and degrade again. Finally, the most conclusive method of proving copper contamination would be to cross-section degraded devices with a FIB and perform cross-sectional SEM or TEM imaging; see the characteristic copper streaks in the n-DBR layers immediately below the active region, and verify with EELS or EDX that the nodules or "streaks" are copper rich.

#### 6.6.2.2 Current Filamentation due to Excessive N-doping

Fifteen years ago, Hong Hou and Kent Choquette from Sandia asked if I'd be interested in studying some of the least reliable oxide VCSELs they'd seen [64]. Apparently the mass flow controller for the n-dopant had broken down and allowed far more n-doping than had been planned in the n-layers of the laser. At the time, we

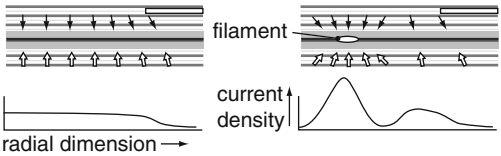



**Fig. 6.25** Cross-sectional schematics of current flow in VCSELs (*above*) and current density plots (*below*), for normal VCSEL (*left*) and filamentary VCSEL (*right*)

were not able to identify exactly why this was detrimental. However, years later when we observed reduced reliability again from increased n-doping, the apparent problem became clearer. Carrier mobility is higher in n-type material than in p-type material. If conductivity is too high, it appears that current filamentation is allowed and highly nonuniform current injection takes place, as shown in Fig. 6.25. The n-cladding layer between the active region and the n-DBR would ideally have high radial resistance and low vertical (longitudinal) resistance. Given that anisotropic conductivity is not easily obtained, compromises and tight control of the spacer layer resistance are important. This is likely to become more critical with the new higher-speed VCSELs under development (14–28 Gbps).

*Failure Analysis*: In one instance, this looked similar to a "dark-line defect" that was always in the same place but had no real dislocations that could be observed in cross-sectional TEM. One test is to run the laser at currents so high that lasing stops (perhaps 1.5× the peak power current). At these currents, good material will appear uniform in electroluminescence (EL) imaging, while highly nonuniform injection will be observed in material that is prone to undesirable filamentation. The bright areas during such aging were the areas that appeared as "dark lines" after aging. When performing EL after degradation, bright areas closely "hug" the "dark lines" with the brightest areas being just one or two micrometers away from the dark line; with devices having normal dark-line defects (DLDs), the bright areas are separated by several micrometers from the DLDs.

#### 6.6.2.3 Excessive Temperature and/or Drive Current

Premature degradation can occur when datasheet parameters are exceeded or even when transceivers are operated at the edge of specified datasheet parameters indefinitely. The failure signature is the same as is seen in normally degraded parts with accelerated aging. The two precautions that should be taken are customer education (make sure they understand that adequate cooling of the parts is needed) and designer education (give the transceiver designers a maximum current that they're allowed to use under normal operation and another maximum that they're allowed to hit under emergency high-temperature operation). Both of these should be sustainable for long periods without significant degradation. We believe future 25- and 28-Gbps VCSELs will test the limits of what is possible and make this failure mechanism more of an issue in the coming decades.

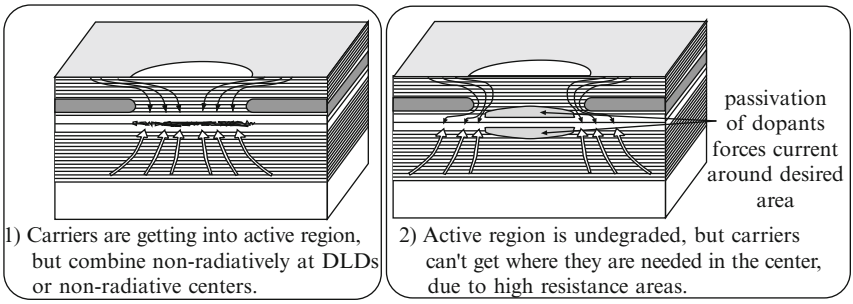



**Fig. 6.26** Two explanations of degradation. Option 1, shown at *left*, is the normal cause of degradation, where recombination centers reduce active region efficiency. Option 2, shown at *right*, is a better fit for the data observed on most VCSELs which have been through accelerated aging and show only normal wearout degradation. Option *2* shows little active region degradation

#### 6.6.2.4 Normal Degradation: The "Current Shunting Failure Mechanism"

The work that has been done on the primary VCSEL wearout has been somewhat inconclusive. However, evidence from a number of sources makes it appear that "normal degradation" in VCSELs is driven by hydrogen migration passivating dopant sites near the center of the p-DBR mirror above the active region [62, 65, 66]. This results in increased resistance in the center and current being shunted off to the edges or to one side or the other and not being uniformly distributed as originally intended. This is shown in Fig. 6.26. Evidence from cathodoluminescence and low-current L-I [62, 66] and low-current imaging with cooled cameras [67] shows no degradation of the active region at low currents where current crowding is less of an issue. However, the case for current shunting is not entirely rock solid, since the expected increases in series resistance are not observed. Further (difficult and expensive) research work would need to be done to make a more conclusive case, but to date, adequate wearout lifetimes were not a problem, so this work hasn't been done. It may be a source of greater interest once 14–28-Gbps VCSELs come further into the development cycle since these higher-speed devices require higher bias currents.

## 6.7 Failure Analysis of VCSELs

As mentioned earlier, our entire approach relies on a deep understanding of the physics of failure for the device. Without this, one is simply "guessing" what is wrong, and fixing the problem involves a lot of trial and error before the solution can be reached. We will discuss the failure analysis techniques recommended for routine use below, in order of importance, and how often they are used. For devices that failed on internal reliability tests, such devices were hopefully suspended shortly after reaching their failure criterion so that they are not "overdegraded"

and do not have extraneous defect structures added on by continued aging. For devices that are customer returns, device history is always useful, particularly the number of hours of service the part has seen before failure and the estimated deployment temperature.

As we discuss in more detail in the separate failure analysis chapter (Chap. 2 of this book), each group will have a failure analysis flowchart that will change over time, depending on what the most common failure mechanisms are at that time and how difficult each of the tests is to perform with the resources available to the group. But in general, every FA starts with an L-I-V scan of the laser and then goes to optical microscope photography and electroluminescence (EL) imaging. Where it goes from there depends on what you observe from these routine FA tests, what failures you have seen in the past, and how much experience you have with the laser design in question. Please see Chap. 2 for a more detailed discussion.

### *6.7.1 Electroluminescence (EL) and Spectrally Filtered Electroluminescence (SFEL)*

The most important technique in VCSEL failure analysis is electroluminescence (abbreviated "EL"). EL is simply taking a microscope photo of the VCSEL while a small current is run through the laser. This current is typically from 1/10th to 1/3rd of the threshold current (a fraction of a milliamp). The purpose of taking the images at this low current is to avoid monitoring mode formation. During lasing, the modes make the near field appear like a "pearl necklace," with bright spots in perhaps a quarter to half of the area, and the rest dark. With such an image, one couldn't tell whether "dark lines" are hiding in the dark area or whether the active region is undegraded. Thus, EL imaging at currents above lasing threshold is only rarely used in failure analysis and should always be preceded by subthreshold imaging.

At low currents, undegraded VCSELs should appear totally uniform across their entire aperture. Devices with normal degradation (perhaps from the "current shunting" failure mechanism) show mild and gradual darkening and require little increase in bias to take properly exposed images. By contrast, devices with climb dislocations show high-contrast darkening and require substantially more drive current to get an EL image. Figure 6.27 compares the two degraded devices just described above. We can quantitatively monitor the luminescence profile across the device and plot them to see the differences more clearly where distinguishing them might be difficult for an untrained eye. This is shown below in Fig. 6.28.

A substantial improvement in contrast can be obtained, particularly for gradually degraded devices, by using a bandpass filter. Common data communication VCSELs have an upper mirror (p-DBR) reflectivity of 99.8% at the lasing wavelength, meaning that the average photon will need to cycle around hundreds of times before it escapes. In the process, it can "walk" sideways before escaping, which has the effect of smearing out the image. By comparison, if a wavelength

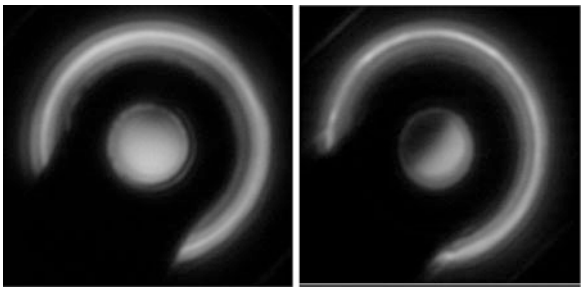


**Fig. 6.27** EL from wearout device on *left* shows only soft-edged degradation, while that with dark-line defects from climb dislocations on the *right* shows much more profound darkening

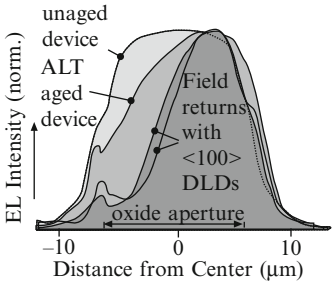



**Fig. 6.28** Traces of brightness versus radial distance show the expected symmetric pattern for undegraded devices (*light gray*), mild darkening to one side for normal wearout degradation (*medium gray*), and very severe darkening to one side for devices with climb dislocations (*dark gray*)

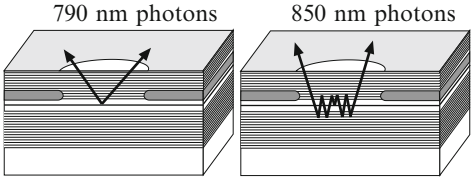



**Fig. 6.29** At 790 nm (*left*), upper mirror reflectivity is low, and photons escape almost immediately. At 850 nm, the average photon will be reflected many times before escaping, resulting in gain or loss, "ray confusion," and lower image contrast

outside the mirror "stop band" can be chosen, a mirror reflectivity below 40% can be obtained, meaning most photons escape on the first pass. This gives a much higher-contrast image. Drawings of what happens with and without the filters are shown in Fig. 6.29. Without filters, the low-contrast 850-nm image dominates. Sample images with and without the spectral filters are shown in Fig. 6.30.

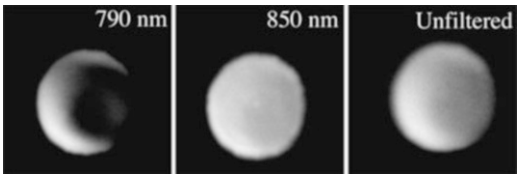



**Fig. 6.30** Images of a degraded VCSEL taken with a 790-nm bandpass filter (BPF) (*left*), an 850-nm BPF (*center*), and no filter (*right*). Note the great improvement in image contrast available using the correct spectral filter

It is desirable to have the electroluminescence camera hooked to a normal microscope, with an *x*-*y* translation stage, a turret objective, and filter slots for the spectral filters. The spectral filters are normally "10-nm bandpass filters," centered at 780 or 790 nm +/−5 nm with a triple cavity to give at >30-dB rejection of wavelengths outside the passband. A 1-in.- or 25-mm-diameter round filter is customarily used and can go into the same microscope slot used for the polarizer. While any sort of objectives can be used at low magnification power, we found that objectives that are optimized for near IR imaging are able to provide resolution nearly 2× better [68] (albeit at double the cost of standard visible objectives). The camera used would ideally have a cooled CCD for low noise and high bit depth and include user control over gain and integration time [69]. Realistically, if images are carefully adjusted, normal cheap silicon CCD cameras and image capture hardware are often adequate for this purpose.

### 6.7.2 *Light and Voltage Measurement as a Function of Drive Current (L-I-V Characterization)*

L-I sweep curves are the most common method of characterization, with every VCSEL being tested with such a sweep as part of at least a die-level test. For all but the cheapest devices, it will go through a header-level test after it's been packaged. A great deal can be discerned by how much threshold current has increased and how much peak power has dropped. Sample L-I curves are shown below in Fig. 6.31.

Voltage is also monitored as the same time as light out, and changes in device resistance during aging can often be an important clue to operation. In difficult cases, using a parameter analyzer with log power and current versus voltage can give useful information. In such cases, the parameter analyzer should be set up to sweep voltage with 0.05-V spacing between points. Max current should be limited to something past where peak power is obtained (which we'll call $I_{peak\ pwr}$) but usually should not be more than 1.5× $I_{peak\ pwr}$, as shown in Fig. 6.31.

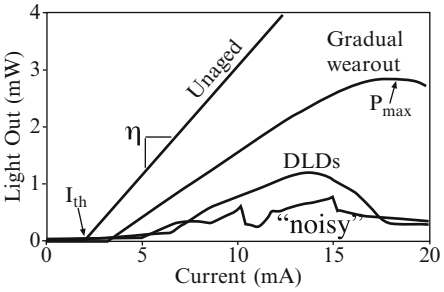



**Fig. 6.31** Light versus current (L-I) curves. Normal aging will result in some reduction in peak power and an increase in threshold current, although usually these are mild unless they've been aged a very long equivalent time. Climb dislocations or dark-line defects (DLDs) result in more severe changes fairly quickly. If a large portion of the lasing area has been degraded, a relatively small area and small number of filaments will be available, resulting in "kinks" in the L-I curve when changing from $n$ to $n + 1$ filaments – something it will do with perhaps every half milliamp increase in current

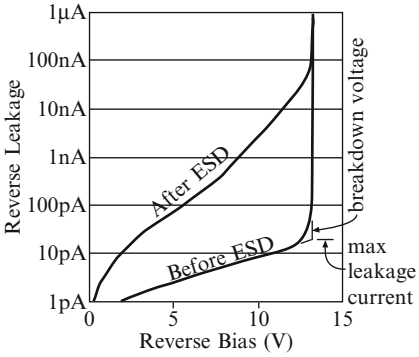



**Fig. 6.32** Reverse biased current versus voltage (I-V) curve, taken by a parameter analyzer, before and after damage by electrostatic discharge (ESD)

### *6.7.3 Reverse I-V*

Reverse I-V characterizes the reverse-leakage current of the laser. Lasers without defects will have very low leakage currents, thanks to their well-defined intrinsic region. Only picoamps will leak through for biases under a few volts, with a steady increase under increasing bias. By contrast, devices with cracks or climb dislocations will typically show leakage 2–3 orders of magnitude order of higher than undamaged parts and no reduction in avalanche breakdown voltage; parts with ESD damage will show 4+ order of magnitude increase in reverse leakage and a reduced voltage to the "breakdown threshold," if the breakdown voltage is defined as a current of 1- or 10-μA reverse leakage. Typical results are shown in Fig. 6.32.

Theoretically, reverse I-V can be scanned at the same time as the forward I-V curve is scanned. In practice, this is a bad idea since the "compliance current limits" should be set far lower – say 10 μA in reverse bias versus 30–50 mA in forward bias. Also, to get useful results comparable to those seen in Fig. 6.32, the correct type of dark-shielded test box with triaxial shielded cables is needed. Conventional test apparatuses without special wiring will produce nA of background noise and will not give the pA results shown here. As discussed earlier, it is wise to include 100 kΩ of series resistance to prevent thermal runaway of devices that have defects. Capacitance-voltage (C-V) profiling has been widely used to profile traps in degraded edge-emitter lasers, but little work to date has been done with degraded VCSELs [70].

### 6.7.4 Nomarski Microscopy

If cracks, epitaxial defects, or other similar problems are present, Nomarski microscopy will do a far better job of being able to see the defects than conventional bright-field optical microscopy. Nomarski microscopy is a form of phase contrast microscopy that allows relatively small height steps (tens of nanometers high) to be easily seen. The only minor disadvantage is that it gives unnatural colors. For final visual inspection of die and for epitaxial qualification, Nomarski microscopy should always be used, not the cheaper but less useful bright-field microscopy.

### 6.7.5 Scanning Electron Microscopy (SEM)

Scanning electron microscopy is, of course, an incredibly useful tool in general. But with VCSELs, the active region of interest is buried under a few micrometers of mirror material (p-DBR). Very little of what you'd like to know can be determined from an SEM scan of the top. In fact, many of the FA reports that I've seen that used SEM pictures were seriously flawed; they'd find a piece of dust, call it "contamination," and blame it for the failure – when in fact it was harmless and normal surface contamination. In general, I've seen misinterpretation of SEM images support many false diagnoses and almost never seen it solve real problems, so caution is recommended in prescribing SEM analysis or interpreting SEM images.

### 6.7.6 Thermally Induced Voltage Alteration (TIVA) and Other Backside Laser Scanning Microscope (LSM) Techniques

One of the most useful and cost-effective techniques for looking for climb dislocations is the use of "backside TIVA," which uses a laser scanning microscope, and its use is commonplace where the defect can be identified as being due to

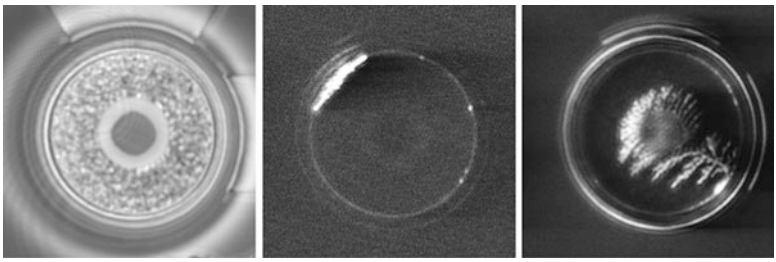


**Fig. 6.33** Images of mesa oxide VCSELs taken from the backside. Reflected light image (*left*) shows a mesa ~42 μm in diameter with a 12-μm oxide aperture. The *middle picture* shows a backside TIVA image of a die that operated normally; however, it shows some evidence of damage at the *upper left* which has not propaged. The TIVA image of the device shown on the *right* failed in reliability testing, due to a dark-line defect that propagated from the mesa edge in its *lower right* and then caused failure when it reached the emitting area in its *center*

climb dislocation networks from the EL imaging. This allows the extent of climb dislocations to be clearly seen; if climb dislocations come from outside the pumped area, the point of origin can be clearly seen. Examples of backside TIVA images can be seen in Fig. 6.33. Other examples have been published earlier [50, 51]. Climb dislocation networks show up as "bright" areas with normal amplifier mapping, while lasers without dislocations only show bright lines along the mesa edges.

The technique was first explored by the author and Daniel Barton in July 2000, examining four carefully chosen degraded VCSELs with different failure mechanism under a range of laser scanning wavelengths. We were looking from the top on samples with no special preparation. We found that wearout failure mechanisms such as gradual aging showed no contrast, while climb dislocations showed high contrast in TIVA. Unlike OBIC, which uses photon energies above the bandgap (e.g., 0.63 μm), TIVA uses subbandgap wavelengths – both the popular 1.06 and 1.34-μm wavelengths worked well. I recognized that in the search for an effective plan-view failure analysis technique, backside TIVA appeared to have great promise, so I obtained appropriate samples with DLDs and asked coworkers in our failure analysis lab to prepare them. The headers were decapped and embedded in epoxy, and the back of the header was polished off until the back of the die was visible. They were then polished and contact reestablished to the back of the p-side wirebond (easy) and to the n-side GaAs contact (difficult due to native oxide buildup). Dan Barton imaged the resulting sample in 2002 [71], and backside TIVA has spread across all the large VCSEL manufacturers in the years since then. Over the years, a large number of root causes for climb dislocations have been identified, including scriber damage, epi defects, cracking from excessive die mounting pressure, scratches, and, of course, the mirror oxide failure mechanism.

Backside TIVA is commonplace as a silicon IC failure localization technique [72]. Many of the precautions that apply are the same. To obtain high resolution, the sample must be thin (<60 μm); otherwise, optical aberrations and absorption will degrade

image quality. This is one of the main problems in sample preparation: the technicians are terrified of polishing too far and thus stop as soon as they see that the header has been polished away, leaving ~150 μm of substrate material on the back. To get successful results, careful polishing with a fine grit and frequent checking are required. I prefer to use an assembly microscope with the puck tilted to the side and compare the thickness-to-width ratio to give a feeling for what the remaining substrate thickness is. It is strongly recommended to try sample preparation first on samples that are *not* very important and verify that you can repeatably control polishing, die thickness, and reestablishing electrical contact before starting to prepare die that one *does* care about.

The laser scanning microscope (LSM) is usually one that is intended for IC failure analysis work; many microscope owners are also willing to provide backside TIVA imaging as a service [73]. Although theoretically, such wavelengths should allow resolution of only 0.6–0.8 μm, but as long as the sample has been thinned to <50 μm of remaining substrate thickness and the instrument threshold settings are properly adjusted, resolution of 0.2 μm has been obtained.

One of the beauties of the backside TIVA technique is that once it has been finished, follow-up work can be done in TEM using FIB liftout on the samples.

### 6.7.7 Transmission Electron Microscopy (TEM)

Transmission electron microscopy (TEM) is the most useful tool for failure analysis of VCSELs and would be used thousands of times a year for failure analysis, if it weren't for its high cost and long painful sample preparation and imaging times. If it had been available during my Ph.D. work, it would have substantially sped up the early understanding of VCSEL degradation. However, it was only the advent of the focused ion beam (FIB) sample preparation of VCSELs in 2000 by Mathes [39], and especially sequential FIB sample prep of plan view followed by cross-sectional liftout in 2002 by Stark and Professor Phillip Russell at NCSU [50], that really allowed the full range of field failure analysis and research work to be performed.

Climb dislocations, the dislocations that nucleate them, and the thermal damage from ESD are all clearly visible. Cumulatively, these cover the majority of field failures and, if properly done, are conclusive as to the root cause of the failures, though not always the fix to prevent a recurrence of the failures.

The FIB sample preparation is shown below in Fig. 6.34. One important point here is that if we know that the damage is in the active region (almost always the case), then plan-view TEM is far more useful than cross-sectional TEM. The FIB sample preparation involves removing all but about 1 μm of material, at which point the sample is thin enough to "see through" with the TEM. For oxide VCSELs, it is therefore traditional to remove most or all of the n-DBR below the active region and most of the p-DBR above the oxide layer. The oxide layer and active layer are left untouched. The device is then imaged, and if desired, a cross section can be prepared based on the results. If a cross section were prepared without a plan-view "failure localization" technique first being used (either backside TIVA or plan-view

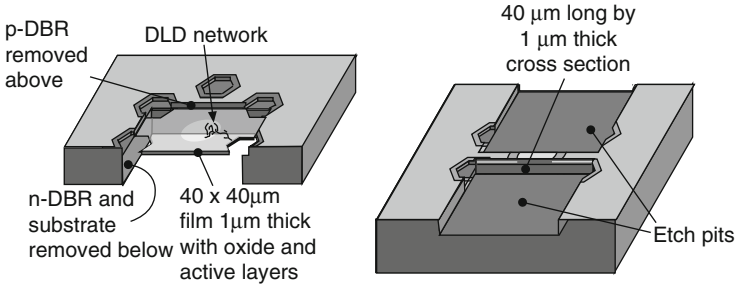



**Fig. 6.34** Images of how plan-view (*left*) and cross-sectional (*right*) TEM samples are prepared by FIB




**Fig. 6.35** A plan-view TEM image (*left*) shows the character of the climb dislocations looking from the *top*, as they extend laterally. A cross-sectional TEM image (*right*) clearly shows the three quantum wells from the side as well as climb dislocations primarily in the top quantum well. The oxide layer can also be seen in the *upper right* of the image

TEM), a randomly chosen cross section would involve the FIB destroying approximately 98% of the area. In all likelihood, this would destroy the area of interest. Therefore, for VCSELs, it is recommended to prepare and examine plan view TEM images before considering taking a cross sectional TEM sample from that device. It is rarely useful to have a cross sectional TEM without the corresponding plan view TEM image. Examples of both plan-view and cross-sectional TEM images are shown in Fig. 6.35.

A number of limitations are present for this technique. It is expensive (usually US $3,000–3,500 for plan-view preparation and imaging as of this writing and sometimes more if a cross section is also desired). Only a few firms are capable of doing this with high quality. There is often a problem with "bending fringes" creating lines in images that need to be distinguished from actual dislocations (multiple tilts must be examined). In response, one firm has gone to using a scanning TEM with a high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) which has minimal sensitivity to sample bending. TEM often shows nothing if a gradually degraded VCSEL is examined; it only shows coherent crystal defect structures such as climb dislocations. Higher concentrations of point defects, which can cause significant degradation, require other methods of detection.

### 6.7.8 Reverse Bias Emission Microscopy (EMMI)

EMMI is used to look for mechanical damage to LEDs [74] and silicon ICs [75]. It does so by applying a reverse bias and imaging the very small amount of photoemission with a photomultiplier or a cooled CCD camera with a long integrating time. It has been used to study ESD-damaged VCSELs [59] as well as VCSELs with climb dislocations from 85/85 aging [76]. It has not been a core FA technique accessible by most manufacturers, but this may be partly due to lack of access to the equipment in-house. However, cameras with cooled CCDs that allow long exposure times and are capable of taking images with just a few microamps of reverse current over an integration time of several seconds can provide good results [69]. It may have usefulness in screening out damaged devices as a complement to device burn-in, if a high-speed EMMI test set could be developed.

### 6.7.9 Electron-Beam-Induced Current (EBIC) and OBIC

One failure analysis tool that is extremely useful in most "in-plane" semiconductor lasers is EBIC. This allows inexpensive plan-view imaging of the entire die image [18]. Unfortunately, typical VCSELs have p-DBRs that are 3–4 μm thick, while typical EBIC SEMs have a penetration depth of only 1–2 μm for most electrons. Therefore, the only useful plan-view EBIC images that have been obtained in VCSELs were by Jiang et al. and involved the use of rare scanning transmission electron microscopes (STEMs). His best results were obtained using an especially rare 400-keV machine at Lucent/Bell Laboratories [77]. Our attempts to use a 200-keV STEM system at Agilent labs showed very low resolution [78]. The backside TIVA technique is superior in every way. The problems involved in plan-view EBIC imaging of VCSELs have been covered elsewhere in greater detail [79, 80, 81].

Cross-sectional (as opposed to plan-view) EBIC can be somewhat useful for research purposes if a significant number of dies are available. As mentioned before in the TEM chapter, cross-sectional analysis should *never* be done on any die where true root cause is needed, until comprehensive plan-view imaging has been collected and shown the region of interest. But if a FIB is used to cut away half the VCSEL while it's still in a header or a mount allowing probing, the die can then be tilted and imaged by EBIC. This allows us to see whether the active region has actually degraded or not and separate out various failure mechanisms [62, 82].

To the best of our knowledge, plan-view OBIC on VCSELs has not been successfully performed. It would require a collimated source with a carefully chosen wavelength, ideally at a mirror transmission wavelength peak, but not at too high a photon energy to avoid upper mirror (p-DBR) absorption. Cross-sectional OBIC would give data similar to that described above for cross-sectional EBIC.

## 6.8 Future VCSEL Reliability Challenges

Future reliability research in VCSELs has a number of opportunities for development. One is the development of 14–40-Gbps VCSELs, where very high current densities are used to obtain fast rise and fall times [6]. These devices will have many wearout lifetime challenges, and epi recipe optimization as well as better thermal management will be key. In addition, customer education will continue to become more important. Customers (and the sales and marketing teams that sell to them) would like to simplify the VCSEL reliability to a single phrase: "Whatever you do to the transceivers or VCSELs, they're so tough you can't hurt them." When there were lots of design margin with 1-Gbps VCSELs, it was close to being true. But in the future, customers will need a datasheet that shows what lifetime they can expect as a function of how well they cool the product. An alternate approach that might have lower current densities required would involve a more complex device with an integrated electro-optic modulator; the laser is simply left on at a constant DC level and thus does not need such high currents to get adequate rise and fall times[83, 84]. Another area of research may be nonhermetic low-cost products. Historically, oxide VCSELs without hermetic packaging had an issue with premature failure unless additional features were added to the die [32, 85]. Higher temperatures in many data center applications and high-density boxes will add to the applications operating at 70–80°C, rather than the historical 50°C case temperature, and create additional demands for longer wearout lifetime in VCSELs. Wafer-level burn-in will continue to become more common as cost reduction pressures continue.

Another opportunity for development will be in AlGaInAs 850-nm strained QW VCSELs, following the lead of what's been performed for 808-nm in-plane lasers [17, 19]. These compressively strained lasers are able to totally resist climb dislocation growth and could eliminate vulnerability to all of the maverick failure mechanisms listed earlier and possibly even make burn-in unnecessary. Such VCSELs were demonstrated ~15 years ago – more for their performance benefits than for their reliability benefits [20]. The reason such devices are not commercially offered today is the relatively narrow window of growth conditions for the quaternary alloy, relative to the ease of growth of the GaAs QW lasers.

VCSELs at other wavelengths may become available. The additional challenges of thermal conductivity, electrical conductivity, and mirror reflectivity make many other material systems not well suited to use for VCSELs. Many areas have proven disappointing, with all manufacturers having backed away from 1,310-nm VCSELs, in spite of hundreds of people having researched them for two decades and some successes [86]. Likewise, extensive work has been done on blue VCSELs using GaN. Recently CW lasing at room temperature has been demonstrated, but they have not yet been commercialized [87]. Red VCSELs have many useful applications, including DVD readers and plastic fiber-optic links. But for a long time, they had very low lifetimes. However, they have finally reached small-scale commercialization a few years ago [88], and positive results continue

to be reported [89]. Quantum dot VCSELs may open up new wavelengths as well [90]. Progress has been reported in high power applications, both using external cavities (abbreviated VECSELs), as well as in 2-D arrays of smaller VCSELs [91].

Enhanced wafer qualification techniques are another opportunity for improvement. Normal device fabrication and aging give a minimum 2-week delay between when devices are grown and when useful feedback about their reliability is obtained. However, a number of techniques exist for measuring the concentration of point defects [92–94] and thereby giving same-day feedback to the epi growers.

Finally, higher-speed screening techniques to replace burn-in would be valuable, and might be possible, using noise screening [93] or reverse bias photoemission to identify lasers with defects.

## 6.9 Where to Go for Further Information

The best specific information about VCSELs and VCSEL reliability is published in the annual SPIE Photonics West conference proceedings; we have cited a small number of the relevant papers here.

The best general information on semiconductor laser reliability has been published in the Artech books by Fukuda [95] and Ueda [96]. The reliability books by Franklin Nash [41] and Finn Jensen [44] are also very relevant. For information on VCSELs, the book of that title is the best resource [9]; the review papers by Iga [10], Choquette and Hou [11], or Larsson [6] are also good general references on VCSELs.

**Acknowledgments** I would like to acknowledge my coworkers at UCSB, HP/Agilent (now Avago), Emcore, Honeywell/AOCD/Finisar, and JDSU for their contributions, as well as collaborators from numerous other universities, national labs, and analytical companies. I also owe particular thanks to the late Dr. Robert Waters, who conveyed the excitement in the breakthroughs that were taking place two decades ago in laser reliability and was my inspiration in pursuing reliability studies.

## References

1. K. Iga, S. Ishikawa, S. Ohkouchi, T. Nishimura, Room-temperature pulsed oscillation of GaAlAs/GaAs surface emitting injection laser. Appl. Phys. Lett. **45**(4), 348–350 (1984)
2. J.L. Jewell, A. Scherer, S.L. McCall, Y.H. Lee, S. Walker, J.P. Harbison, L.T. Florez, Low-threshold electrically pumped vertical-cavity surface-emitting microlasers. Electron. Lett. **25** (17), 1123–1124 (1989)
3. R.S. Geels, S.W. Corzine, J.W. Scott, D.B. Young, L.A. Coldren, Low threshold planarized vertical-cavity surface-emitting lasers. IEEE Photon. Technol. Lett. **2**(4), 234–236 (1990)
4. K. Wirbel, Components, lasers in for long, short haul, *E.E. Times*, 4 Mar 1996
5. Private conversation with Kent Choquette and Martin Grabherr, Jan 2012

6. A. Larsson, Advances in VCSELs for communication and sensing. IEEE J. Sel. Top. Quantum Electron. **17**(6), 1552–1567 (2011)
7. D.B. Young, J.W. Scott, F.H. Peters, M.G. Peters, M.L. Majewski, B.J. Thibeault, S.W. Corzine, L.A. Coldren, Enhanced performance of offset-gain high-barrier vertical-cavity surface-emitting lasers. IEEE J. Quantum Electron. **29**(6), 2013–2022 (1993)
8. W.J. Fritz, L.B. Bauer, C.S. Miller, Analysis of aluminum gallium arsenide laser diodes failing due to nonradiative regions behind the facets, in *Proceedings of 27th International Reliability Physics Symposium*, vol. 27, 1989, pp. 59–64
9. C.W. Wilmsen, H. Temkin, L.A. Coldren, *Vertical-Cavity Surface-Emitting Lasers: Design, Fabrication, Characterization, and Applications* (Cambridge University Press, Cambridge, 2001)
10. K. Iga, Surface-emitting laser-its birth and generation of new optoelectronics field. IEEE J. Sel. Top. Quantum Electron. **6**(6), 1201–1215 (2000)
11. K.D. Choquette, H.Q. Hou, Vertical-cavity surface emitting lasers: moving from research to manufacturing. Proc. IEEE **85**(11), 1730–1739 (1997)
12. R. Morgan, E. Strzelecki, Spatially modulated reflector for an optoelectronic device, U.S. Patent 6,727,520, 27 Apr 2004
13. J.W. Scott, D.B. Young, B.J. Thibeault, M.G. Peters, L.A. Coldren, Design of index-guided vertical-cavity lasers for low temperature-sensitivity, sub-milliamp thresholds, and single-mode operation. IEEE J. Sel. Top. Quantum Electron. **1**(2), 638–648 (1995)
14. J.D. Dow, R.E. Allen, Role of dangling bonds and antisite defects in rapid and gradual III-V laser degradation. Appl. Phys. Lett. **41**(7), 672–674 (1982)
15. S. Yamakoshi, M. Abe, O. Wada, S. Komiya, T. Sakurai, Reliability of high radiance InGaAsP/InP LEDs operating in the 1.2-1.3 μm wavelength. IEEE J. Quantum Electron. **17**(2), 167–173 (1981)
16. R.G. Waters, S.L. Yellen, N.F. Ruggieri, D.P. Bour, Inhibited dark-line defect formation in strained InGaAs/AlGaAs quantum well lasers. IEEE Photon. Technol. Lett. **2**(8), 531–533 (1990)
17. S.L. Yellen, A.H. Shepard, C.M. Harding, J.A. Baumann, R.G. Waters, D.Z. Garbuzov, V. Pjataev, V. Kochergin, P.S. Zory, Dark-line-resistant, aluminum-free diode laser at 0.8 um. IEEE Photon. Technol. Lett. **4**(12), 1328–1330 (1992)
18. R.G. Waters, Diode laser degradation mechanisms: a review. Prog. Quantum Electron. **15**(3), 153–174 (1991)
19. J.A. Baumann, Strained quaternary InAlGaAs 810-nm lasers, in *Proceedings of SPIE*, Los Angeles, CA, vol. 1850, 1993, pp. 203–214
20. J. Ko, E.R. Hegblom, Y. Akulova, B.J. Thibeault, L.A. Coldren, Low-threshold 840-nm laterally oxidized vertical-cavity lasers using AlInGaAs-AlGaAs strained active layers. IEEE Photon. Technol. Lett. **9**(7), 863–865 (1997)
21. R.W. Herrick, PhD Dissertation, ibid, Sections 8.1 and 8.2
22. F. Jensen, N.E. Petersen, *Burn-in* (Wiley, New York, 1982)
23. R.W. Herrick, Failure analysis and reliability of optoelectronic devices, in *Microelectronics Failure Analysis Desk Reference*, 5th edn. (ASM International, Materials Park, Ohio, 2004), pp. 230–254
24. C. Xie, N. Li, C. Lei, X. Sun, W. Luo, L. Zhao, C. Helms, C. Liu, D. Jessen, R. Carson, New high speed VCSEL product development at Emcore, in *Proceedings of SPIE*, San Francisco, CA, vol. 7952, 2011, pp. 7952–7956
25. R. Herrick, C. Roxlo, O. Sjolund, T. Sudo, High-power optical burn-In, U.S. Patent 7,795,896, 14 Sept 2010
26. K. Maeda, M. Sato, A. Kubo, S. Takeuchi, Quantitative measurements of recombination enhanced dislocation glide in gallium arsenide. J. Appl. Phys. **54**(1), 161 (1983)
27. *Telcordia GR-468, issue 2* (Telcordia Technologies, Piscataway, 2004)
28. R.W. Herrick, Oxide VCSEL reliability qualification at Agilent Technologies, in *Proceedings of SPIE*, San Jose, CA, vol. 4649, 2002, pp. 130–141

29. R.A. Hawthorne III, J.K. Guenter, D.N. Granville, M.K. Hibbs-Brenner, R.A. Morgan, Reliability study of 850 nm VCSELs for data communications, in *Proceedings of International Reliability Physics Symposium*, vol. 34, 1996, pp. 203–210
30. L. Li, J. Xue, Reliability of microelectronics packaging in the era of energy wise and borderless networks, in *Proceedings of International Reliability Physics Symposium*, Anaheim, CA, 2010, pp. 440–445
31. R.W. Herrick, Reliability of fiber optic datacom modules at Agilent Technologies, in *Proceedings of Electronic Components and Technology Conference (ECTC)*, San Diego, CA, USA, vol. 52, 2002, pp. 532–539
32. S. Xie, R.W. Herrick, D. Chamberlin, S.J. Rosner, S. McHugo, G. Girolami, M. Mayonte, S. Kim, W. Widjaja, Failure mode analysis of oxide VCSELs in high humidity and high temperature. J. Lightwave Technol. **21**(4), 1013 (2003)
33. H. Deng, J.J. Dudley, S.F. Lim, C. Lei, B. Liang, M. Tashima, L.A. Hodge, X. Zhang, J. Herniman, R.W. Herrick, 850-nm oxide VCSEL development at Hewlett-Packard, in *Proceedings of SPIE,* San Jose, CA, vol. 3627, 1999, pp. 23–28
34. Weibull++, a software package from Reliasoft Corporation, Tucson, Arizona, USA
35. D. Hodgson, B. Olsen, Protecting your laser diode (Application Note AN-3, rev. 01). ILX lightwave website (2001), www.ilxlightwave.com
36. J.A. Tatum, Commercialization of Honeywell's VCSEL technology, in *Proceedings of SPIE*, San Jose, CA, vol. 3946, 2000, pp. 2–13
37. Y.M. Cheng, R.W. Herrick, P.M. Petroff, M.K. Hibbs-Brenner, R.A. Morgan, Degradation mechanisms of vertical cavity surface emitting lasers, in *International Reliability Physics Symposium*, vol. 34, 1996, pp. 211–213
38. R.W. Herrick PhD Dissertation, op cit., Section 4.3 on TEM
39. D. Mathes, Materials issues for VCSEL operation and reliability, PhD Dissertation, University of Virginia, Charlottesville, VA, 2002
40. R.W. Herrick, unpublished work. We have looked a dozens of TEM cross sections of field returns and failures from accelerated aging, without seeing the sorts of TEM features published at UCSB from very high currents and very deep aging
41. F.R. Nash, *Estimating Device Reliability: Assessment of Credibility* (Springer, New York, 1993)
42. F. Nash, *Notes for Short Course 108 at the 1999 Optical Fiber Conference*. OFC, 1999
43. Overview of the crisis: firestone tire recall-ford explorer (Article at www.firestone-tire-recall.com) [Online]. Available: http://www.firestone-tire-recall.com/pages/overview.html. Accessed 29 Jan 2011
44. F. Jensen, *Electronic Component Reliability: Fundamentals, Modelling, Evaluation, and Assurance*, 1st edn. (Wiley, New York, 1996)
46. ALTA, a software package from Reliasoft Corporation, Tucson, AZ
47. R.G. Waters, R.K. Bertaska, Degradation phenomenology in (Al) GaAs quantum well lasers. Appl. Phys. Lett. **52**(3), 179–181 (1988)
48. F. Siegelin, C. Brillert, Failure analysis of vertical cavity surface emission laser diodes, in *ISTFA 2003: Proceedings of the 29th International Symposium for Testing and Failure Analysis*, Santa Clara Convention Center, Santa Clara, CA, 2–6 Nov 2003, p. 426
49. C. Helms, I. Aeby, W. Luo, R.W. Herrick, A. Yuen, Reliability of oxide VCSELs at emcore, in *Proceedings of SPIE,* vol. 5364, 2004, pp. 183–189
50. T.J. Stark, P.E. Russell, C. Nevers, 3-D defect characterization using plan view and cross-sectional TEM/STEM analysis, in *Proceedings of ISTFA*, vol. 31, 2005, p. 344
51. C. Lei, N. Li, C. Xie, R. Carson, X. Sun, W. Luo, L. Zhao, C. Helms, D. Jensen, C. Liu, Emcore VCSEL failure mechanism and resolution, in *Proceedings of SPIE*, San Francisco, CA, vol. 7615, 2010, p. 761504
52. G.T. Dangelmayer, *Esd Program Management: A Realistic Approach to Continuous Measurable Improvement in Static Control*, 2nd edn. (Springer, New York, 1999)
53. ESD Symposium, an annual event held every September by the ESD Association. See www.esda.org for further information.

54. See for example www.prostat.com/training for programs with Stephen Halperin and Ron Gibson [Online]. Available: http://www.prostatcorp.com/training/. Accessed 3 Feb 2011
55. See for example ESD auditing kits – Prostat Corporation, www.prostat.com/kits.asp [Online]. Available: http://www.prostatcorp.com/Kits.asp. Accessed 31 Jan 2011
56. H. Snyder, A. Wallash, A study of methods to eliminate metal contact in GMR head manufacturing, in *Proceedings of the ESD Symposium*, 2000, pp. 190–192
57. See articles found at www.wallash.com [Online]. Available: http://www.wallash.com/. Accessed 14 Feb 2011
58. H. Meier, R. Santschi, S. Odermatt, B. Witzigmann, S. Eitel, F. Nallet, G. Letay, A TCAD approach to robust ESD design in oxide-confined VCSELs, in *Proceedings of SPIE,* San Jose, CA, vol. 6484, 2007, p. 648405
59. J.J. Krueger, Studies of ESD-related failure patterns of Agilent oxide VCSELs, in *Proceedings of SPIE,* San Jose, CA, vol. 4994, 2003, pp. 162–172
60. S.A. McHugo, A. Krishnan, J.J. Krueger, Y. Luo, N. Tan, T. Osentowski, S. Xie, M.S. Mayonte, R.W. Herrick, Q. Deng, and others, Characterization of failure mechanisms for oxide VCSELs, in *Proceedings of SPIE,* vol. 4994, 2003, pp. 55–66
61. D. Mathes, J. Guenter, B. Hawkins, B. Hawthorne, C. Johnson, An atlas of ESD failure signatures in vertical cavity surface emitting lasers, in *Proceeding of ISTFA*, vol. 31, 2005, p. 336
62. R. Herrick, Degradation in vertical cavity lasers, PhD Dissertation, Department of ECE (published with University Microfilms, or available from the author), UCSB, 1997
63. J. Guenter, D. Mathes, B. Hawkins, J. Tatum, Developments at Finisar AOC [6908-04], in *Proceedings-SPIE the International Society for Optical Engineering*, vol. 6908, 2008, p. 6908
64. R.W. Herrick, PhD Dissertation, ibid, see section 8.3
65. R. Biard, J. Guenter, Analysis of degraded VCSEL characteristics (an internal company memo), 1995
66. R.W. Herrick, P.M. Petroff, Gradual degradation in 850-nm vertical-cavity surface-emitting lasers. IEEE J. Quantum Electron. **34**(10), 1963–1969 (1998)
67. R. Herrick, D. Barton, L. Wong, Electroluminescence study at Sandia National Labs using LN2 cooled CCD camera. Agilent Internal Memo, 2000
68. R.W. Herrick, PhD Dissertation, op. cit., Appendix G, section G.1.2.2
69. Aqua Ex-i Camera by Q-Imaging of British Columbia, Canada
70. R.W. Herrick, PhD Dissertation, University Microfilms, Santa Barbara, CA, Appendix H, 1997
71. R. Herrick, D. Barton, Y. Luo, M. Mayonte, Results from first backside TIVA images. Agilent Internal Memo, 31 May 2002
72. E. Cole, Beam-based defect localization methods, in *Microelectronics Failure Analysis: Desk Reference,* 5th ed. (ASM International, Materials Park, Ohio, 2004), pp. 406–416
73. The most popular place for scanning optical microscopy among VCSEL firms has been Semicaps in San Jose California or Singapore: www.semicaps.com/som.htm [Online]. Available: http://www.semicaps.com/som.htm. Accessed 14 Feb 2011
74. R.S. Mann, D.K. McElfresh, Categorizing light output degradation failures in LEDs using the relationship between defect revealing mechanisms responsible for EL, CL, EBIC, and reverse bias photoemission (RP), in *Proceedings IEEE International Reliability Physics Symposium*, Las Vegas, NV, 1995, pp. 177–186
75. C. Boit, Fundamentals of photon emission (PEM) in silicon – electroluminescence for analysis of electronic circuit and device functionality, in *Microelectronics Failure Analysis: Desk Reference*, 5th edn. (ASM International, Material Park, Ohio, 2004), pp. 356–368
76. D.K. McElfresh, L.D. Lopez, D. Vacar, Reverse-bias emission sheds light on the failure mechanism of degraded vertical-cavity surface-emitting lasers. J. Appl. Phys. **99**(12), 123113 (2006)
77. Private conversation with Wenbin Jiang, about work performed for Motorola

78. L. Wong, R.W. Herrick, G. Girolami, HVEBIC test (powerpoint slide set). Agilent Internal Communication, 25 May 2000
79. R.W. Herrick, ASM Microelectronics FA Desk Reference, ibid, 2004; see section 5.3.2
80. R.W. Herrick, PhD Dissertation, ibid, Appendix E
81. B. Bastek, P. Veit, J. Christen, S. Petzold, F. Bertram, C. Berger, A. Dadgar, A.J. Krost, Nano-scale correlation of structural and optical properties of DBRs using STEM CL, in *Proceedings of SPIE*, 2011, vol. 7939, pp. 7939–7946
82. R.W. Herrick, P.M. Petroff, Improved reliability of red GaInP vertical-cavity surface-emitting lasers using bias-induced annealing. Appl. Phys. Lett. **72**(15), 1799–1801 (1998)
83. A. Paraskevopoulos, High-bandwidth VCSEL devices, in *Proceedings of LEOS*, vol. 19th annual conference, 2006, pp. 400–401
84. N.N. Ledentsov, F. Hopfer, D. Bimberg, High-speed quantum-dot vertical-cavity surface-emitting lasers. Proc. IEEE **95**(9), 1741–1756 (2007)
85. S.X. Xie, R.W. Herrick, G.N. De Brabander, W.H. Widjaja, U. Koelle, A.N. Cheng, L.M. Giovane, F.Z. Hu, M.R. Keever, T. Osentowski, and others, Reliability and failure mechanisms of oxide VCSELs in non-hermetic environments, in *Proceedings of SPIE,* vol. 4994, 2003, pp. 173–180
86. L.A. Graham, M. Schnoes, K.D. Maranowski, T.R. Fanning, M.V. Crom, S.A. Feld, M.H. Gray, K. Bowers, S.L. Silva, K. Cook, and others, New developments in 850 and 1300 nm VCSELs at JDSU, in *Proceedings of SPIE*, vol. 7229, 2009, p. 72290B
87. D. Kasahara, D. Morita, T. Kosugi, K. Nakagawa, J. Kawamata, Y. Higuchi, H. Matsumura, T. Mukai, Demonstration of blue and green GaN-based vertical-cavity surface-emitting lasers by current injection at room temperature. Appl. Phys. Express **4**(7), 072103 (2011)
88. G. Duggan, D.A. Barrow, T. Calvert, M. Maute, V. Hung, B. McGarvey, J.D. Lambkin, T. Wipiejewski, Red vertical cavity surface emitting lasers (VCSELs) for consumer applications, in *Proceedings of SPIE*, vol. 6908, 2008, p. 69080G–1
89. K. Johnson, M. Hibbs-Brenner, W. Hogan, M. Dummer, K. Dogubo, G. Berg, Record high temperature high output power red VCSELs, presented at the *Vertical-Cavity Surface-Emitting Lasers XV*, San Francisco, CA, 2011, p. 795208–795208–12
90. N.N. Ledentsov, D. Bimberg, Z.I. Alferov, Progress in epitaxial growth and performance of quantum dot and quantum wire lasers. J Lightwave Technol. **26**(11), 1540–1555 (2008)
91. J.F. Seurin, G. Xu, V. Khalfin, A. Miglo, J.D. Wynn, P. Pradhan, C.L. Ghosh, L.A. D'Asaro, Progress in high-power high-efficiency VCSEL arrays, in *Proceedings of SPIE*, vol. 7229, p. 722903–1
92. J.W. Tomm, A. Bärwolff, A. Jaeger, T. Elsaesser, J. Bollmann, W.T. Masselink, A. Gerhardt, J. Donecker, Deep level spectroscopy of high-power laser diode arrays. J. Appl. Phys. **84**, 1325 (1998)
93. X. Chen, J.A. Johansen, C. Liu, Temperature dependence of low-frequency electrical noise and reliability of semiconductor lasers (Proceedings Paper), in *Proceedings of SPIE*, vol. 4602, 2001, pp. 128–133
94. For example, extreme high current aging, followed by C-V or log I-V profiling to look for the presence of point defects, or the migration of point defects
95. M. Fukuda, *Reliability and Degradation of Semiconductor Lasers and LEDs* (Artech House, Boston, 1991)
96. O. Ueda, *Reliability and Degradation of III-V Optical Devices* (Artech House, Boston, 1996)

# Chapter 7
# Structural Defects in GaN-Based Materials and Their Relation to GaN-Based Laser Diodes

**Shigetaka Tomiya**

**Abstract** Reduction of structural defects in GaN-based laser diodes is critically important for high-efficiency and highly reliable performance. Therefore, it is very important to understand their nature. Since their crystal structure is different from that of conventional zinc-blende-based III-V materials such as GaAs, different kinds of structural defects are nucleated in the epitaxial films for GaN-based laser diodes. Here we describe on transmission electron microscope analysis of various kinds of structural defects observed in these epitaxial films. Since these structural defects are frequently observed in GaN-based laser diodes, we also discuss their relation to GaN laser diodes in terms of device reliability and degradation.

## 7.1 Introduction

Considerable efforts have been expended on the development of GaN-based laser diodes (LDs) since the first demonstration of a continuous-wave (CW) laser in 1996 [1], since GaN-based LDs are tremendously desirable for many optoelectronic applications such as optical data storage, color displays, photochemical processing, and environmental monitoring [2]. Near-UV emission GaN-based LDs have already been realized and are now mass-market products with applications in "Blu-ray" for optical data storage. Blue-emission LDs are also appearing in applications such as laser projection display systems. In addition, there have recently been increased requirement to improve the emission efficiencies of green light-emitting diodes (LED) and to realize green LDs.

No matter how perfect semiconductors are, defects still exist in varying degrees. Reducing the defects is the most important task in semiconductors in their early

S. Tomiya (✉)
Advanced Materials Laboratories, Sony Corporation, 4-14-1 Asahi, Atsugi, Kanagawa 243-0021, Japan
e-mail: shigetaka.tomiya@jp.sony.com

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_7,
© Springer Science+Business Media New York 2013

stage of development. GaN-based semiconductors are obviously no exception and furthermore, a variety of defects is generated as compared to conventional compound semiconductors. Since GaN-based semiconductors have hexagonal crystal structure and their alloys are heavily lattice mismatched systems, these materials systems are quite different from conventional compound semiconducting materials whose crystal structure is zinc-blende structure such as GaAs-based and InP-based materials. In addition, mainly due to the nonavailability of native GaN substrates in the early stage of the development of LDs, GaN epitaxial layers have been grown on heavily lattice mismatched substrate such as sapphire and SiC. (Sapphire still remains the dominant substrate for LED mainly due to the commercial sense.) These substrates are heavily lattice mismatched to the GaN epitaxial layers, and thus, structural defects are inevitable in the layers.

In semiconductor optical devices, in order to improve the device characteristics and reliability, it is obviously natural to reduce density of defects that usually act as nonradiative recombination centers and/or lead to cause of high resistivity. Therefore, it should be required to analyze defects in detail and to understand their generation mechanism. In this chapter, we describe various kinds of structural defects observed in GaN-based epitaxial films and discuss their relation to reliability of GaN-based LDs.

## 7.2 Threading Dislocations: Defects Generated at Epi/Subinterface

As mentioned above, sapphire substrates are mainly used for epitaxial growth in the early stage of development of the GaN-based LDs. Due to the large lattice mismatch and the difference between the thermal-expansion coefficients of GaN-based materials and the sapphire substrates, a high density of dislocations (typically $10^8$–$10^{10}$/cm$^2$) was inevitable. Despite such a high density of threading dislocations, emission efficiency of GaN-based LEDs was much better than that of the conventional zinc-blende-based LEDs and had brighter a relatively long lifetime [3].

However, in the early stage of development, it was rather difficult to fabricate reliable LDs, due to their high threshold current density, high operating voltage, and high density of dislocations. Therefore, a significant reduction in power consumption ($I_{op} \times V_{op}$) is important for extending the lifetime [4]. More importantly, reducing the dislocation density is also very effective [5]. A reduction in the dislocation density down to the order of $10^6$/cm$^2$ by using the epitaxial lateral overgrowth (ELO) technique has extended the lifetimes of such LDs to over 15,000 h under 30 mW CW operation at 60°C [4, 6]. It is known that the rapid degradation observed in conventional GaAs- and InP-based LDs can be attributed to dislocation multiplication by nonradiative recombination at threading dislocations [7]. Therefore, the dislocation density should be minimized to $10^3$–$10^4$/cm$^2$, which means that there are essentially no dislocations

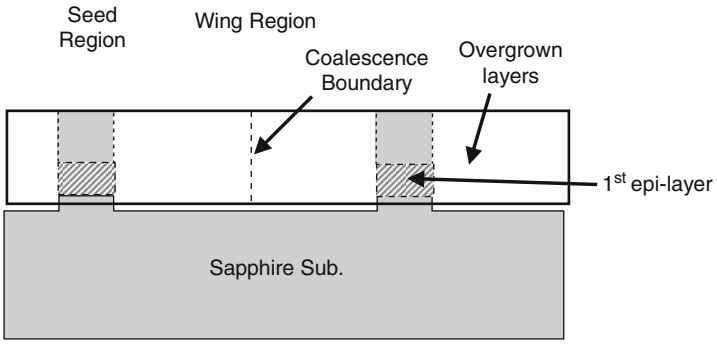



**Fig. 7.1** Schematic illustration of ELO GaN layers

in the stripe region. However, in GaN-based LDs fabricated using the ELO technique, something on the order of tens of threading dislocations exists in the stripe region. Therefore, even if dislocations act as nonradiative centers in GaN, the dislocations seem to be benign in terms of determining the device lifetime.

Some questions arise: do dislocations in GaN LDs propagate in the same way as in conventional GaAs- and InP-based LDs? How low should we reduce the dislocation density in order to obtain reliable LDs? What characteristics of the dislocations are harmful to the device lifetime? The following section is to describe dislocations in GaN-based materials and LDs and to clarify these issues.

### *7.2.1 Dislocations in ELO GaN Layers*

The definition of ELO that is used here is that low dislocation regions (called "wing regions" hereafter) do not contact with the substrate and/or the underlying GaN layers; more specifically, the wing regions are suspended and are supported by high dislocation regions (called "seed regions" hereafter). This method was originally proposed by Zheleva et al. [8, 9], who used 6H-SiC as substrates, rather than the (0001) c-face sapphire substrates.

ELO GaN layers were formed as follows. A GaN layer was grown on a c-face sapphire substrate using metal-organic chemical vapor deposition (MOCVD), after which silicon dioxide ($SiO_2$) was deposited and patterned, forming stripe masks in the GaN $<1\bar{1}00>$ direction. The GaN window region was then etched out by reactive ion etching (RIE), and the $SiO_2$ mask was subsequently removed by wet etching. GaN was then laterally overgrown using the rectangular GaN stripes as seed regions. A schematic illustration of the cross section of the ELO material is shown in Fig. 7.1. It should be noted that a lateral growth front is exhibited close to the inclined $\{11\bar{2}2\}$ facets.

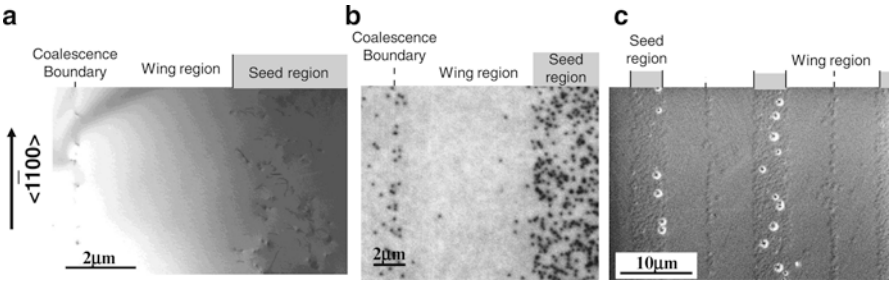



**Fig. 7.2** (**a**) Plan-view TEM image (**b**) Monochromatic CL image of ELO GaN layer. CL image was taken at 366 nm and at room temperature. (**c**) SEM image of the HCl gas etched ELO GaN surface

Figure 7.2a shows a typical plan-view TEM image of some ELO GaN layers. The threading dislocation density in the seed regions is about $4 \times 10^8/\text{cm}^2$. No dislocations were observed in the wing regions in this particular image. Threading dislocations aligned along the $<1\bar{1}00>$ directions were observed at the coalescence boundaries, and their density was estimated to be about $3 \times 10^8/\text{cm}^2$, which is similar to that of the seed regions. In GaN layers grown on sapphire substrates, three types of threading dislocations were formed with Burgers vectors of $1/3<11\bar{2}0>$ (**a**-type dislocations, edge character), (0001) (c-type dislocations, screw character), and $1/2<11\bar{2}3>$ (**a** + **c**-type dislocations, mixed character). In general, the Burgers vector of the c-axis component can hardly be determined from plan-view TEM observations, since the c-axis component of **g•b** always satisfies the zero condition. Thus, we also investigated ELO GaN layers using cathodoluminescence (CL) microscopy and the HCl gas etching method [10].

A monochromatic CL image of the ELO GaN layers is shown in Fig. 7.2b. The ELO GaN layers in the observed area were doped by Si. The image was taken at 366 nm and at room temperature. CL observations were also carried out at low accelerating voltages such as 2–3 kV, such that high-resolution CL images could be obtained. The density of dark spots (which corresponds to the density of threading dislocations) was greatly reduced in the wing regions from that in seed regions. Some of the dark spots were slightly darker than others. The slightly darker spots were observed in the seed regions and in the coalescence boundaries but were not observed in the wing regions. The photoluminescence (PL) intensity of the Si-doped GaN layers decreased with the number of **a** + **c**-type dislocations and **c**-type dislocations, although the PL intensity did not change significantly when the number of **a**-type dislocations was changed [10]. Based on the observation in Hino et al. [10], the slightly darker spots may correspond to **a** + **c**-type dislocations, and many of the other dark spots correspond to **a**-type dislocations. Therefore, almost all of the dislocations in the wing regions are concluded to be **a**-type dislocations.

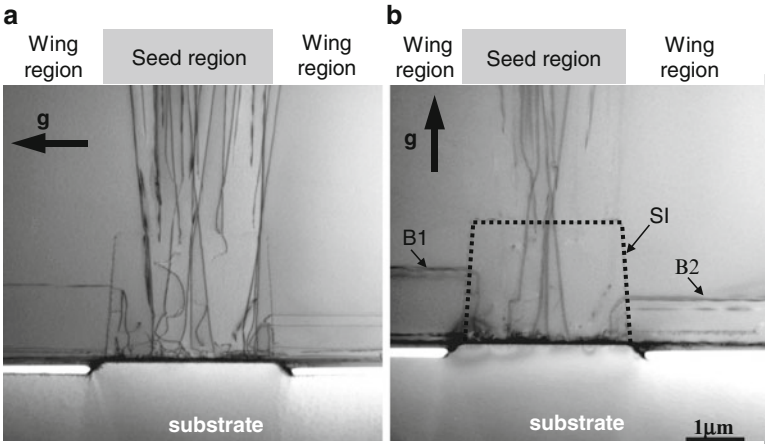



**Fig. 7.3** Cross-sectional TEM $<1\bar{1}00>$ images taken from the area in the vicinity of the seed region of ELO GaN layers with (**a**) $g = 11\bar{2}0$ and (**b**) $g = 0002$. "SI" indicates the side interface between first epi-layer of the seed region and the overgrown epi-layers

In order to characterize the spatial distribution of threading dislocations of different types in ELO GaN layers, it is appropriate to apply the HCl gas etching method, since the HCl gas etching technique can reveal different types of threading dislocations as different shapes of etch pits [10]. Figure 7.2c shows a scanning electron microscope (SEM) image of an etched ELO GaN surface. Three different kinds of etch pits are observed. The largest ones, the medium-sized ones, and the smallest ones correspond to **c**-type dislocations, **a** + **c**-type dislocations, and **a**-type dislocations, respectively. All three types of dislocations were observed in the seed regions. At the coalescence boundaries, **a** + **c**-type dislocations appeared at the coalescence boundaries. In the wing regions, only **a**-type dislocations were observed. Thus, we conclude that threading dislocations in the wing regions of ELO GaN layers are **a**-type dislocations. Both **a**-type dislocations and **a** + **c**-type dislocations were observed at the coalescence boundaries. This result is consistent with the deductions made from the CL observations.

### *7.2.2 Origin of the Threading Dislocations in the Wing Regions of ELO GaN Layers*

It is strongly desirable to fabricate a laser stripe on the region whose dislocation density as low as possible. However, in the early stage of development, low-dislocation-density substrates and/or templates were not available. Hence, laser stripe are fabricated on the wing regions of ELO GaN layers. Thus, it is important to know the origin of threading dislocations in the wing regions. Figure 7.3 displays some cross-sectional TEM images showing representative aspects of the defect

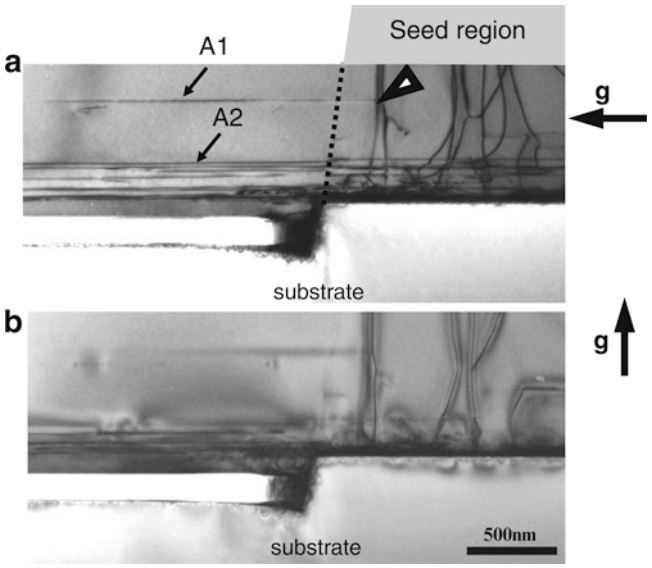



**Fig. 7.4** Cross-sectional TEM $<1\bar{1}00>$ images taken from the area near the interface between the epi-layers and the substrate with (**a**) $g = 11\bar{2}0$ and (**b**) $g = 0002$

structures in the vicinity of the seed regions of ELO GaN layers. The dotted lines in Fig. 7.3b indicate the interface between the first epilayer and the second overgrown epilayer. In the seed regions, the threading dislocations propagate almost vertically through the upper interface between the first epilayer and the overgrown layer. It is also observed that the dislocations extend laterally from the side interface, as indicated by SI in Fig. 7.3b. From the invisibility criterion of the TEM diffraction contrast, the dislocations that extend laterally are determined to be both **a**-type and **a** + **c**-type.

The lateral extension of the dislocations can be categorized into three types of behavior. The first type of behavior is due to dislocations bending at the side interface. The dislocations (e.g., those indicated by B1 and B2 in Fig. 7.3b) first propagate almost vertically but then bend to adopt a horizontal (0001) basal plane. The bending occurs at slightly inclined side interfaces. It should be noted that some dislocations first bend in an intermediate inclined direction before complete bending in the basal plane. The first type of behavior occurred for **a** + **c**-type dislocations. Sakai et al. reported similar behavior for $<11\bar{2}0>$-oriented stripe ELO GaN layers [11]. The second type of behavior is similar to the first type but occurs with **a**-type dislocations. A typical example is indicated by A1 in Fig. 7.4a. The vertically propagated dislocations suddenly bend by 90° into the basal plane near the side interfaces. The dislocation bending points are not exactly at the side interface but are not far from it, as shown by the arrowhead in Fig. 7.4a. Although this behavior seems to be less correlated to the inclined interface, the existence of the interface may be

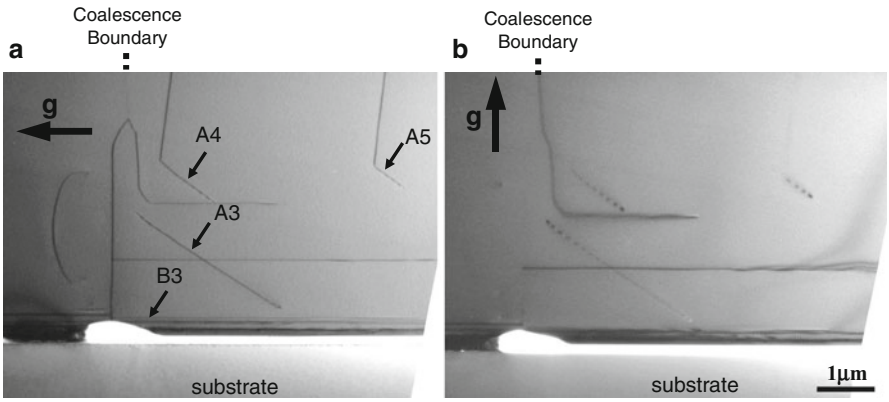



**Fig. 7.5** Cross-sectional TEM $<1\bar{1}00>$ images taken from the area in the vicinity of the coalescence boundary of ELO GaN layers with (**a**) $g = 11\bar{2}0$ (**b**) $g = 0002$

necessary in order for the dislocations to experience the stress field in order to cross slip to the basal plane.

The third behavior is simply the continuous lateral extension of dislocations lying in the basal plane in the seed regions of the first epilayer, as indicated by A2 in Fig. 7.4a. Heavily defective regions are formed in the bottom of the first GaN epilayer near the epi/substrate interface. The heavily defective regions mainly consist of stacking faults and of **a**-type dislocations lying in the basal planes. These defects simply propagate into the wing regions toward the coalescence boundaries. We note that some **a**-type dislocations on the basal plane are dissociated to Shockley partial dislocations separated by stacking faults [12]. Thus, stacking faults accompanied with Shockley partials and **a**-type dislocations on the basal plane are thought to be more or less similar kinds of defects.

At the coalescence boundaries, both **a**-type and **a** + **c**-type dislocations bend upward and thread almost parallel to the c-axis, as shown in Fig. 7.5. These are the origins of dislocations at the coalescence boundaries. Some dislocations running near the bottom of the wing regions vanished when they encountered the triangular voids that formed at the bottom of the coalescence boundaries (see, e.g., the dislocation denoted by B3 in Fig. 7.5a).

As shown at A3 in Fig. 7.5a, dislocations propagate at an angle of ~32° to the basal plane in the wing regions. These angled dislocations bend upward into the episurface, as shown by A4 and A5 in Fig. 7.5a, for example. In Fig. 7.6, we can see how some horizontal dislocations bend upward at an angle of ~32° to the basal plane. These angled dislocations are the origin of the dislocations in the episurface of the wing regions. It is important to note that this bending only occurred for **a**-type dislocations. As we concluded in the previous section, almost all of the dislocations in the wing regions are **a**-type dislocations. This is consistent with the results obtained in this section.

The propagation angle is almost perpendicular to the inclined lateral growth fronts that are close to the $\{11\bar{2}2\}$ facets. [The angle between the $\{11\bar{2}2\}$ facets and

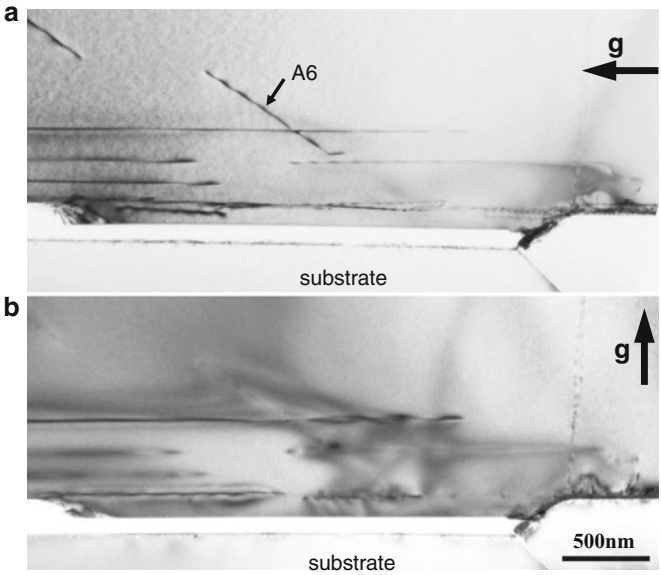



**Fig. 7.6** Cross-sectional TEM $<1\bar{1}00>$ images showing that dislocation marked as "A6" bend upwards at an angle of ~32°. The images were taken with (**a**) $g = 11\bar{2}0$ (**b**) $g = 0002$

the basal plane is 58°]. Therefore, we think that this type of bending is closely related to the growth process on the $\{11\bar{2}2\}$ facet plane on which the horizontal dislocations intersect. However, judging from investigations of a number of samples, we think that this type of bending occurs contingently due to some instability during lateral growth.

### *7.2.3 Comparison Between Short-Lived and Long-Lived LDs in Terms of Dislocation Distribution*

In the early stage of development of GaN-based LDs, it was found that the device lifetime shows a strong correlation with power consumption ($I_{op} \times V_{op}$) [4]. However, some LDs exhibited a short lifetime, although their power consumption was similar to that of LDs with a long lifetime. Both of these categories of devices were grown on the ELO GaN layers, and both were fabricated in the same manner, such as sharing the same seed/wing aspect ratio. Here, we compare the long lifetime devices with the short lifetime ones in terms of defects and consider the cause of the difference.

The lifetime of our initial development type of LDs under 30 mW CW operation at 50°C is plotted against power consumption ($I_{op} \times V_{op}$) in Fig. 7.7. The lifetime of the LDs denoted as Group-A tended to increase with decreasing power consumption, but the lifetime of the LDs denoted as Group-B was an order of magnitude smaller than that of the LDs in Group-A, in spite of having almost the same power consumption.

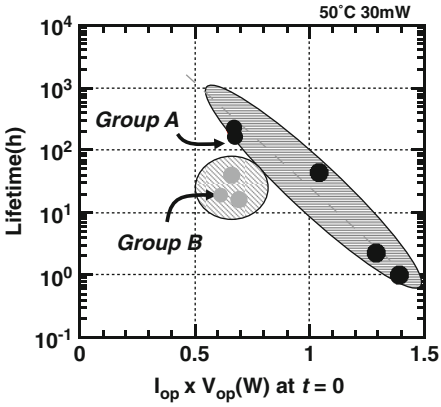



**Fig. 7.7** Lifetime versus power consumption under 30 mW CW operation at 50°C

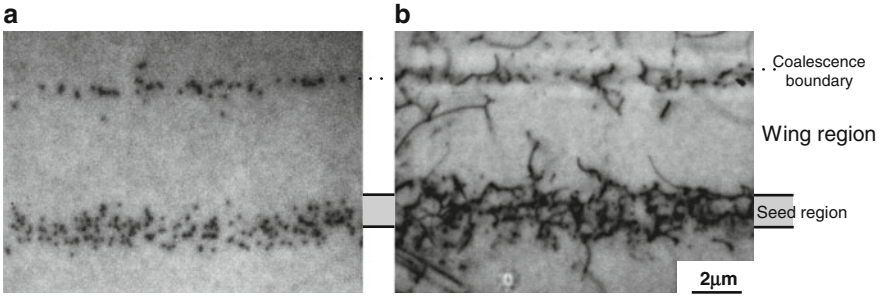



**Fig. 7.8** Monochromatic CL images of the n-GaN contact layers of (**a**) Group-A LD and (**b**) Group-B LD, taken at 366 nm

In order to understand the phenomena shown in Fig. 7.8, we compared Group-A LDs and Group-B LDs in terms of defects. Figure 7.9 shows plan-view monochromatic CL images of the n-GaN contact region of the LDs taken at a wavelength of 366 nm. The n-GaN contact regions were formed as follows: Ti/Pt/Au metal was evaporated on the n-GaN layer after the n-GaN layer was exposed to RIE. For CL observation, the n-type electrode metals were removed by chemical etching. As shown in Fig. 7.8a, the CL image of a Group-A LD has similar features to the image shown in Fig. 7.2b. On the other hand, in the CL image of a Group-B LD (Fig. 7.8b), lateral dark lines, which are thought to be dislocations in the basal plane [13, 14], are observed in the wing regions in addition to the dark spots such as observed in Fig. 7.8a. The lateral dark lines propagate from the seed regions and disappear in the wing regions. Some dark lines appeared independently in the wing regions and did not seem to be correlated with the seed regions.

If the dislocations in the basal plane are confined in the n-GaN layer and do not thread upward, the dislocation density in the active region of the Group-B LDs may be the same as that of the Group-A LDs. To confirm the vertical spatial configuration of the defect distribution, cross-sectional TEM observations of the Group-A

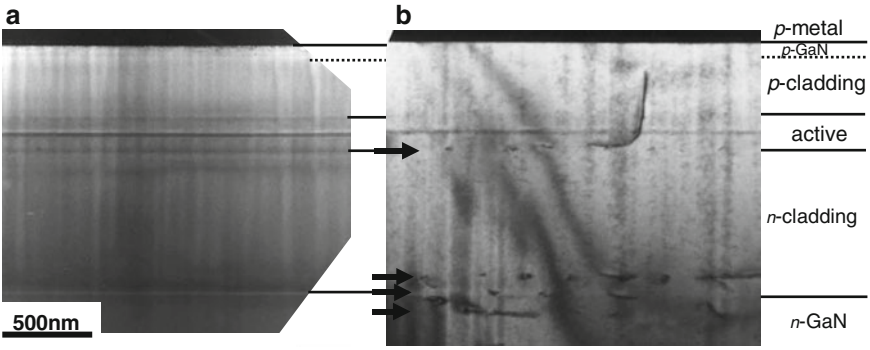



**Fig. 7.9** Cross-sectional TEM $<11\bar{2}0>$ images of (**a**) Group-A LD and (**b**) Group-B LD. The *arrows* indicate the positions that horizontal dislocations lying in the basal plane are seen

and the Group-B LDs were carried out (Fig. 7.9). We note here that the specimens used for the TEM studies in this section are not identical with those used in the aging tests but were sampled from the nearest laser chips of the same wafer lot. Few or no defects were observed in the n-GaN layers for Group-A LDs (Fig. 7.9a), while many horizontal dislocations lying in the basal plane were observed for Group-B LDs, both near to the interface between the n-cladding layer and the n-GaN layers and in the active regions, as indicated by the arrows in Fig. 7.9b. These dislocations are **a**-type dislocations, as determined by their diffraction contrast invisibility. They occur in pseudoperiodic arrays with periods ranging from 130 to 270 nm. One possible mechanism for the introduction of the dislocations in the basal plane is the relaxation of the lattice and/or thermal mismatch. Supposing that the lattice mismatch between the n-GaN layers and the n-cladding layers ($Al_{0.02}Ga_{0.92}N$) is fully relaxed by the introduction of pure edge dislocations with perpendicular Burger's vectors $b = 1/3<11\bar{2}0>$, we estimate that the periodicity of the dislocation array is roughly 166 nm. This value is within the observed value. The positions where the basal plane dislocations are introduced are not only exactly at the interface between the n-GaN layers and the n-cladding layers but are also below/above this interface. In addition, the basal plane dislocations also occurred at the interface between the n-cladding layer and the active layers. These observations cannot be solely explained by lattice and/or thermal mismatch relaxation. Since the growth procedures at the interface regions are complicated, some instability can occur, such as gas flow turbulence. We therefore think that growth front instability such as surface roughening (or faceting) during the growth can also lead to basal plane dislocations.

Another possible mechanism for the formation of the basal plane dislocations is threading dislocations coming from the bottom bend, which propagate laterally and then bend upward again. Dislocations are frequently observed that bend upward parallel with the c-axis and then thread through the active regions, as indicated by the arrows in Fig. 7.9b. As mentioned above, we think that dislocation bending is also caused by growth front instability, such as surface roughening (or faceting) during the growth.

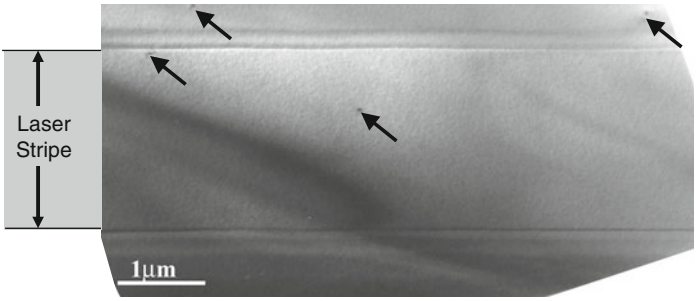



**Fig. 7.10** Plan-view TEM image of the stripe region of a degraded GaN LD

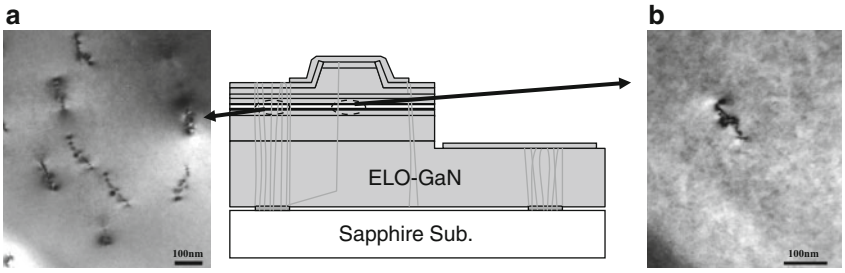



**Fig. 7.11** Magnified plan-view TEM images of a degraded GaN-based LD. (**a**) *Inside* the laser stripe. (**b**) *Outside* the laser stripe in the seed regions

In any case, the result is that these dislocations increase the dislocation density at the active layers in the wing regions and are detrimental to device lifetime.

### *7.2.4 Dislocations in Degraded LDs*

In this section, we describe the dislocation behavior after the degradation and discuss dislocation motion in GaN LDs. Figure 7.10 shows a plan-view TEM image of the stripe region of an LD degraded within the order of a 100 h under 30 mW continuous operation at 25°C. This particular LD has our initial development type of active layer structure [4]. The arrows in the figure indicate the threading dislocations. Some dislocations are inside the laser stripe and some are outside. No significant difference was observed in dislocation contrast between the inside and the outside of the laser stripe. Figure 7.11 shows magnified images of the threading dislocations inside the laser stripe and above the ELO seed region, respectively.

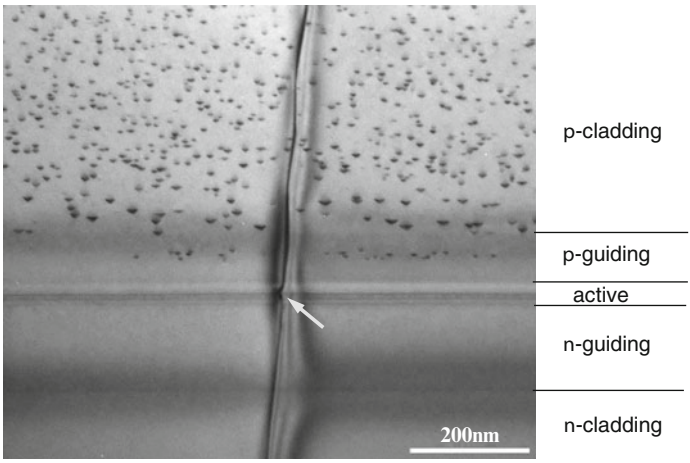



**Fig. 7.12** Cross-sectional TEM image of GaN-based LD. *White dotted circle* indicates a slight winding of the dislocations at the active layer

First, we discuss whether the plan-view TEM thin specimen shown in Fig. 7.11 includes the active region of the LD or not. Figure 7.12 shows a cross-sectional TEM image near the active region of the LD before the aging test. A threading dislocation was observed in the center of the image, which was an **a**-type dislocation, as determined by the invisibility criterion. The threading dislocation was wound at the active layer. Winding at the active layer was caused by stress concentration due to the lattice mismatch. The dislocations in Fig. 7.11 exhibited a dotted line-like contrast, which means that the dislocations thread through from the backside of the plan-view TEM thin specimen to the front side [15]. Moreover, additional contrast was observed in the middle of the dislocations. This contrast was caused by strain due to winding of the dislocations inside the thin TEM specimen. Therefore, we can conclude that the plan-view TEM thin specimen shown in Fig. 7.11 includes the active layer.

Now, let us compare the threading dislocations inside and outside of the laser stripe. Since the additional contrast in the middle of the dislocations due to winding is almost the same inside and outside of the stripe, there is no significant difference between these two regions. Furthermore, no dislocation multiplication from the threading dislocations was observed. Therefore, we conclude that there is no structural change in the threading dislocations after degradation, even in the case of the threading dislocations that exist inside the laser stripe.

The lifetime of LDs fabricated from III-V arsenides and phosphides or ZnSe-based II-VI materials is strongly related to the densities of any dislocations that are present [7, 16]. These devices are rapidly degraded by the growth of nonradiative regions at the active layers, which consist of dislocation networks from preexisting defects such as threading dislocations. Therefore, the existence of dislocations is fatal in terms of reliability. It is important to consider the recombination

enhancement effect for dislocation motion in light-emitting devices, since this effect is one of the primary causes of defect growth in III-V arsenides and phosphides or ZnSe-based II-VI materials for light-emitting devices by carrier injection. There are two modes for dislocation motion: dislocation glide and dislocation climb. Here, we focus on the dislocation glide motion. These devices have the zinc-blende (cubic) structure, whose slip systems are in the $\{111\}<110>$ directions. Thus, dislocation multiplication by recombination-enhanced dislocation glide (REDG) occurs along the $<110>$ directions, since the intersections between the (001) planes and the $\{111\}$ slip planes are in the $<110>$ directions. In these materials systems, 60° dislocations whose Burger's vectors are lying in the $\{111\}$ slip plane are quite common. An applied shear stress is needed for dislocations to glide and is mainly caused by misfit strain. Using Peach-Keohler's force [17], the applied shear stress for 60° dislocations on the $\{111\}$ plane is not zero, since the primary slip planes of the dislocations are inclined to the (001) plane of the device interface.

In the case of GaN LDs, their crystal structure is of the wurtzite (hexagonal) type, and their primary slip system is in the $\{0001\}<11\bar{2}0>$ directions. The lattice misfit does not give rise to a shear stress on the primary slip systems, since the primary slip planes are parallel to the device interface (0001) plane. Thus, the dislocations are thought to hardly glide at all. However, recombination enhancement effects accelerate dislocation motion, and it is important to consider this aspect.

To investigate the recombination enhancement effect on dislocations in GaN, we performed in situ carrier injection experiments in a TEM, in which we used the electron beam as a source of electronic excitation in the defective regions of GaN layers. The beam current density in this experiment was of the order of $10^2$A/cm$^2$, which is lower than the laser operating current density.

The arrowheads in Fig. 7.13 indicate a $<11\bar{2}0>-$oriented dislocation half loop that lies in the (0001) basal plane and which is an **a**-type dislocation. The tip of the loop is contained in the $\{1\bar{1}0n\}$ planes. Gliding of this dislocation half loop was observed. The displacement shown in Fig. 7.13 occurred within ~5 s. This leads to the conclusion that dislocation mobility is of the order of $10^{-5}$ cm/s. This value is considerably larger than the value of around $10^{-13}$ cm/s that Sugiura estimated for the nonenhancement case at room temperature [18]. Maeda et al. also reported the REDG effect on dislocations in single-crystalline powder of hexagonal GaN using in situ TEM [19]. We therefore think that our observation identifies the recombination enhancement effect. It is important to note that the recombination enhancement effect observed here even occurs below the laser operation current density. As described above, ideally the applied shear stress on the $\{0001\}$ plane is zero. However, in this case, the dislocation half loop lies close to the epi/subinterface, whose crystalline quality is quite poor. Thus, the shear stress for this dislocation half loop is not negligible, since the crystallographic axis is not thought to be well aligned. Our observation suggests that REDG can easily occur in GaN under the laser operation current density at the room temperature if a driving force is present.

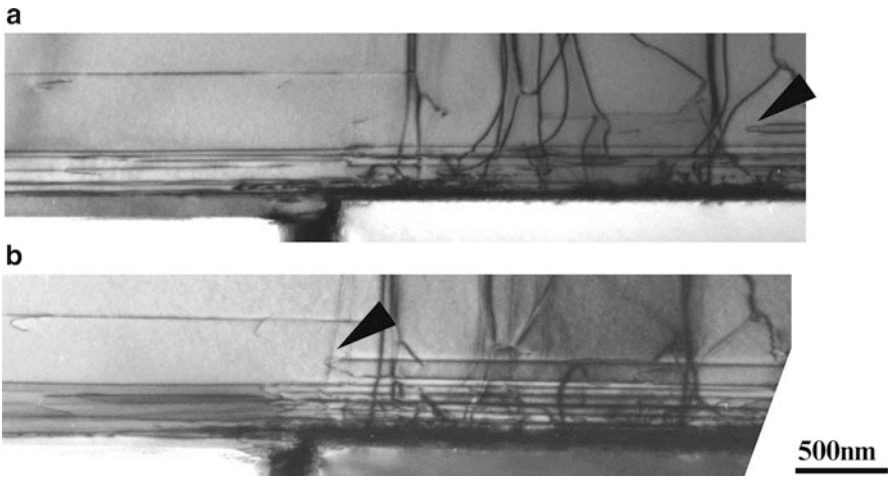



**Fig. 7.13** In-situ carrier injection experiments in a TEM using the electron beam as a source of electronic excitation in defective regions of GaN layers; (**a**) before and (**b**) after electron beam irradiation. The *arrowheads* indicate $<11\bar{2}0>$–oriented dislocation half loop lying in the (0001) basal plane

On the contrary, it is also important to note that no other dislocation motion was observed besides the dislocation half loops mentioned above. In order to grow the defects from the threading dislocations along the c-axis, there are two possible mechanisms. One is a direct defect growth process induced by a climbing motion. The other mechanism requires two processes: dislocation along the c-axis cross slip into the basal plane and growth by glide motion in the basal plane. In order to directly degrade the active layer of LDs by defect multiplication, one of these two mechanisms ought to be required. However, neither was observed under the conditions that we examined. We therefore conclude that this may be the reason why no dislocation multiplication from the threading dislocations was observed in the degraded LDs.

### *7.2.5 Degradation Rates of LDs*

The relationship between the dark spot density (DSD) and the mean time-to-failure (MTTF) of the LDs is shown in Fig. 7.14a. The LD structure grown on the ELO GaN layer here consists of a GaInN interlayer inserted between a GaInN multiple quantum well (MQW) active layer and an Mg-doped AlGaN electron-blocking layer. The DSD was estimated using a $\mu$-photoluminescence ($\mu$-PL) system. It is noted that the DSD by $\mu$-PL agrees with the DSD estimated by CL and the threading dislocation density estimated by TEM. Although the spatial

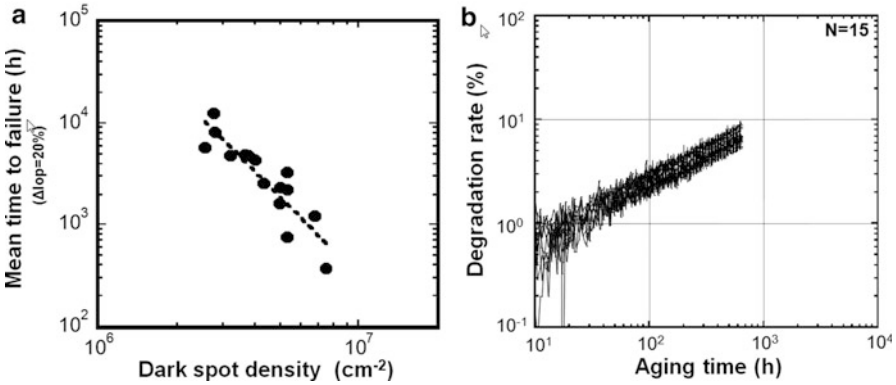



**Fig. 7.14** (**a**) Relationship between dark spot density (DSD) estimated by the m-PL system and mean time to failure (MTTF) of LDs. (**b**) Relationship between degradation rate and aging time. Both were taken under 30 mW CW operation at 60°C

resolution of the $\mu$-PL system is poorer than that of the CL system and the TEM, the advantage of the use of the $\mu$-PL system is that it is less time-consuming and it is easy to estimate a wider area at the same time. The lasers were aged under 30 mW CW operation at 60°C, and the lifetime was defined as the time at which the operating current increased by 20%. The power consumption of all of the lasers used to estimate the MTTF was about 0.33 W and was independent of the dislocation density. As we described previously, the reduction of threading dislocations in the active region is very effective for the enhancement of device reliability. The result obtained here provides additional evidence for the strong correlation between DSD and MTTF. In addition, the lifetime of the LDs grown homoepitaxially on GaN substrates, whose dislocation density is less than $1 \times 10^6/\text{cm}^2$, is expected to exceed 100,000 h [20]. This result also provides additional evidence.

In the present experiments, the degradation rate was found to be almost proportional to the square root of the aging time, as shown in Fig. 7.14b. Similar results were also reported by Marona et al. [21]. These results indicate that the degradation is governed by a diffusion process. Since degradation is not caused by dislocation multiplication at the active layers, degradation may be caused by diffusion of point defects toward the active layers. In addition, judging from the fact that minimization of the threading dislocation density is critical for the longevity of LDs, the diffusion process discussed here is enhanced by the threading dislocations. These observations lead to the following model: degradation proceeds by the diffusion of point defects toward the vicinity of the threading dislocations at the active layer.

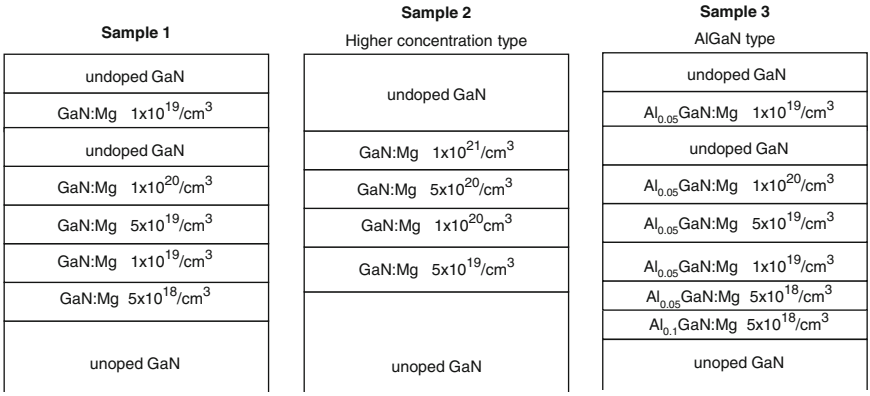



**Fig. 7.15** Schematic illustration of Mg step-doped samples. The figures for Mg concentration [Mg] indicated in each layer are the designed values

## 7.3 Defects in Mg-Doped GaN-Related Materials: Pyramidal Defects

P-type doping in GaN and related materials had been a big obstacle for light-emitting devices because efforts to achieve p-type GaN led to compensated high-resistivity materials. However, in 1989, Akasaki and Amaono made the unprecedented breakthrough by using low-energy electron beam irradiation (LEEBI) on the Mg-doped GaN [22]. Hole concentrations as high as $3 \times 10^{18}/\text{cm}^3$ and a resistivity of 0.2 Ω cm were obtained. Afterward, Nakamura et al. also obtained p-type GaN by thermal annealing at 700°C under an $N_2$ [23]. By realizing favorable p-type doping, development of GaN-based LEDs had been dramatically progressed.

Minimizing the power consumption is of critical importance for not only improving the device performance but also extending the device lifetime [4]. In general, LDs require higher current density than LEDs and thus require higher free hole concentrations and lower p-resistivity for p-conducting layer. One of the key issues is the optimization of the Mg doping conditions for GaN and AlGaN layers. Heavy doping is often tried in attempt to achieve high hole concentrations. However, as Mg concentration [Mg] increases beyond a certain concentration, free hole concentration is not obtained higher but rather starts to decreased. One of the reasons for this is that inverse pyramidal-shaped structural defects (which are hereafter referred to simply as "pyramidal defects") are generated when [Mg] is high enough. For example, in early stage of development of LDs, pyramidal defects were often found in Mg-doped layers by TEM observations as shown in Fig. 7.12 [24]. Such defects have also been reported by several researchers [25–27]. In this section, structural characterization of these pyramidal defects is described mainly based on our study [28].

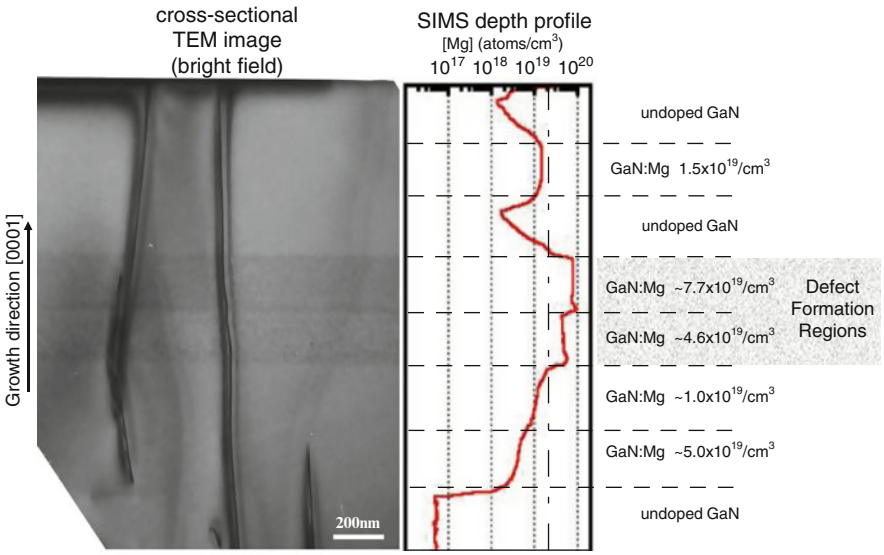



**Fig. 7.16** Cross-sectional TEM image, together with SIMS depth profile of Mg for sample 1

### *7.3.1 Dependence of Defect Generation on Mg Doping Concentrations*

In order to study the effects of Mg doping on the defect generation, as shown in Fig. 7.15, Mg step-doped structures were grown on c-face sapphire substrates using MOCVD at growth temperature ~1,000°C. Growth was carried out under an $H_2$ ambient.

These samples are characterized by secondary ion mass spectrometry (SIMS) and TEM. The samples were not annealed prior to TEM and SIMS studies. Figure 7.16 shows a cross-sectional bright field TEM image together with a SIMS depth profile of Mg for sample 1. Besides threading dislocations, dark bands can be observed in the center of the image. The dark doted bands consist of many pyramidal defects, although each pyramidal defect is not well resolved in this image because of the low magnification. The dark doted bands are observed only in layers 4 and 5 of sample 1. This indicates that the pyramidal defects form when [Mg] is higher than ~$4 \times 10^{19}/cm^3$.

Sample 2 was doped with a higher concentration of Mg than sample 1. Figure 7.17 shows a cross-sectional weak beam (0002) dark-field TEM image and a SIMS depth profile of Mg for sample 2. Magnified images of the 1st and 4th Mg-doped layers are shown in Fig. 7.18a. Projecting triangular shapes are clearly visible, and the tips of the defects are pointing toward the substrate. The size and density dependence on [Mg] are also shown in Fig. 7.18b. The density of the

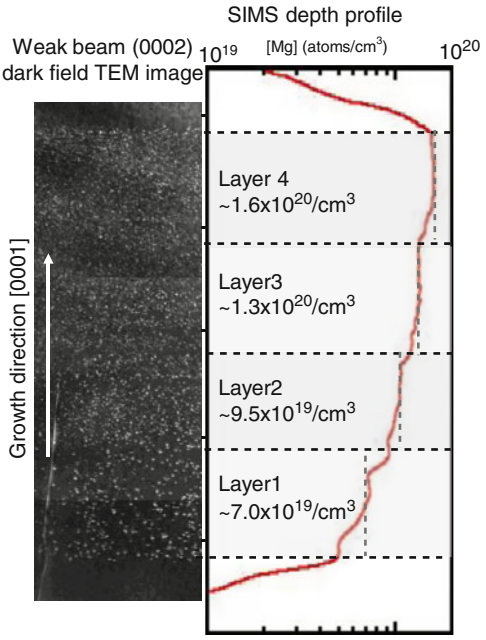



**Fig. 7.17** Cross-sectional TEM image and SIMS depth profile of Mg for sample 2

pyramidal defects increased as [Mg] increased, but their size decreased as [Mg] increased.

Figure 7.19 shows a cross-sectional bright field TEM image and a SIMS depth profile of Mg and H for sample 3. The dark bands were also observed only in layers 4 and 5 of sample 3. This situation is similar to that of sample 1. We have also measured the concentration of hydrogen in the layers, [H]. In undoped GaN layers (layer 1), the residual [H] is about $5 \times 10^{17}/\mathrm{cm}^3$. [H] increases with increasing [Mg] but becomes saturated at $\sim 5 \times 10^{19}/\mathrm{cm}^3$, while [Mg] can attain over $5 \times 10^{19}/\mathrm{cm}^3$. Comparing the TEM image to the SIMS depth profile, [Mg] tends to be higher than [H] in the defect formation layer. This result suggests that defect formation requires a higher [Mg] compared to [H].

### *7.3.2 Microstructure of Pyramidal Defect*

Figure 7.20 shows various kinds of dark-field images of the pyramidal defects. The defects appear brighter than the matrix in the multiple (0002) dark-field image and appear darker than the matrix in the multiple (0002) dark-field image, as shown in Fig. 7.20a and b, respectively. These contrast changes indicate that there is an inversion operation relationship between the pyramidal defects and the matrix [27, 29]. In the weak beam (0002) dark-field image (Fig. 7.20c), projecting triangular shapes are clearly visible, and the pyramidal defects appear bright, which may be

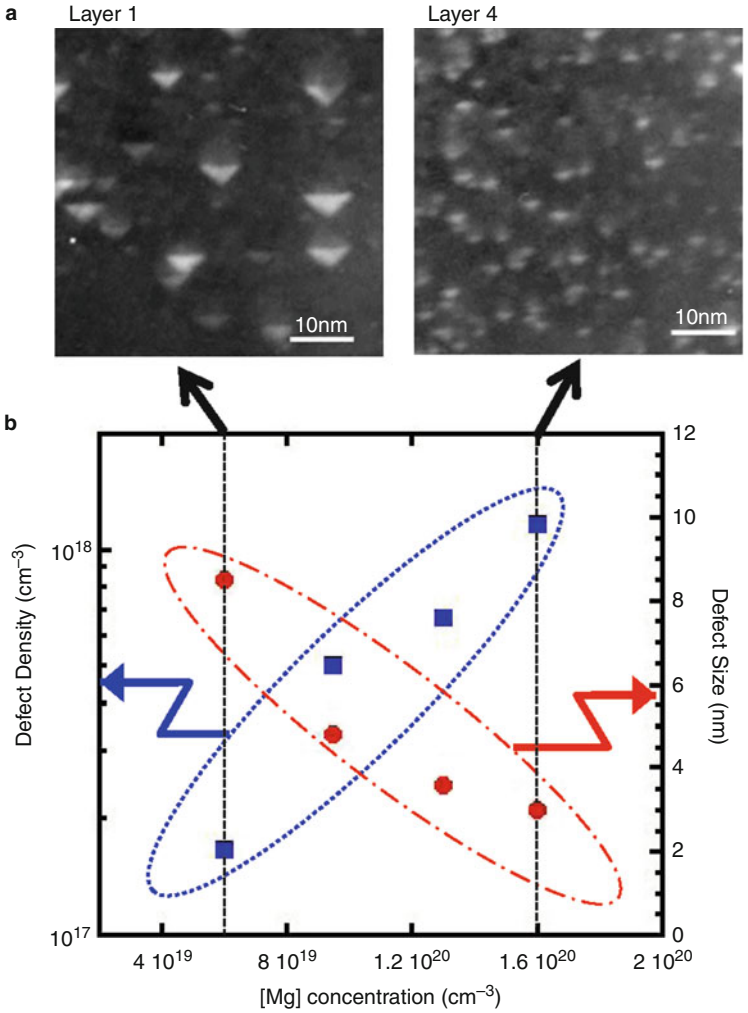



**Fig. 7.18** (**a**) Magnified weak beam (0002) dark field image of layer 1 and layer 4 of sample 2. (**b**) The dependence of defect size and defect density on Mg concentration

due to a small lattice translation between the pyramidal defects and the matrix. This is supported by the results of high-resolution TEM observation (Fig. 7.21), which show that the lattice planes shift ~c/8 (~1/4 monolayer) between the inside of the defects and the outside. This result agrees with first-principles pseudopotential density functional calculations for Mg-rich inversion domain boundaries conducted by Northrup [30]. His model boundary is of the form of GaNMgNGa layers that are stacked in *abcab* registry, with at least 3/4 of the available c sites occupied by Mg atoms. An additional 1/4 monolayer of Mg can be incorporated in this layer provided additional compensating Mg acceptors are located nearby. Figure 7.20d

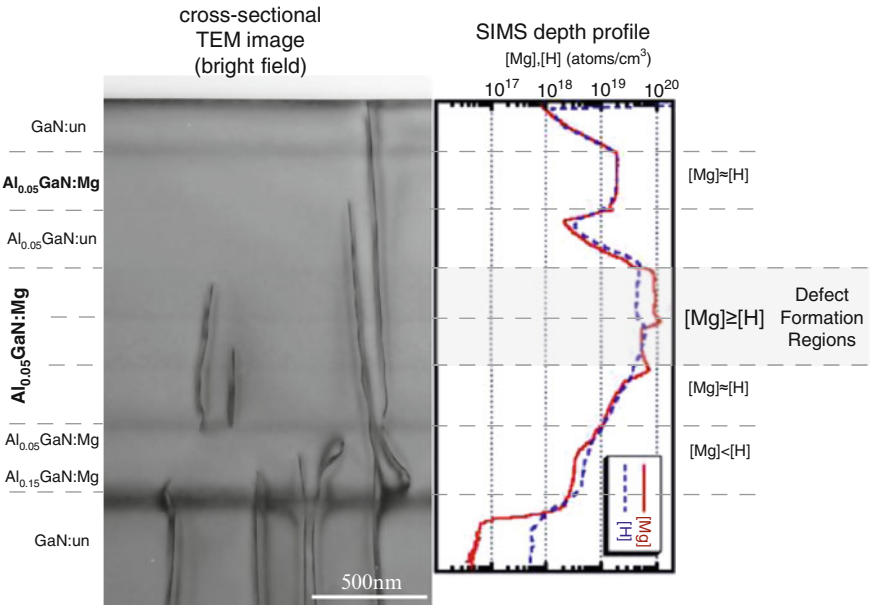



**Fig. 7.19** Cross-sectional TEM image, together with SIMS depth profile of Mg for sample 3

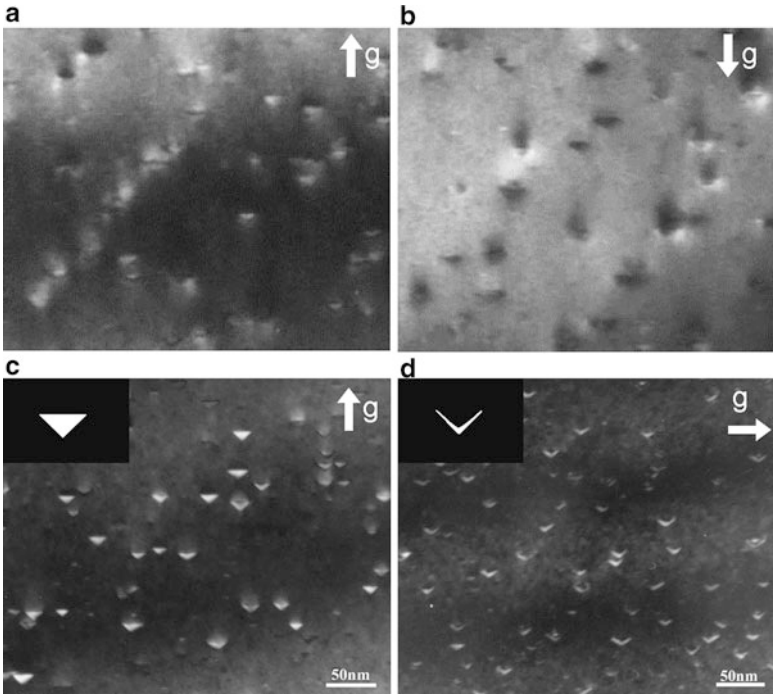



**Fig. 7.20** Dark field images of the pyramidal defects. (**a**, **b**) are the multiple (0002) dark field image and the multiple $(000\bar{2})$ dark field image, respectively. (**c**, **d**) are the weak beam (0002) dark field image and the weak beam $(000\bar{2})$ dark field image

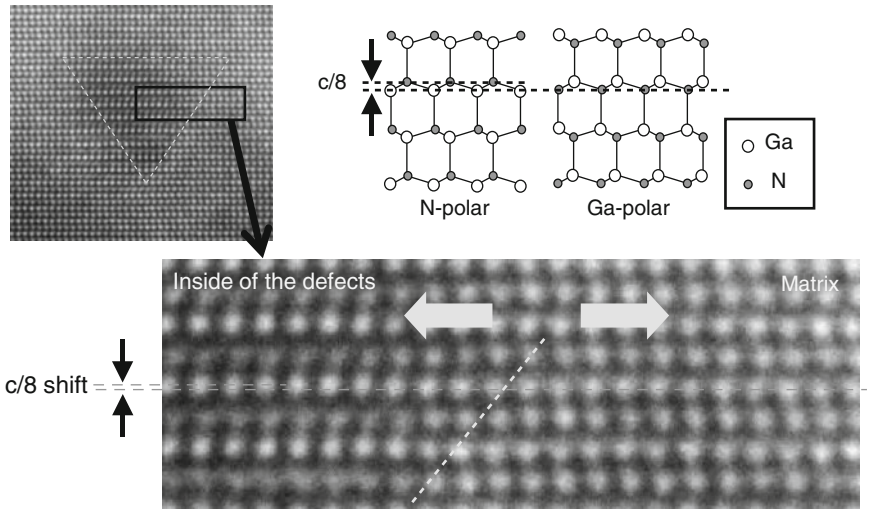



**Fig. 7.21** HREM image of pyramidal defects

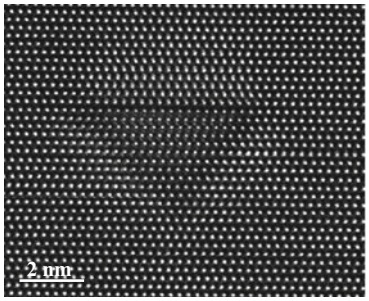



**Fig. 7.22** HAADF-STEM image of pyramidal defects by using probe Cs-corrected TEM

shows the weak beam $(11\bar{2}0)$ dark-field image. Two sides of the projecting triangular shapes are brighter. Since weak beam imaging is sensitive to lattice displacement, the lattice planes at the boundaries are thought to be slightly changed. We therefore speculate that the boundaries of the pyramidal defects are Mg-rich. To confirm this, HAADF-STEM observation of the pyramidal defects in the GaN epitaxial layer was performed, as shown in Fig. 7.22, since the intensity of the HAADF-STEM image depends on the atomic number. The defects themselves are darker than the matrix, and moreover, the boundaries are much darker than the inside of the defects. Since Mg is lighter than Ga, the boundaries are thought to be more Mg-rich than the inside of the defects and that part of the boundaries may consist of Mg-rich layer such as $Mg_3N_2$ and $(Ga_xMg_{0.75(1-x)})N$. The structural model based on these results is shown in Fig. 7.23.

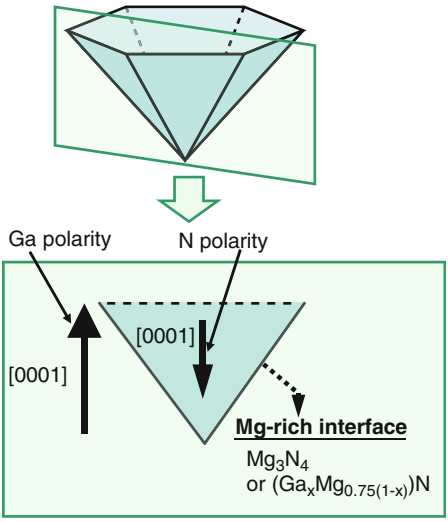



**Fig. 7.23** Schematic illustration of the model of pyramidal defects

Since the size of the pyramidal defects is usually very small, it is rather difficult to obtain direct evidence for the structural model by conventional analytical techniques. In order to prove the structural model shown in Fig. 7.23, we grew Mg-doped epitaxial layers with overdoping Mg interfaces. Figure 7.24 shows a cross-sectional bright field TEM image, which was taken along the $[1\bar{1}00]$ axis. Because of the overdoping of Mg, the interface is faceted. Above the interface, the apexes of the pyramidal defects lie in the opposite directions. Convergent beam electron diffraction (CBED) patterns are also shown in Fig. 7.25. The CBED patterns were taken in the matrix regions along the $<1\bar{1}00>$ direction, above and below the interface. The obtained CBED patterns indicate that the polarity has inverted across the interface. This suggests that the directions of the pyramidal defects are related to the film polarity and that there is an inversion operation relationship between the pyramidal defects and the matrix. In addition, Mg was also detected at the interface by energy dispersive X-ray spectrometry (EDX) (Fig. 7.25). These results support the structural model shown in Fig. 7.23.

### *7.3.3 Formation Mechanism of Pyramidal Defect*

As shown in Fig. 7.18, the defects typically are of 2.5–8 nm in size. From the structural model, the amount of magnesium situated on the surface of the defect can be estimated to be around 150–1,000 atoms, when the 3:1 coverage with magnesium and gallium such as $(Ga_xMg_{0.75(1-x)})N$ is supposed. Since the volume of such a defect contains 500–20,000 atoms, the average magnesium concentration in the

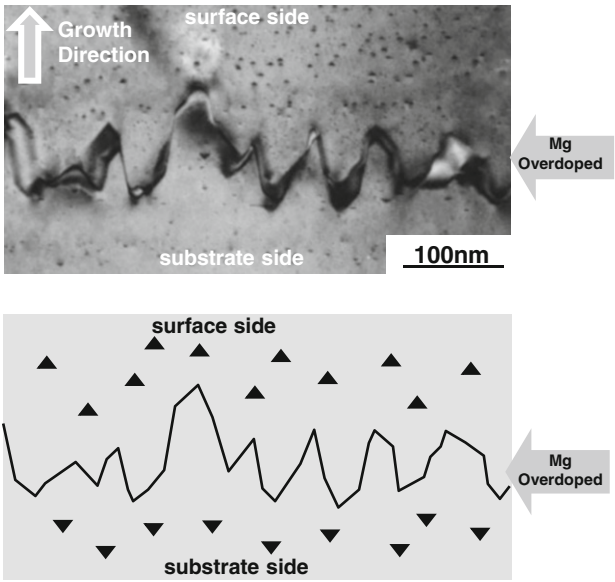



**Fig. 7.24** Bright field TEM image of Mg-overdosed GaN layers. *Black triangles* indicate the pyramidal defects in schematic illustration shown in below BF-TEM image. Directions of apex of the pyramidal defects are inversed between substrate side and surface side

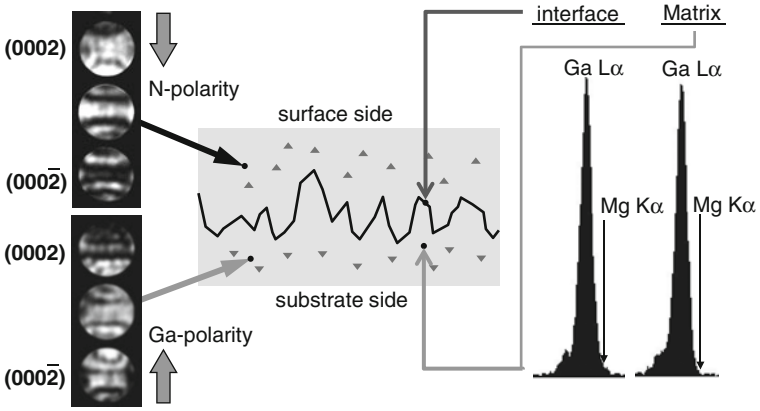



**Fig. 7.25** CBED and EDX analysis of Mg-overdosed GaN layers. CBED patterns were taken at the matrix both from *upper* and *lower* regions of the interface. EDX spectra were taken at the interface and the matrix of the lower region

defect is around 20–30%, which is two orders of magnitude higher than in the surrounding matrix (~$10^{19}$/cm$^3$). Thus, formation of the defects has to be required a huge amount of magnesium in growing surface.

During Mg-doped GaN growth on (0001) plan, accumulation of Mg occurs with stable accumulations of close to monolayer of Mg on the Ga-polarity surface [31]. Other reported evidence for surface accumulation of Mg for Mg-doped GaN has been based on Auger spectroscopy of layers after growth [32] and on the observation that Mg acts as a persistent surfactant for the growth of GaN [33]. We accept these as being valid for MOCVD growth. As the amount of Mg precursor injected into the reactor increases in order to dope more heavily with Mg, Mg atoms start to segregate at the growth surface.

An amount of hydrogen is also key role to form the defects as was understood from Fig. 7.19. When the concentration of hydrogen at the surface is sufficient, Mg will incorporate uniformly into the layer. However, when the concentration of hydrogen is insufficient, magnesium nitride clusters may be created at the growth surface. High exposure to Mg forms a faceted surface, and the subsequent layer is of inverted polarity as shown in Fig. 7.24 and ref. [34]. Similar phenomena occurred in this case. Growth is suppressed at the point of the magnesium nitride clusters, and consequently small pits are formed, just like forming a faceted surface. The polarity is also changed inside the pits. Since the growth rate of an N-polar surface is significantly slower than that of a Ga-polar surface in MOCVD growth, the pits are soon buried by lateral growth from the surrounding Ga-polar matrix. The formation of the pyramidal defects is then terminated. It has been reported that a higher concentration of Mg in the vapor phase enhances the ratio of the lateral growth rate to the vertical growth rate [35]. When [Mg] increases, the defects are consequently buried faster due to an enhancement of the lateral growth rate. This leads that the size of the pyramidal defects becoming smaller.

### *7.3.4 Effects on Conductivity by Pyramidal Defect*

In GaN:Mg grown by MOCVD, concentration of free holes at room temperature reaches its maximum value of about $10^{18}$/cm$^3$ for the Mg concentration of about $3 \times 10^{19}$/cm$^3$, and it decreases with farther increase of Mg concentration [36]. This result matches up precisely with formation of the pyramidal defects.

$Mg_{Ga}$ exhibits a relatively shallow acceptor [37–39]. However, before invention of LEEBI treatment method [22] and thermal treatment method [23], attempts of Mg doping were not successful presumably because of (1) high hydrogen presence in MOCVD grown samples and (2) self-compensation effect due to heavily Mg doping.

Although hydrogen enhances Mg incorporation, hydrogen passivates Mg by forming electrically and optically inactive [40]. Meanwhile, passivation with H during growth in fact favorable for obtaining high-conductivity p-type GaN because the hydrogen inhibits the formation of native donors which self-compensate the acceptors [41, 42]. However, once amount of hydrogen depletes compared to that of magnesium, magnesium segregates on growing surface, and pyramidal

defects are formed. Ab initio calculations on the stability of the facets with different Mg content suggest a termination of the surface with magnesium and gallium at a ratio of 3 to 1, respectively [30]. This ratio satisfies the electron counting rule, but additional gallium on the vicinal surface would provide donor-like states. These donor-like states owing to formation of the pyramidal defects compensate and, then, degrade p-type conductivity.

In addition, Mg-doped GaN shows large optical absorption at the lasing wavelength in contrast to Si-doped or intrinsic GaN [43]. Higher Mg concentration may exhibit larger pronounced absorption. Since internal loss of the LDs is increased by Mg-doped layers [44], the pyramidal defects, which form as a consequence of high Mg doping, can increase operating current of the LDs as a result of increase of the internal loss.

Hansen et al. reported that the formation of the pyramidal defects increased the threshold voltage of the LDs by 30% [27]. Since device lifetime has a strong correlation with power consumption ($I_{op} \times V_{op}$) [4], as described the previous section, not only the reduction of threading dislocations but also the reduction of the pyramidal defects also contributes to extending the device lifetime.

## 7.4 Defects in InGaN Multiple Quantum Wells

Near-UV emission GaN-based LDs have already been realized and are now mass-market products with applications in "Blu-ray" for optical data storage. Blue-emission LDs are also appearing in applications such as laser projection display systems. In addition, there have recently been increased requirement to improve the emission efficiencies of green LEDs and to realize green LDs.

Because of these changing circumstances, it is important to understand the material properties of the ternary alloy GaInN, since this alloy is inevitably used as one element of active layers in GaN-based optical devices. This alloy makes it possible to cover a wide spectral range by simply changing the indium concentration. However, one technological problem that effects the growth of GaInN alloys with higher In concentrations is the low miscibility of InN and GaN [45, 46]. This leads to phase separations in the alloys, which in turn leads to the formation of In-rich precipitates and clusters [47, 48]. Another type of structural defect that is frequently reported in GaInN alloys is the so-called V defect [49–54], which takes the form of an open hexagonal, inverted pyramid with $\{10\bar{1}1\}$ side walls [51]. Indium incorporation can vary at or near these V defects. This variation in indium incorporation enhances the effects of compositional fluctuations and/or phase separation. Since structural imperfections such as this can deteriorate emission efficiency, it is important to minimize these structural imperfections.

Above-mentioned structural imperfection has been successfully minimized to achieve high-power pure-blue LDs [55]. However, a new type of structural defects, which was not reported before, has proved very hard to reduce [56]. These defects

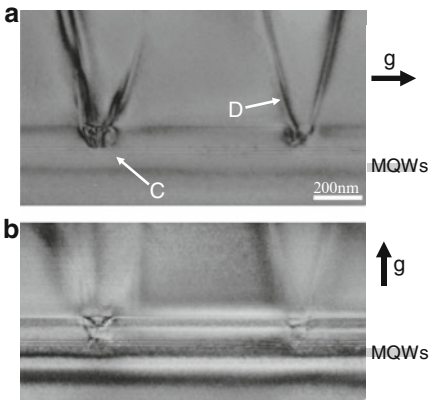



**Fig. 7.26** Cross-sectional TEM $<1\bar{1}00>$ images of the multiple defects with (**a**) $g = 1-1\bar{2}0$ and (**b**) $g = 0002$

nucleate at GaInN multiple quantum wells and are combination of columnar defects with dislocations [57]. (These defects are hereafter simply are referred to as "multiple defects.") In this section, a detailed structural analysis of these multiple defects is described, and their formation mechanism is also discussed. In this section, we first describe on structural analysis of the multiple defects. Subsequently, the degradation analysis of pure-blue LDs is described.

### *7.4.1 Structural Defects in GaInN Multiple Quantum Wells*

To investigate the multiple defects that we had newly observed nucleating in GaInN MQWs with increased In concentrations, a series of MQW structures and LD structures were grown on GaN substrates by MOCVD. The indium concentrations in the quantum wells were higher than 12%. For the barrier layer materials that formed part of the MQW structures, we examined the use of $Ga_{0.98}In_{0.02}N$, GaN, and $Al_{0.02}Ga_{0.98}N$. The densities of threading dislocations in the GaN substrates were less than $1 \times 10^6/cm^2$.

Figure 7.26 shows cross-sectional TEM images of the multiple defects formed in the vicinity of GaInN MQWs. Columnar defects are observed at the MQWs, as indicated by C in Fig. 7.26a. These columnar defects did not originate at threading dislocations generated from the underlying layers. However, dislocations that nucleated from the columnar defects were observed, as indicated by D. These dislocations denoted by D run upward toward the episurface. As seen in Fig. 7.26, the multiple defects consist of a combination of columnar defects and dislocations. In Fig. 7.26b, the contrast caused by the dislocations has disappeared.

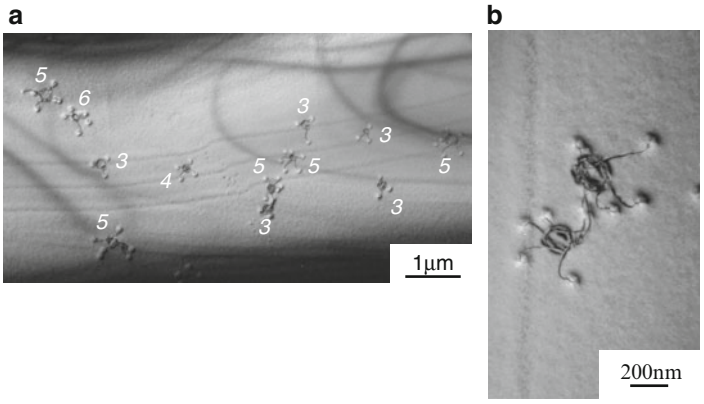



**Fig. 7.27** Plan-view TEM image of the multiple defects

From the invisibility criterion of the TEM diffraction contrast, the Burger's vector of the dislocations is determined to be $1/3<11\bar{2}0>$ (**a**-type dislocations).

Figure 7.27 shows plan-view TEM images of the multiple defects. Although the lateral size of the columnar defects depends on various parameters such as growth conditions and indium concentrations, their typical sizes ranges from about 50 to about 200 nm. In the cross-sectional images in Fig. 7.26, several dislocations are observed as being generated from the boundary edges of the planar defects. As shown in Fig. 7.27a, the number of **a**-type dislocations generated from a single multiple defect is somewhere between three and six.

In order to investigate the columnar defects, weak beam $(11\,\bar{2}\,0)$ dark-field observations were performed (Fig. 7.28a). The three images are rotated from one another along the c-axis. The region of bright contrast in the columnar defects varies among these images. This indicates the region of bright contrast in the columnar defects varies among these images. This indicates that the columnar defects consist of subgrains that are twisted slightly with respect to one another. Moreover, the columnar defects themselves are also twisted from their surroundings in the normal matrix regions. This twisting creates the **a**-type dislocations from the boundaries of the planar defects.

A high-resolution lattice image of the columnar defects is shown in Fig. 7.29. The columnar defects are formed at the interfaces between the quantum well layers and the barrier layers. To determine the polarity of the columnar defects, CBED experiments were performed together with HAADF-STEM observations, as shown in Fig. 7.30. The CBED patterns were taken from the $[1\bar{1}00]$ zone axis. The CBED patterns indicate that the position from inside the defects and the position from outside the defects have opposite polarities, as seen from the different contrasts in the 0002 and 000$\bar{2}$ disks, due to the strong dynamical effects of electron diffraction (Fig. 7.30a and c). Figure 7.30d shows Bloch-wave simulated CBED patterns with thicknesses of about 90 nm, which match the experimentally obtained CBED

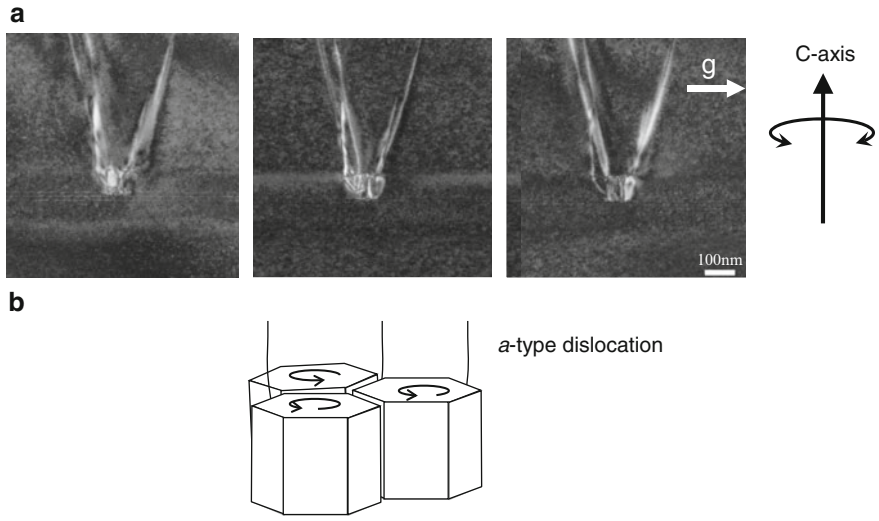



**Fig. 7.28** (**a**) Weak beam ($11\bar{2}0$) dark field images of the multiple defects at MQWs taken under different conditions, which are slightly rotated among the three images. (**b**) Schematic illustrations of the columnar defects, which consist of sub-grains

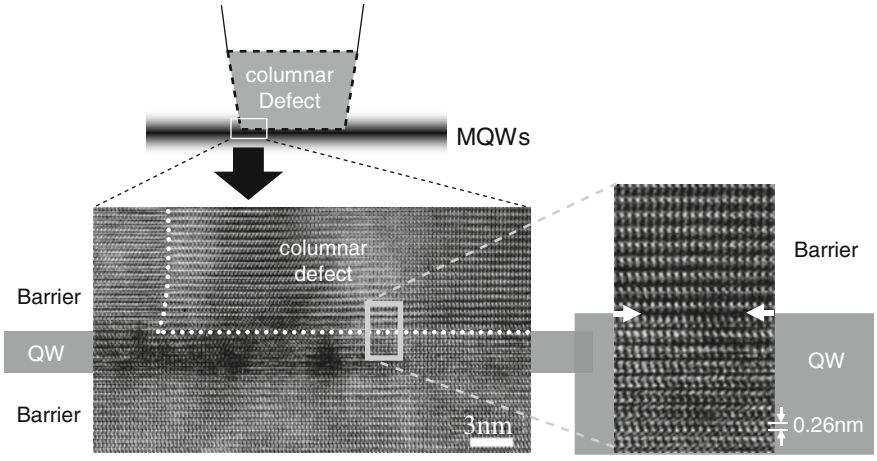



**Fig. 7.29** $<11\bar{2}0>$ High-resolution lattice image of the columnar defects

patterns. Thus, the columnar defects have inverse polarity from the matrix and are thought to constitute inversion domains.

In-rich precipitates are frequently observed in the vicinity of the multiple defects, as shown in Fig. 7.31. This observation suggests that the In-rich precipitates are strongly related to the formation of the columnar defects. If we

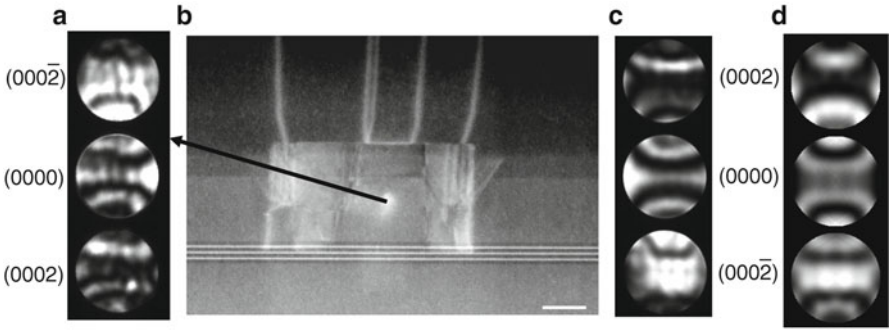



**Fig. 7.30** (**a**) [1$\bar{1}$00] Zone axis convergent-beam electron diffraction patterns from the inside of columnar defects and (**c**) normal (matrix) regions of the p-GaN layers. (**b**) Corresponding HAADF-STEM image. (**d**) Simulated CBED patterns with TEM foil thickness of about 90 nm

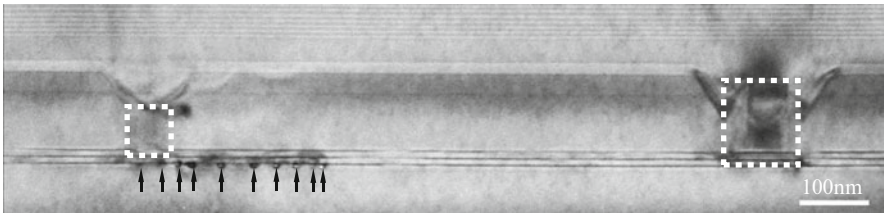



**Fig. 7.31** Cross-sectional bright field TEM image of the multiple defects. The columnar defects and the dislocations are out of contrast. The columnar defects are indicated by the *dotted lines*. The *arrows* indicate In-rich precipitates

suppose that the stacking sequence of the closed-packed plane of the wurtzite structure is a-A-b-B-a-A-b-B-, then the stacking sequence of (0001) plane along the c-axis has to be a-A-b-B-a-A-b-B-B-b-A-a-B-b in order to form the inversion domains (IDs). It is noted that B-B is the interface of the inversion domain boundary. Since the In-rich precipitates were nucleated near the multiple defects, the excess B-B bonds are thought to be due to In-In parings. A schematic illustration of the structural model is shown in Fig. 7.32. We believe that the ID defects are generated due to the formation of excess In-In bonds at the interface between the quantum wells and the barrier layers.

On the basis of these results, it is important to suppress the formation of In-In bonds, particularly at the interface between the quantum well layer and the barrier layer in order to reduce the density of multiple defects. By focusing on the optimization of the growth procedure of the interface between the quantum well layer and the barrier layer and the use of an AlGaN ternary alloy as the barrier layer instead of GaN and InGaN ternary alloys, the density of the multiple defects can be reduced to less than ~5 $\times$ $10^5/cm^2$ [55]. This result also suggests that the presence of Al atoms prevents the formation of In-In bonds at the interface.

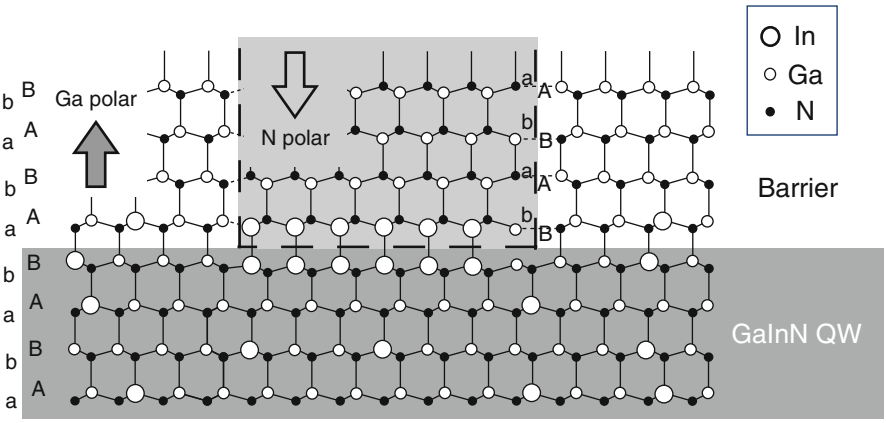



**Fig. 7.32** Schematic illustration of structural model of the columnar defects nucleated at the interface between quantum well layer and barrier layer

### *7.4.2 LDs with Current Injection-Free Regions*

By simultaneously reducing the multiple defects and the operating current density by optimization of the device structures, GaN-based LDs with emission wavelengths of 440–450 nm and output powers of 0.75 W under CW operation at 35°C had been achieved [61]. However, when the cavity lengths of the LDs were changed, sudden failures frequently occurred. By introducing current injection-free (CIF) regions near the facets of the LDs, we have confirmed that LDs incorporating CIF regions exhibit improved suppression of sudden failures [58].

After the design cavity length of the LDs has changed from 2.0 to 1.4 mm, sudden failures have frequently occurred [58]. Damages to the front facets of the degraded devices were observed by using cross-sectional scanning electron microscopy. Thus, we speculated that the phenomena observed here were due to the facet-related degradation. As a countermeasure against facet-related degradation, nonabsorbing facets (NAFs) have been used in high-power AlGaInP-based LDs [59]. The NAFs were fabricated by a local band-gap widening technique near the facets involving impurity-induced disordering. Another attempt to improve high-power AlGaInP-based LDs was the introduction of a current-blocking region by a second epitaxial growth to deposit an n-type layer [60]. Compared with these techniques, LDs with a current injection-free (CIF) region near the laser facet are relatively simple and effective for high-power AlGaInP-based LDs [61]. We therefore introduced a CIF region into our LDs [58]. The CIF region was fabricated by the removal of the p-electrode metals near the laser facets. The stripe width was 12 μm, and their cavity length was 1.4 mm. The lengths of the CIF regions from the facets were 15–45 μm. We also fabricated the LDs without a CIF region for reference.

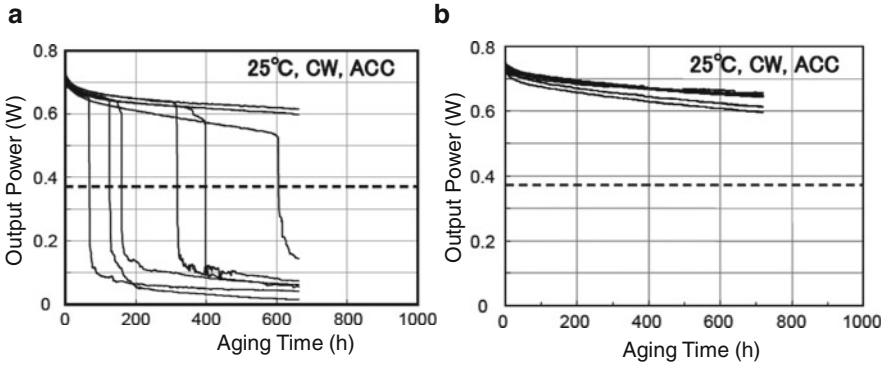



**Fig. 7.33** Aging test results of (**a**) laser diodes with conventional structure, and (**b**) with current injection-free structure

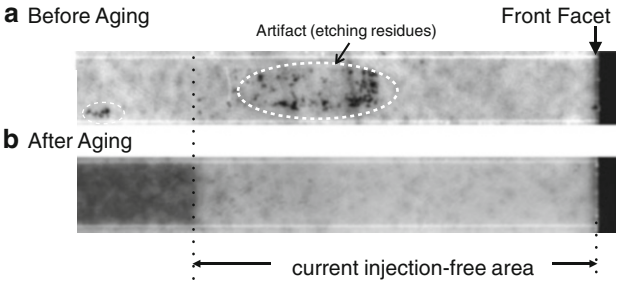



**Fig. 7.34** Fluorescence microscope images of laser chips with current injection-free structure (**a**) before and (**b**) after aging test. *Dark spotted contrasts* inside *white circles* in (**a**) is due to residues formed during the etching process for p-metal removal

Figure 7.33 shows aging curves for lasers without and with the CIF structures, respectively. The aging tests conducted in constant mode under an initial power of 0.75 W CW operation at 35°C. Sudden failures frequently occurred in lasers without the CIF region. Meanwhile, the LDs with a CIF region did not experience sudden failures beyond more than 700 h operation, and their lifetime was estimated to be more than 10,000 h.

The LDs discussed here used identical epitaxial wafers for devices both with and without the CIF region. Therefore, the densities of the structural defects were identical. The driving current densities under a power of 0.75 W were about 6.2 KA/cm$^2$ both with and without CIF region, and there were only subtle differences between them. We then performed fluorescence microscope observations of the LDs by removing the p-electrode metals by chemical etching. Figure 7.34 shows fluorescence microscope images of LDs with the CIF region

before and after the aging test. It should be noted here that the sample shown in Fig. 7.34b was still "alive" and did not degrade by sudden failure. After the aging tests, no significant reduction in emission intensity was observed in the area of the current injection-free region. On the other hand, the emission intensity was reduced in the area around the current injection region in LDs with the CIF region. Although sudden failures did not occur in the LDs with the CIF region during our experiments, their output power gradually decreased during the aging test. This gradual change was caused by darkening of the "non"-current injection-free region (i.e., the current injection region) observed in Fig. 7.34b.

Darkening in the current injection region is due to increase of nonradiative recombination centers. If we suppose that the optical field is the dominant cause of this process, then darkening observed in fluorescence microscope image could be initiated from the laser facet since the optical field near the laser facets is stronger compared to the other areas. However, dark regions were observed only in the current injection region. Therefore, degradation is thought to be caused by current injection. By inspecting a wide set of stress tests on GaN-based LDs, Meneghini et al. concluded that the degradation process is an electrothermally activated one and that the impact of the optical field is weak [62]. Their results agreed well with the results described here.

As discussed just above, the density of nonradiative recombination centers in the active regions was increased mainly by current injection in the LDs. Increase of nonradiative recombination centers leads to increase of optical absorption. Once the number of nonradiative recombination centers in the active regions exceeds a certain density, then optical absorption at/near the laser facets cannot be controlled. Then, positive feedback, such as heat generation due to nonradiative recombination, band-gap shrinkage (and/or further increase in the number of nonradiative recombination centers), further increases in optical absorption and, so on, occurs and drives the devices toward sudden failure, which is usually referred to as catastrophic optical damage (COD) degradation. Since the introduction of the CIF region suppresses the increase in nonradiative recombination centers near the laser facets, the positive feedback toward COD degradation can be curtailed. Therefore, sudden failures can be suppressed in this way.

## 7.5 Degradation Mechanisms of GaN-Based LDs

In conventional III-V-based LDs, there are classified into three major degradation modes, rapid degradation, gradual degradation, and catastrophic failures [7].

Rapid degradation occurs within 100 h of operation in the most cases of the conventional III-V-based LDs. Main cause of rapid degradation is explained by recombination-enhanced dislocation glide (REDG) or recombination-enhanced climb (REDC) at preexisting structural defects [7]. However, as described

previously, we could not observe REDG and/or REDC in the degraded GaN-based LDs, although when there contains a certain amounts of structural defects in the stripe regions.

Catastrophic failure in conventional III-V-based LDs takes place accidentally due to a current surge. Then, degradation happens catastrophically. This mode is occurred at the mirror facet due to strong optical absorption and, thus, is called catastrophic optical damage (COD). Essential cause of COD is that heat generation at the mirror facet is proliferated by increase of optical absorption, which is caused by nonradiative recombination at surface states of cleaved facets. Since this mode occurred at mirror facet, this is sometime called facet (related) degradation. In the previous section, we described facet degradation in our high-power pure-blue LDs and showed the way to suppress this type of degradation mode. Several groups have been observed facet degradation in GaN-based LDs [63–65]. In these cases, degradation is simply due to the delamination of the dielectric coatings or the formation of deposits on the dielectric coatings [63–65].

When rapid degradation and catastrophic failure are eliminated, long-term degradation (or wear-out failures) determines the maximum lifetime of the LDs. In conventional III-V-based LDs, this is called gradual degradation. Major cause of this mode is related to point defect reaction in epitaxial layers [7]. Thus, this mode is sometime called bulk degradation. In this section, we discuss gradual degradation of GaN-based LDs. We described the following results in regard to the degradation previously:

1. Degradation is not due to dislocation multiplication since no growth or formation of new structural defects was observed in the degraded LDs. However, the density of threading dislocations should be minimized as much as possible.
2. Degradation is governed by a diffusion process, since the degradation rate is almost proportional to the square root of the aging time.
3. Excess magnesium doping leads to segregation of magnesium and leads to form pyramidal defects. These can increase the operating current as a result of increase of the internal loss and/or the threshold voltage of the LDs. Thus, Mg doping has to be optimized.
4. Degradation proceeds with an increase of nonradiative recombination centers since, for example, luminescence intensity in the current injection region of pure-blue LDs decrease uniformly.
   Marona et al. also reported that threshold current variation is strongly correlated to the decrease of the subthreshold emission intensity, indicating that degradation is related to the increase of the nonradiative recombination components [21].
5. Optical field does not influence gradual degradation. If so, darkening in fluorescence microscope image could be initiated from the laser facet in the pure-blue LDs with CIF regions because the optical field near the laser facets is stronger as compared to the other region. However, such phenomenon was not observed.

6. Carrier injection is the main cause of gradual degradation. This is supported by the fact that luminescence intensity was decreased only in the current-injected region in the aged pure-blue LDs.

On the basis of these results, gradual degradation of GaN-based LDs is thought to proceed with an increase in the capture cross section of nonradiative recombination centers at the active regions, due to the diffusion of point defects and/or impurities during laser operation.

As mentioned before, reduction of threading dislocation density is very important for extending device lifetime. Suppose that gradual degradation is governed by the diffusion of point defect and/or impurities, this diffusion is enhanced by existence of threading dislocation, since the diffusion rate at dislocation cores is much faster than that in the bulk, which is known as pipe diffusion effect.

P-type doping in GaN and related materials is much more difficult than n-type doping because their behavior is very complicated. Therefore, optimization of magnesium doping should be carefully done. Otherwise, it can lead not only to worsen p-type conductivity but also to increase optical absorption as a result of Mg segregation. This deterioration affects increase of the threshold voltage of the LDs [27] and then affects device lifetime.

In addition, we also think that magnesium doping directly affects gradual degradation via point defect/impurities diffusion during device operation. In our pure-blue LDs, threading dislocations from the n-type regions is quite low ($<1 \times 10^6/cm^2$) because these devices were grown on low-dislocation-density GaN substrates. Thus, we do not have to consider effect of threading dislocations in this case. A certain amount of the multiple defects was formed in the early stage of development. Reduction of the multiple defects greatly improves the device lifetime. The multiple defects consist of the inversion domains at active regions and dislocations that generate at active regions and extend upward into p-type regions (magnesium-doped regions). Although reduction of structural defects at active regions reduces initial luminescence efficiency, structural defects at active regions do not multiply (significantly) during device operation. Therefore, significant improvement of device lifetime due to reduction of the multiple defects in our pure-blue LDs can be mainly owing to suppression of diffusion path from the p-type region to active regions via dislocations, which are accompanied with the multiple defects. In this context, we think that magnesium doping plays important role of point and/or impurities diffusion during device operation.

The deep-level transient spectroscopy (DLTS) study of degraded LDs by Meneghini et al. revealed that increase in threshold current is strongly correlated to the increase in the concentration of a deep level (DL) which is located at 0.35–0.45 eV below the conduction band [66]. This study also pointed out that this DL has similar properties to the DL which generally labeled as E2 [67–69]. Although the origin of level $E_2$ is still unknown, possible candidate of origin of $E_2$ is related residual Mg concentration in GaN [69]. $E_2$ level has also been proposed to responsible for $Mg_{Ga}$-$V_N$ complex [70]. These are consistent with above-mentioned hypothesis.

By recent photoluminescence studies of Mg-doped GaN, there are two separate accepters, called A1 and A2 [71]. A2 is stable for annealing behavior, while A1 is unstable against annealing as well as UV or electron excitation. A2 is likely to be regular Mg substitutional accepter $Mg_{Ga}$. A1 may have Mg-related complex form, which possibly has combined with hydrogen H. Unstable behavior of A1 against thermal annealing at $>500°C$ can be explained by mobile H [71]. It also notes here that H in p-type GaN acts as a donor [$H^+$] in p-type GaN [72], and this behavior is obviously unfavorable.

Another possible form of A1 may be $Mg_{Ga}$-$V_N$ complex. This type of complex is experimentally confirmed by positron annihilation spectroscopy [73]. In MOCVD growth of Mg-doped GaN layers, deficiency of hydrogen leads to increase $V_N$ concentrations. Since $V_N$ is positively charged and acts as donor, formation of $V_N$ is obviously also harmful for obtaining good p-type GaN layers.

Although A1 is not clearly identified so far, both $H^+$ and/or $V_N{}^+$-complexed Mg-related point defects can be candidates for source of gradual degradation (bulk degradation). During device operation, these Mg-related point defects are decomposed into positively charged $H^+$ and/or $V_N{}^+$ point defects. These positively charged point defects may diffuse into active regions and, then, deteriorate the device lifetime. There is some possibility for device degradation due to diffusion by interstitial Mg and/or Mg clusters. Our preliminary three-dimensional atom probe studies of degraded LDs could not detect Mg atom diffusion so far. This candidate may not be suitable at this moment. In the either case, further studies are required deep understanding the nature of degradation mechanisms.

## References

1. S. Nakamura, M. Senoh, S. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, Y. Sugimoto, H. Kiyoku, Room-temperature continuous-wave operation of InGaN multi-quantum-well-structure laser diodes with a long lifetime. Appl. Phys. Lett. **70**, 868 (1997)
2. H. Morkoc, *Nitride Semiconductors and Devices* (Springer, Berlin, 1999)
3. S.D. Lester, F.A. Ponce, M.G. Craford, D.A. Steigerwald, High dislocation densities in high efficiency GaN-based light emitting diodes. Appl. Phys. Lett. **66**, 1249 (1995)
4. T. Tojyo, T. Asano, M. Takeya, T. Hino, S. Kijima, S. Goto, S. Uchida, M. Ikeda, GaN-based high power blue-violet laser diodes. Jpn. J. Appl. Phys. **40**, 3206 (2001)
5. T. Sugahara, H. Sato, M. Hao, Y. Naoi, S. Kurai, S. Tottori, Y. Yamashita, K. Nishino, L.T. Romano, S. Sakai, Direct evidence that dislocations are non-radiative recombination centers in GaN. Jpn. J. Appl. Phys. **37**, L398 (1998)
6. S. Nagahama, N. Iwasa, M. Senoh, T. Matsushita, Y. Sugimoto, H. Kiyoku, T. Kozaki, M. Sano, H. Matsumura, H. Umemoto, K. Chocho, T. Mukai, High-power and long-lifetime InGaN multi-quantum-well laser diodes grown on low-dislocation-density GaN substrates. Jpn. J. Appl. Phys. **39**, L647–650 (2000)
7. O. Ueda, *Reliability and Degradation of III-V Optical Devices* (Artech House, Boston, 1996)
8. T.S. Zheleva, S.A. Smith, D.B. Thomson, K.J. Linthicum, P. Rajagopal, R.F. Davis, Pendeo-epitaxy: a new approach for lateral growth of gallium nitride films. J. Electron. Mater. **28**, L6–L8 (1999)

9. K.J. Linthicum, T. Gehrke, D. Thomson, E. Carlson, P. Rajagopal, T. Smith, D. Batchelor, R.F. Davis, Pendeo epitaxy of gallium nitride thin films. Appl. Phys. Lett. **75**, 196 (1999)
10. T. Hino, S. Tomiya, T. Miyajima, K. Yanashima, S. Hashimoto, M. Ikeda, Characterization of threading dislocations in GaN epitaxial layers. Appl. Phys. Lett. **76**, 3421 (2000)
11. A. Sakai, H. Sunakawa, A. Usui, Defect structure in selectively grown GaN films with low threading dislocation density. Appl. Phys. Lett. **71**, 2259 (1999)
12. K. Suzuki, S. Takeuchi, Electron microscopy of dislocations introduced into GaN by plastic deformation. Philos. Mag. Lett. **79**, 423 (1999)
13. S.J. Rosner, G. Girolami, H. Marchand, P.T. Fini, J.P. Ibbetson, L. Zhao, S. Keller, U.K. Mishra, S.P. DenBaars, J.S. Speck, Cathodoluminescence mapping of epitaxial lateral overgrown in gallium nitride. Appl. Phys. Lett. **74**, 2035 (1999)
14. P. Hacke, K. Domen, A. Kuramata, T. Tanahashi, O. Ueda, Origin of the nonradiative <11–20> line defect in lateral epitaxy-grown GaN on SiC substrates. Appl. Phys. Lett. **76**, 2547 (2000)
15. P. Hirsh, A. Howie, R.B. Nicholson, D.W. Pashley, W.J. Whelan, *Electron Microscopy of Thin Crystals*, 2nd edn. (R.E. Kreiger Publishing Co., Huntington, 1977)
16. S. Tomiya, E. Morita, M. Ukita, H. Okuyama, S. Itoh, K. Nakano, A. Ishibashi, Structural study of defects induced current injection to II-VI blue light emitter. Appl. Phys. Lett. **66**, 1208 (1995); S. Tomiya, H. Noguchi, Y. Sanaka, T. Hino, S. Taniguchi, A. Ishibashi, Microstructural study of optically degraded ZnCdSe quantum wells. Appl. Phys. Lett. **74**, 1824 (1999); S. Tomiya, H. Okuyama, A. Ishibashi, Relation between interface morphology and recombination-enhanced defect reaction phenomena in II–VI light emitting devices. Appl. Surf. Sci. **159–160**, 243 (2000)
17. M. Peach, J.S. Koehler, The forces exerted on dislocations and the stress fields produced by them. Phys. Rev. **80**, 436 (1950)
18. Sugiura Lisa, Dislocation motion in GaN light-emitting devices and its effect on device lifetime. J. Appl. Phys. **81**, 1633 (1997)
19. K. Maeda, K. Suzuki, M. Ichihara, S. Nishiguchi, K. Ono, Y. Mera, S. Takeuchi, Electronically induced dislocation glide motion in hexagonal GaN single crystals. Phys. B **273–274**, 134 (1999)
20. S. Goto, K. Tamamura, O. Matsumoto, T. Tojyo, T. Sasaki, Y. Yabuki, K. Naganuma, T. Asatsuma, S. Tomiya, S. Uchida, M. Ikeda, presented at the *Late News Abstract, 2003 Electronic Materials Conference*, Santa Barbara, CA, 2002, V9
21. L. Marona, P. Wisniewski, P. Prystawko, I. Grzegory, T. Suski, S. Porowski, P. Perlin, R. Czernecki, M. Leszczynski, Degradation mechanisms in InGaN laser diodes grown on bulk GaN crystals. Appl. Phys. Lett. **88**, 201 (2006)
22. H. Amano, M. Kito, K. Hiramatsu, I. Akasaki, P-type conduction in Mg-doped GaN treated with low-energy electron beam irradiation (LEEBI). Jpn. J. Appl. Phys. **28**, L2112 (1989)
23. S. Nakamura, T. Mukai, M. Senoh, High-power GaN P-N junction blue-light-emitting diodes. Jpn. J. Appl. Phys. **30**, L1998 (1991)
24. S. Tomiya, T. Hino, S. Goto, M. Takeya, M. Ikeda, Dislocation related issues in the degradation of GaN-based laser diodes. IEEE J. Quantum Electron. **10**, 1277 (2004)
25. Z. Liliental-Weber, M. Benamara, J. Washburn, I. Grzegory, S. Porowski, D.J.H. Lambert, C.J. Eiting, R.D. Dupuis, Mg-doped GaN: similar defects in bulk crystals and layers grown on $Al_2O_3$ by metal-organic chemical-vapor deposition. Appl. Phys. Lett. **75**, 4159 (1999)
26. P. Vennéguès, M. Benaissa, B. Beaumont, E. Feltin, P. De Miery, S. Dalmasso, M. Leroux, P. Gilbart, Mg-doped GaN: similar defects in bulk crystals and layers grown on Al2O3 by metal–organic chemical-vapor deposition. Appl. Phys. Lett. **77**, 880 (2000)
27. M. Hansen, L.F. Chen, S.H. Lim, S.P. DenBaars, J.S. Speck, Mg-rich precipitates in the p-type doping of InGaN-based laser diodes. Appl. Phys. Lett. **80**, 2469 (2002)
28. S. Tomiya, S. Goto, M. Takeya, M. Ikeda, Structural defects in Mg-doped GaN and AlGaN grown by MOCVD. Mater. Res. Soc. Symp. Proc. **743**, L12.7.1 (2003)

29. L.T. Romano, J.E. Northrup, M.A. O'Keefe, Inversion domains in GaN grown on sapphire. Appl. Phys. Lett. **69**, 2394 (1996)
30. J.E. Northrup, Magnesium incorporation at (0001) inversion domain boundaries in GaN. Appl. Phys. Lett. **82**, 2278 (2003)
31. A.J. Ptak, T.H. Myers, L.T. Romano, C.G. Van de Walle, J.E. Northrup, Magnesium incorporation in GaN grown by molecular-beam epitaxy. Appl. Phys. Lett. **78**, 285 (2001)
32. T.S. Cheng, S.V. Novikov, C.T. Foxon, J.W. Orton, Mechanisms of magnesium incorporation into GaN layers grown by molecular beam epitaxy. Solid State Commun. **109**, 439 (1999)
33. V. Ramachandran, R.M. Feenstra, J.E. Northrup, D.W. Greve, MRS internet. J. Nitride Semicond. Res. **5S1**, W3.65 (2000)
34. L.T. Romano, J.E. Northrup, A.J. Ptak, T.H. Myers, Faceted inversion domain boundary in GaN films doped with Mg. Appl. Phys. Lett. **77**, 2479 (2000)
35. B. Beaumont, S. Haffouz, P. Gilbart, Magnesium induced changes in the selective growth of GaN by metalorganic vapor phase epitaxy. Appl. Phys. Lett. **72**, 921 (1998)
36. H. Obloh, K.H. Bachem, U. Kaufmann, M. Kunzer, M. Maier, A. Ramakrishnan, P. Schlotter, Self-compensation in Mg doped p-type GaN grown by MOCVD. J. Cryst. Growth **195**, 270 (1998)
37. I. Akasaki, H. Amano, Widegap column-III nitride semiconductors for UV/blue light emitting devices. J. Electrochem. Soc. **141**, 2266 (1994)
38. W. Götz, N.M. Johnson, J. Walker, D.P. Bour, R.A. Street, Activation of acceptors in Mg-doped GaN grown by metalorganic chemical vapor deposition. Appl. Phys. Lett. **68**, 667 (1996)
39. P. Kozodoy, H. Xing, S.P. DenBaars, U.K. Mishra, A. Saxler, R. Perrin, S. Elhamri, W.C. Mitchel, Activation of acceptors in Mg-doped GaN grown by metalorganic chemical vapor deposition. J. Appl. Phys. **87**, 1832 (2000)
40. S. Nakamura, N. Iwasa, M. Senoh, T. Mukai, Hole compensation mechanism of p-type GaN films. Jpn. J. Appl. Phys. **31, Part 1**, 1258 (1992)
41. J.A. Van Vechten, J.D. Zook, R.D. Hornig, B. Goldenberg, Defeating compensation in wide gap semiconductors by growing in H that is removed by low temperature de-ionizing radiation. Jpn. J. Appl. Phys. **31**, 3662 (1992)
42. J. Neugebauer, C.G. Van de Walle, Role of hydrogen in doping of GaN. Appl. Phys. Lett. **68**, 1829 (1996)
43. P. Kiesel, F. Renner, M. Kneissl, N.M. Johnson, G.H. Dohler, Electroabsorption spectroscopy – direct determination of the strong piezoelectric field in InGaN/GaN heterostructure diodes. Phys. Stat. Sol. (a) **188**, 131 (2001)
44. M. Kuramoto, C. Sasaoka, N. Futagawa, M. Nido, A.A. Yamaguchi, Reduction of internal loss and threshold current in a laser diode with a ridge by selective re-growth (RiS-LD). Phys. Stat. Sol. (a) **192**, 329 (2002)
45. I. Ho, G.B. Stringfellow, Solid phase immiscibility in GaInN. Appl. Phys. Lett. **69**, 2701 (1996)
46. L.T. Romano, M.D. McCluskey, C.G. Van de Walle, J.E. Northrup, D.P. Bour, T. Suski, J. Jun, Phase separation in InGaN multiple quantum wells annealed at high nitrogen pressures. Appl. Phys. Lett. **75**, 3950 (1999)
47. H.W. Shim, R.J. Choi, S.M. Jeong, L. Van Vinh, C.-H. Hong, E.-K. Suh, H.J. Lee, Y.-W. Kim, Y.G. Hwang, Influence of the quantum-well shape on the light emission characteristics of InGaN/GaN quantum-well structures and light-emitting diodes. Appl. Phys. Lett. **81**, 3552 (2002)
48. C.J. Sun, M.Z. Anwar, Q. Chen, J.W. Yang, M.A. Khan, M.S. Shur, A.D. Bykhovski, Z. Liliental-Weber, C. Kisielowski, M. Smith, J.Y. Lin, H.X. Xiang, Quantum shift of band-edge stimulated emission in InGaN-GaN multiple quantum well light-emitting diodes. Appl. Phys. Lett. **70**, 2978 (1997)
49. Y. Chen, T. Takeuchi, H. Amano, I. Akasaki, N. Yamada, Y. Kaneko, S.Y. Wang, Pit formation in GaInN quantum wells. Appl. Phys. Lett. **72**, 710 (1998)

50. X.H. Wu, C.R. Elsass, A. Abare, M. Mack, S. Keller, P.M. Petroff, S.P. DenBaars, J.S. Speck, S.J. Rosner, Structural origin of V-defects and correlation with localized excitonic centers in InGaN/GaN multiple quantum wells. Appl. Phys. Lett. **72**, 692 (1998)
51. I.H. Kim, H.S. Park, Y.J. Park, T. Kim, Formation of V-shaped pits in InGaN/GaN multiquantum wells and bulk InGaN films. Appl. Phys. Lett. **73**, 1634 (1998)
52. N. Sharma, P. Thomas, D. Tricker, C. Humphreys, Chemical mapping and formation of V-defects in InGaN multiple quantum wells. Appl. Phys. Lett. **77**, 1274 (2000)
53. S. Tomiya, K. Kanitani, T. Tanaka, T. Ohkubo, K. Hono, Atomic scale characterization of GaInN/GaN multiple quantum wells in V-shaped pits. Appl. Phys. Lett. **98**, 181904 (2011)
54. O. Goto, S. Tomiya, Y. Hoshina, T. Tanaka, M. Ohta, Y. Ohizumi, Y. Yabuki, K. Funato, M. Ikeda, High power pure-blue semiconductor lasers. Proc. SPIE **6485**, p. 64850Z (2007)
55. M. Ohta, Y. Ohizumi, Y. Hoshina, T. Tanaka, Y. Yabuki, K. Funato, S. Tomiya, S. Goto, M. Ikeda, High-power pure blue laser diodes. Phys. Stat. Sol. (a) **204**, 2068 (2007)
56. S. Tomiya, O. Goto, Y. Hoshina, T. Tanaka, M. Ikeda, Multiple defects in GaInN multiple quantum wells grown on ELO GaN layers and on GaN substrates. Phys. Stat. Sol. (c) **3**, 1779 (2006)
57. O. Goto, Y. Ohizumi, M. Shoji, T. Tanaka, Y. Hoshina, M. Ohta, Y. Yabuki, S. Tomiya, and M. Ikeda, Suppression of COD in pure-blue laser diodes with current injection-free region near the laser facet (in Japanese), IEICE Technical Report ICICE-107, 2007, pp. 35–37
58. Y. Ueno, H. Fujii, K. Kobayashi, K. Endo, A. Gomyo, K. Hara, S. Kawata, T. Yuasa, T. Suzuki, Novel window-structure AlGaInP visible-light laser diodes with non-absorbing facets fabricated by utilizing GaInP natural superlattice disordering. Jpn. J. Appl. Phys. **29**, L1666 (1990)
59. H. Hamada, M. Shono, S. Honda, R. Hiroyama, K. Matsukawa, K. Yodoshi, T. Yamaguchi, High-power operation of 630 nm-band transverse-mode stabilized AlGaInP laser diodes with current-blocking region near facets. Electron. Lett. **27**, 661 (1991)
60. Y. Xu, Y. Li, Q. Gan, Q. Cao, G. Song, L. Guo, L. Chen, High-power AlGaInP laser diodes with current injection free region near the laser facet. Opt. Eng. **45**, 034205–1 (2006)
61. M. Meneghnini, G. Meneghesso, N. Trivellin, E. Zanoni, K. Orita, M. Yuri, D. Ueda, Extensive analysis of the degradation of blu-ray laser diodes. IEEE Electron Device Lett. **29**, 578 (2008)
62. T. Schoedl, U.T. Schwarz, V. Kummler, M. Furitsch, A. Leber, A. Miler, A. Lell, V. Harle, Facet degradation of GaN heterostructure laser diodes. J. Appl. Phys. **97**, 123102 (2005)
63. C.C. Kim, Y. Choi, Y.H. Jang, M.K. Kang, M. Joo, M.S. Noh, Proc. SPIE **6894**, p. 689400 (2008)
64. L. Marona, P. Wiśniewski, L. Leszczyński, I. Grzegory, T. Suski, S. Porowski, R. Czernecki, A. Czerwinski, M. Pluska, J. Ratajczak, P. Perlin, Why InGaN laser-diode degradation is accompanied by the improvement of its thermal stability. Proc. SPIE **6894** 689400 (2008)
65. M. Meneghini, C. de Santi, N. Trivellin, K. Orita, S. Takigawa, T. Tanaka, D. Ueda, G. Meneghesso, E. Zanoni, Investigation of the deep level involved in InGaN laser degradation by deep level transient spectroscopy. Appl. Phys. Lett. **99**, 093506 (2011)
66. H.K. Cho, C.S. Kim, C.-H. Hong, Electron capture behaviors of deep level traps in unintentionally doped and intentionally doped *n*-type GaN. J. Appl. Phys. **94**, 1485 (2003)
67. H.K. Cho, F.A. Khan, I. Adesida, Z.-Q. Fang, D.C. Look, Deep level characteristics in n-GaN with inductively coupled plasma damage. J. Phys. D: Appl. Phys. **41**, 155314 (2008)
68. A. Hierro, S.A. Ringel, M. Hansen, J.S. Speck, U.K. Mishra, S.P. DenBaars, Hydrogen passivation of deep levels in *n*–GaN. Appl. Phys. Lett. **77**, 1499 (2000)
69. U. Kaufmann, M. Kunzer, M. Maier, H. Obloh, A. Ramakrishan, B. Santic, P. Schlotter, Nature of the 2.8 eV photoluminescence band in Mg doped GaN. Appl. Phys. Lett. **72**, 1326 (1998)
70. B. Monemar, P.P. Paskov, G. Pozina, C. Hemmingsson, J.P. Bergman, T. Kawashima, H. Amano, I. Akasaki, T. Paskova, S. Figge, D. Hommel, A. Usui, Evidence for two Mg related acceptors in GaN. Phys. Rev. Lett. **102**, 235501 (2009)

71. J. Neugebauer, C.G. Van de Walle, Hydrogen in GaN: novel aspects of a common impurity. Phys. Rev. Lett. **75**, 4452 (1995)
72. S. Hautakangas, J. Oila, M. Alatalo, K. Saarinen, L. Liszkay, D. Seghier, H.P. Gislason, Vacancy defects as compensating centers in Mg-doped GaN. Phys. Rev. Lett. **90**, 137402 (2003)
73. J. Neugebauer, C.G. Van de Walle, Atomic geometry and electronic structure of native defects in GaN. Phys. Rev. B. **50**, 8067 (1994)

# Chapter 8
# InGaN Laser Diode Degradation

**Piotr Perlin and Łucja Marona**

**Abstract** We discuss the current knowledge of degradation processes in InGaN laser diodes. It is quite surprising that after quite a few years of intensive studies, there is still no clear picture of physical mechanisms lying behind these aging processes. First of all, in contrast to GaAs counterparts, the nitride laser degradation seems to be independent from extended defects movement and multiplication, is uniform across the device surface, and is rather unrelated with optical phenomena.

Involvement of point defects in this process is very tempting but not yet sufficiently experimentally supported hypothesis. Bipolar GaN devices will very likely teach us new things about the physics and technology of wideband gap devices – a truly new thing in the optoelectronics.

## 8.1 Introduction

Since the first appearance of nitride optoelectronic devices, the problem of their reliability and understanding the mechanism of degradation processes turned out to be crucial. The early studies that focused on light-emitting diodes (LEDs) revealed that the degradation of these devices is dominated by the packaging issues (epoxy darkening and carbonization) [1, 2]. Packaging improvement enabled manufacturing blue-green LEDs operating at current densities close to 20 A/cm$^2$ and of the lifetime of 50–100 kh, the value satisfying for the most of the manufacturers.

P. Perlin (✉)
Institute of High Pressure Physics, Semiconductors Laboratory, Polish Academy of Sciences, Sokołowska 29/37, 01-142 Warsaw, Poland

TopGaN Ltd., Warsaw, Poland
e-mail: piotr@unipress.waw.pl

Ł. Marona
Institute of High Pressure Physics, Semiconductors Laboratory, Polish Academy of Sciences, Sokołowska 29/37, 01-142 Warsaw, Poland

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_8,
© Springer Science+Business Media New York 2013

In contrast to LEDs, nitride laser diodes (LDs) reliability issues were found to be much more complex. The early laser diodes were grown on sapphire substrates which are not lattice matched to GaN. Therefore, these devices were characterized by a very large density of defects of the order of $10^9$ to $10^{10}$ dislocation per square centimeter. It was not too surprising that the lifetime of the first InGaN laser diodes was very short. For example, the first CW operated laser demonstrated by Nakamura lived only 1 s [3]. An obvious approach was to increase the lifetime of laser diodes by reducing the dislocation density. Epitaxial lateral overgrowth (ELOG) technology allowed to reduce the dislocation density approximately by two orders of magnitude [4]. In parallel, availability of freestanding GaN substrates (dislocation densities $<10^4$ $cm^2$) was gradually improving. Using of both, ELOG technique and GaN substrates, allows to increase lifetime of nitride laser diodes to desirable 10,000 h [4]. Though the described progress in the longevity of nitride laser diodes was impressive, the mass production of these devices started with a significant delay (6–7 years), which may mean that the production yield and reliability problems were not completely solved at that time. At present, though a limited number of companies apparently possess a long-living nitride laser diodes know-how, our understanding of the degradation processes in this important material system is still very limited.

## 8.2 Structure of InGaN Laser Diode

Nitride-based laser diode structures are grown by two epitaxial method: mostly by metal-organic vapor phase epitaxy (MOVPE) [4] but also by molecular beam epitaxy (MBE) [5] on bulk GaN substrates. There are only few GaN substrates suppliers, and they use different fabrication technologies. Sumitomo Electrics uses "DEEP" method [6], Ammono Ltd. uses ammonothermal growth method [7], TopGaN Ltd. is specializing in high-pressure gallium nitride growth [8], and Furukawa, Hitachi Cables, Mitsubishi, Lumilog, and Kyma use various modifications of hydride vapor phase epitaxy (HVPE) to manufacture GaN substrates. Gallium nitride substrates have many quality disadvantages like crystal bowing, dislocations, and too small available sizes. Substrate problems influence epitaxial growth often leading to cracking, defects propagation, and poor uniformity of growth. Ability to grow good quality nitride layers is a crucial issue for obtaining satisfying parameters of the device.

The conventional epistructure of InGaN laser diode is a separate confinement heterostructure (SCH) consisting of 1–3 InGaN quantum wells [4]; see Fig. 8.1. Usually on top of the active layers, a *p*-type AlGaN layer, acting as a electron blocker, is placed. In minority cases, EBL can be moved up and put between the upper waveguiding layer and the upper cladding layer. Active region is covered by waveguides, usually formed by GaN layers of the thickness of approximately 100 nm. Waveguiding layers may be Si- or Mg-doped, but they may be also left

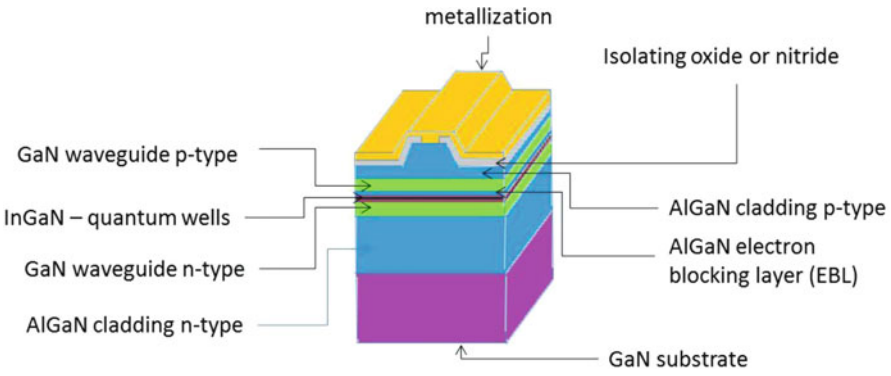



**Fig. 8.1** Scheme of nitride laser diode epitaxial structure

undoped. Whole structure is sandwiched by *n*- and *p*-doped AlGaN claddings to avoid optical mode leakage.

## 8.3 Packaging of InGaN Laser Diodes

Nitride laser diodes are typically mounted into 5.6 mm TO cans (TO18 package) (Fig. 8.2) or quite rarely into 9 mm cans. In most cases, the laser diodes are assembled in a p-up configuration. Depending on the manufacturer, laser chips are mounted on the top of a highly thermally conducting submount (heat spreader) or directly on the stem of the package. Very high thermal conductivity of the GaN allows for heat-spreader elimination.

The important aspect of nitride laser diodes is hermeticity of the package. Since the high photon energy of the light emitted from a laser enhances the chemical reactions on the mirrors, it is important to keep moisture and hydrocarbons out of the operating devices. This issue will be discussed in more detail in the next chapters.

## 8.4 Symptoms of the Degradation

Degradation of the semiconductor laser diodes can be divided into three groups depending on the device region where a damage occurs as shown in the Table 8.1 below.

In the next sections, problems with facet, active layer, and contact degradation will be discussed in more details.

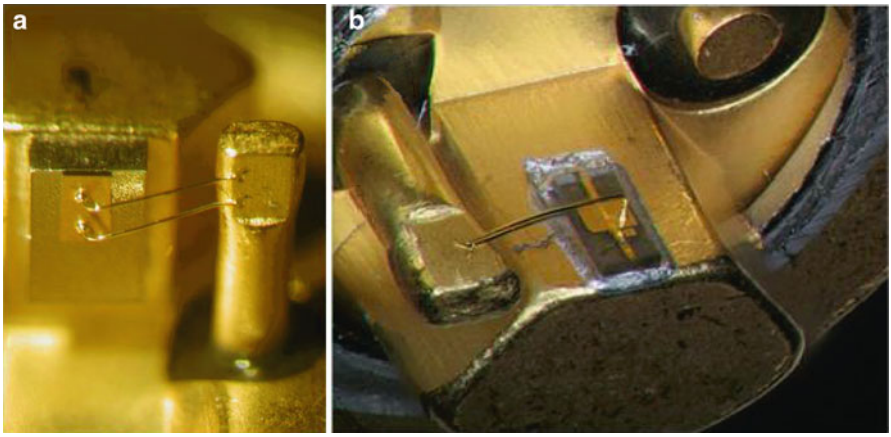



**Fig. 8.2** Packaged nitride laser diode: (**a**) nitride laser diode installed on top of diamond heat spreader in TO 18 package, (**b**) laser chip soldered directly to the stem of the package (A courtesy of TopGaN Ltd.)

**Table 8.1** Degradation mechanisms in nitride laser diodes

| Region of a device | Phenomena | Factor |
|---|---|---|
| Laser facet | Catastrophic mirror damage (COD), oxidation, different chemical reactions | Humidity, light, atmospheres, surface states, temperature |
| Active layer | Nonradiative recombination centers | Current flow, temperature |
| Contacts | Metal diffusion | Current flow, temperature |

### *8.4.1 Facet Degradation*

Quite frequently, the first symptom of laser diodes degradation occurs at their facets. Mirror degradation is a phenomenon well known in case of arsenide devices [9]. We can roughly divide it into two classes: first one is a catastrophic optical mirror damage (COMD), a process which manifests through the sudden increase of the mirror's temperature leading to melting of the facets. Increase of the temperature is caused by an optical absorption in the near-facet area of a device and a positive feedback loop between the increase of device temperature and the increase of the optical absorption. The threshold of COMD was determined to be around 40–57 MW/cm$^2$ [10, 11] for InGaN laser diodes. These values are an order of magnitude higher than for GaAs counterparts, thus showing the potential of nitride lasers for high optical power emission. However, a mechanism of COMD is far from being well understood in the nitride system.

The second effect related with mirror degradation is the formation of carbon deposits on the surface of the output mirror. It was discovered that under the operation in the oxygen-free atmosphere, a complicated process of photo-assisted hydrocarbons decomposition occurs leading to the formation of carbon deposit and fast laser diode

degradation. Similar processes were observed in GaAs-based lasers and were called package-induced failure (PIF). This term was introduced by Julia Sharps to name the phenomenon of very fast degradation of 980 nm, high power laser diode operated in hermetic packages under neutral gas atmosphere [12]. The fingerprint of this degradation mode was the formation of carbon deposit on the mirrors of the laser during its operation and high, positive, susceptibility to the presence of oxygen in the laser package [13]. It was shown that the reaction is related to the hydrocarbon photoinduced decomposition and the source of the hydrocarbons may be the package atmosphere or the package material [13]. Quite the same observations have been made recently by groups studying the facet degradation of nitride laser diode. Schoedl et al. [14] found that operating laser diode in dry nitrogen atmosphere leads to the formation of a thick deposit. This deposit can be altered or even remove by adding oxygen to the atmosphere. The authors also claim that the presence of a water vapor speeds up the mirror oxidation. Observation of the carbon deposit formation was also reported by Kim et al. [15]. They limited this process by an additional cleaning and careful sealing of the package. Our experiments performed on laser diodes in argon or nitrogen atmospheres with or without an addition of oxygen show the following:

1. The fastest degradation occurs in dry nitrogen atmosphere.
2. The addition of the oxygen slows down the degradation but does not eliminate it completely until the proper cleaning (oxygen plasma asher and UV ozone cleaning) of the structure is performed.
3. The facet degradation involves the formation of carbon deposit not only in the region of the laser stripe but also around the chip at the level of *p-n* junction.
4. The progress of degradation is irregular; the oscillation of the threshold current may occur, related to the deposit growth and delamination.
5. The source of hydrocarbons is most likely the electroplated gold on the laser chips but also very likely gold layers on the commercial package material.

The carbon deposit formed on the facet of our laser diode is presented in Fig. 8.3. Judging from our experience, the carbon deposit formation may limit the lifetime of devices from few tens up to hundreds of hours. The chemistry of this process is still not well understood, and the resemblance to classical 980 nm PIF mechanism has to be investigated.

### *8.4.2 Active Layer Degradation*

Elimination of facet degradation reveals much more fundamental process occurring in the whole volume of the active area of a nitride laser diode. As a first step, let us analyze symptoms observed during the aging of a device. Degradation in InGaN-based laser diodes occurs the most frequently through the increase of the threshold current like it is shown in Fig. 8.4 [16, 17]. What can be easily seen is that the strong increase of the threshold current is not accompanied by the proportional change in the slope efficiency.

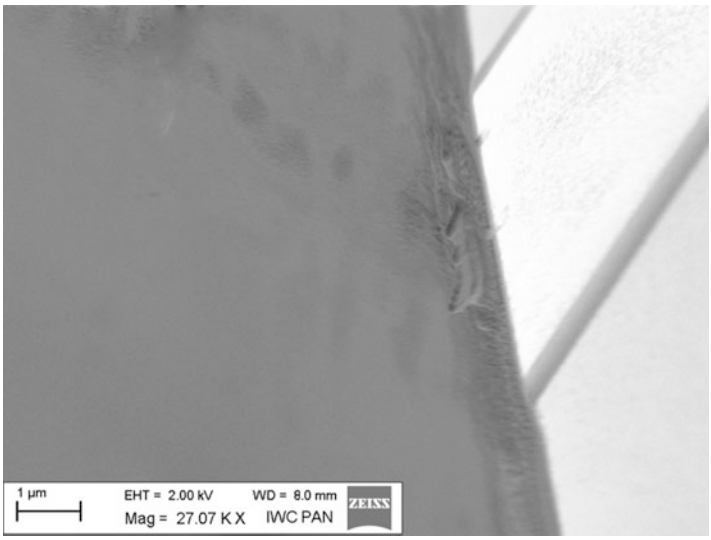



**Fig. 8.3** Carbon deposit protruding from the facet of an InGaN laser diode (Courtesy of TopGaN Ltd.)

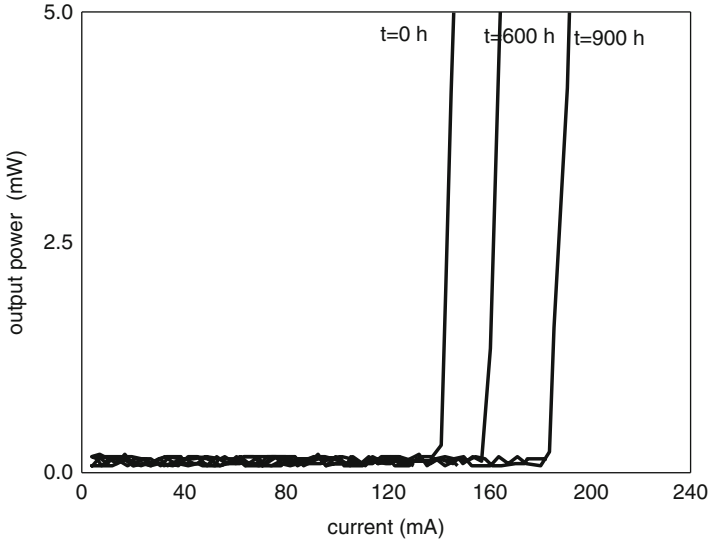



**Fig. 8.4** The evolution of light-current curves with the degradation time

Very slow evolution of slope efficiency of the degraded devices seems to be quite reproducible feature of nitride LDs [16, 17]. This behavior agrees with the simple model of the increase of the nonradiative recombination within the active layer [16]. According to this model, constant slope efficiency implies constant

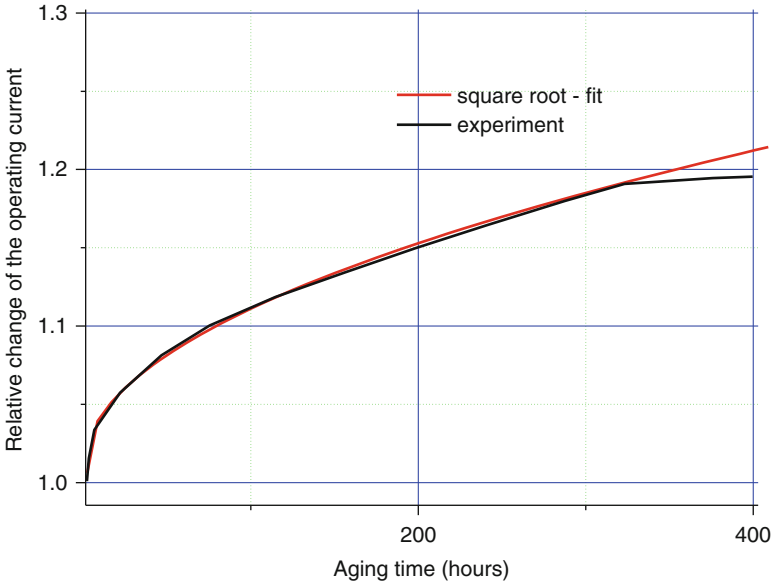



**Fig. 8.5** The increase of operation current as a function of time in comparison with a square root function

injection efficiency and no increase of optical losses (in nitride devices cavity losses are dominated by the optical absorption on magnesium states). The slope efficiency (SE) can be expressed by:

$$\mathrm{SE} = \eta_\mathrm{j} \frac{\alpha_\mathrm{m}}{\alpha_\mathrm{i} + \alpha_\mathrm{m}}, \tag{8.1}$$

where $\eta_\mathrm{j}$ is injection efficiency, $\alpha_\mathrm{m}$ mirror losses, and $\alpha_\mathrm{i}$ internal losses.

The threshold current was reported [16–18] to increase with the square root of time like it is shown in Fig. 8.5.

The square root dependence was quite frequently associated with diffusion according to the Einstein-Smoluchowski formula:

$$\Delta x = \sqrt{Dt}, \tag{8.2}$$

where $\Delta x$ is change of the stoichiometry along one direction, $D$ is diffusion coefficient, and $t$ is time. However, the question arises: what actually diffuses? Though the interdiffusion of magnesium was a tempting candidate for the degradation mechanism, there is only one very indirect evidence provided by secondary ion mass spectroscopy (SIMS)-smeared magnesium profile measured at large area light-emitting diode [19]. The lack of SIMS profiles measured at the laser diodes tested under standard conditions is related to the difficulty in performing

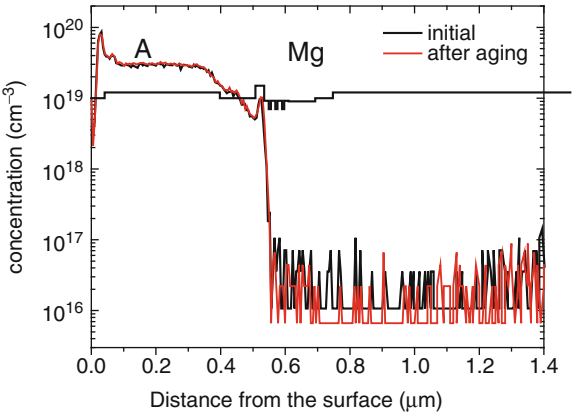



**Fig. 8.6** Magnesium SIMS profiles measured over the aged area and virgin area of the heavily stressed laser diode (10,000 h of operation)

SIMS measurements over very tiny area of the laser stripe. However, we recently managed to accurately measure SIMS profiles of many dopants (Mg, H, Si) [20] over a stressed area of a laser diode (Fig. 8.6).

In contrast to previously reported data, our new results show no magnesium nor hydrogen profile changes with the degradation time. Thus, we can safely say that these two elements are most likely not involved directly in the relevant diffusion process. This of course does not exclude local rearrangements of atoms around defects centers, complexes, etc. Additionally, there is still open question concerning possibility of native defects diffusion.

Though the square root dependence of degradation dominates in some cases, also linear behavior of the current versus time is observed like the one shown in Fig. 8.7. At the present moment, it is not clear why in some cases linear degradation mode dominates and if it means that the mechanism of the degradation is different.

Degradation of laser diodes is often characterized by a value of so-called activation energy ($E_A$). This value can be estimated by using Arrhenius formula:

$$D = C \exp\left(\frac{E_A}{kT}\right), \tag{8.3}$$

where $D$ is degradation rate, $C$ is a constant, $E_A$ activation energy, $T$ temperature, and $k$ Boltzmann constant. Degradation rate can be defined in various ways, as, for example, change of the threshold current in the unit of time or a number of accumulated failures as a function of time. Arrhenius equation shows that degradation is thermally activated process. The values of the activation energy for nitrides are reported to be between 0.3 [11, 16] and 0.5 eV [21].

Finally, the additional argument that the degradation is related to the active area of the device (volume) and is nor related with complicated laser diode topology was provided by Meneghini et al. [17]. They showed that if instead of laser diode you fabricate a completely planar light-emitting diode (LED) of the identical epistructure to the laser, then it degrades in the identical manner scaled only by

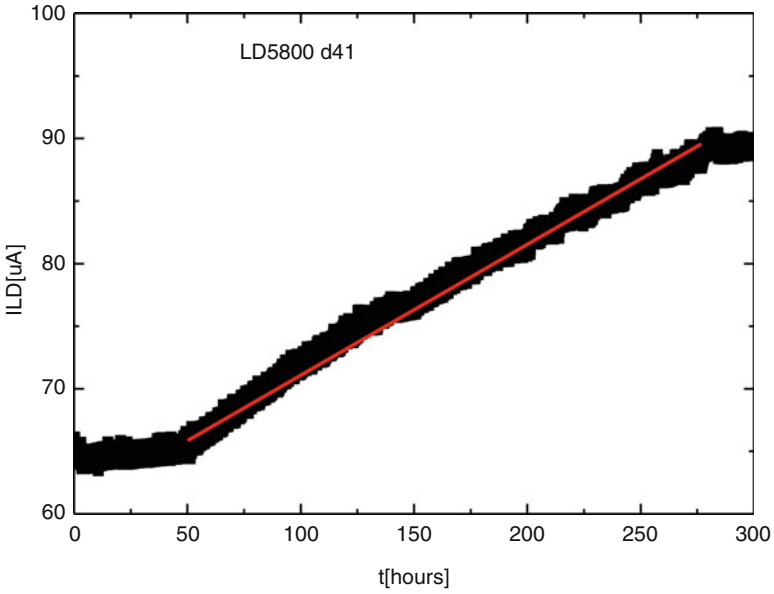



**Fig. 8.7** An example of a close to linear degradation behavior of nitride laser diode. Vertical axis shows a device operating current measured at constant optical power of 10 mW

the value of the operating current density (much lower for LEDs). The authors establish also a linear relation between the aging rate and the current density for all tested devices.

#### 8.4.2.1 Optical Gain in Stressed Laser Diodes

Optical gain is a decisive parameter which defines the laser threshold. Optical gain is a function of the carrier density in the active area of the laser diode and thus depends on the recombination kinetics. One of the first studies of the optical gain in the stressed laser diode was performed by Goddard [22] from Xerox Palo Alto Group. They demonstrate that the laser diode degradation (an increase of the threshold current) is accompanied by a gradual decrease of the optical gain. We performed systematic study of the gain spectra taken during laser diode aging. On Fig. 8.8, we can see the results of our study. The gain spectra were recorded using a Hakki-Paoli method.

We see that after degradation a magnitude of the gain peak is decreased when measured under constant current conditions. However, if current is increased again, we can see that gain curve returns to its original strength and form. This result can be interpreted in such a way that under same operating current, due to an increase nonradiative recombination or carrier escape from the quantum wells, the number

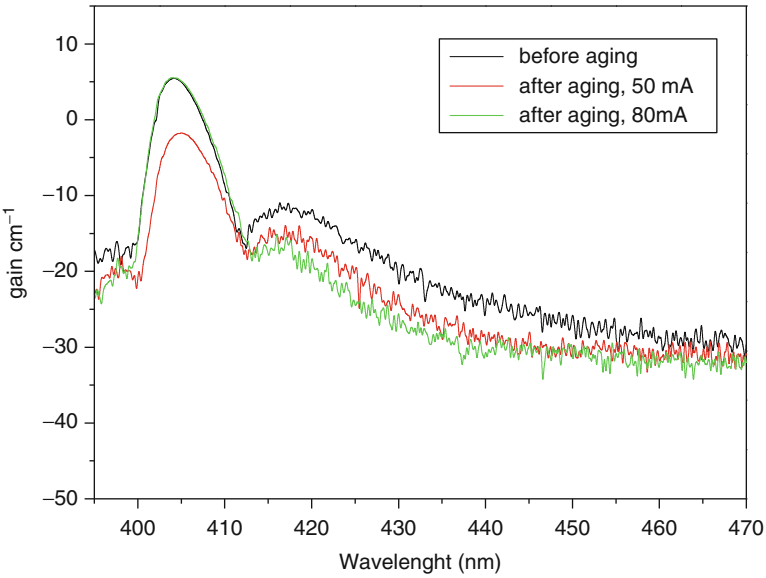



**Fig. 8.8** Gain spectra in the aged device. *Black curve* shows an initial gain peak, a *red curve* presents the spectrum measured at the same current, and a *green curve* shows the gain for the device with increased current to compensate the gain loss

of carriers in the InGaN quantum wells is lower which lowers the gain value correspondingly. The stability of the gain spectral shape confirms the material stability excluding, for example, the quantum well interdiffusion.

### 8.4.2.2 Nonradiative Lifetime Measurements

As we mentioned before, an expected reason for laser diode degradation is generation of nonradiative recombination centers within the active layer of the stressed device. This is why it is so important to determine the nonradiative recombination time (or rate) during the aging process. Direct measurements of the recombination rates on completely processed and packaged devices are hard to perform. Therefore, an important approach to determine these parameters is to use an analysis of subthreshold light-current characteristics of laser diodes. The starting point for such an analysis is always the so-called A-B-C rate equation [23]:

$$\frac{\eta_j I}{qV} = AN + BN^2 + CN^3, \tag{8.4}$$

where $\eta_j$ is an injection efficiency, $I$ is a current, $q$ is a charge on electron and $V$ is an active area volume, $A$ is nonradiative recombination rate coefficient, $B$ is

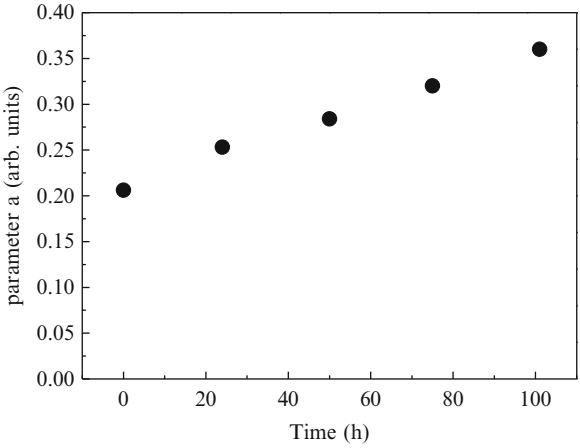



**Fig. 8.9** Depending of nonradiative recombination on the aging time of the device

bimolecular radiative rate coefficient, and $C$ is typically related with Auger recombination or similar mechanism.

Optical power is related with the recombination coefficient through the following equation:

$$P = kBN^2. \tag{8.5}$$

Using the above two formulas, relation between optical power and current can be derived:

$$\frac{I}{\sqrt{P}} = a + b\sqrt{P}, \tag{8.6}$$

where parameters $a$ and $b$ are defined by:

$$a = \frac{qV}{\eta_j}\frac{A}{\sqrt{kB}}, \tag{8.7}$$

$$b = \frac{qV}{\eta_j}\frac{1}{k}. \tag{8.8}$$

The application of the Eq. 8.6 for the determination of the nonradiative lifetime in nitrides laser diode was demonstrated by Ryu et al. [24].

On Fig. 8.9, we show the measured $a$ parameter which is basically proportional to nonradiative recombination rate $A$ (we assume here that bimolecular radiative recombination $B$ is constant). As it is visible, $a$ parameter steadily grows suggesting a steady increase of nonradiative recombination rate.

The increase of the nonradiative recombination in the active layer of the device can be intuitively associated with the increasing density of deep levels

(within Schottky-Read model). One of the few methods suitable for determining the concentration of these nonradiative recombination centers is deep level transient spectroscopy (DLTS). In this method, we measure the junction-related capacitance transient in relatively broad temperature range. Meneghini et al. [25] performed DLTS measurements during the degradation of InGaN laser diodes. They detected a new signal of the amplitude increasing with the degradation time. This signal can be attributed to the electron trap situated 0.35–0.45 eV below the conduction band minimum. Meneghini et al. pointed out that such a level was previously attributed to $V_{Ga}$-$(O_N)_3$ complex [26]; however, this important problem needs further studies.

#### 8.4.2.3 Role of the Extended Defects

From the very beginning of nitride laser diode fabrication, the dislocations were considered as a key factor influencing the lifetime of these devices. As we mentioned before, the first CW operated violet laser of Nakamura et al. lased only for one second [3] and contained most likely around $10^9$ dislocations per centimeter squared. Systematic study performed by Takeya et al. [18] demonstrated very steep dependence of the laser diodes' lifetime on the number of threading dislocations existing in the laser diode structure. The lifetime of the device varies from hundreds to ten thousand hours when defect density decreases from $10^7$ down to $10^6$ $cm^{-2}$. Though the strong dependence of the laser diode lifetime on the defect density is a fact, the physical interpretation of this relation is far from being clear. In the classical III-V systems, the dislocation glide and multiply, leading to the formation of so-called dark lines [9] is a main cause of degradation. Movement and multiplication lead to the effective increase of the number of nonradiative recombination centers. In the nitride laser diodes, the dislocation does not multiply. Gallium nitride, with its large cohesion energy and strong, short interatomic bonds, should not be too much susceptible to these phenomena. Indeed, most of the existing papers claim that the dislocations in the stressed and unstressed regions of the aged laser diodes look the same, like it is demonstrated by Tomiya et al. [27] by using electron transmission microscopy (TEM) characterization. It means that no new dislocations are created, though the authors of that paper reports on the observed recombination-enhanced dislocation glide (REDG) mechanism. The dislocation activity was observed in situ in TEM microscope by irradiating the laser structures with electrons. These dislocations were, most likely, the dislocation loops lying in (0001) basal plane. However, the authors of the previously discussed paper claim that this mechanism cannot lead to multiplication of threading dislocations being extended along the *c* axis of the crystal.

The only reports on the formation of the localized dislocation network come from UK Sharp group [28, 29]. This group fabricated their devices by using a gas source MBE method. After a few-hour-long operation, the laser diodes were deprocessed and examined by TEM. The electron microscopy revealed the network of dislocations confined in the active region of the device. The authors claim that the mechanism of dislocation network formation may be similar to REDG

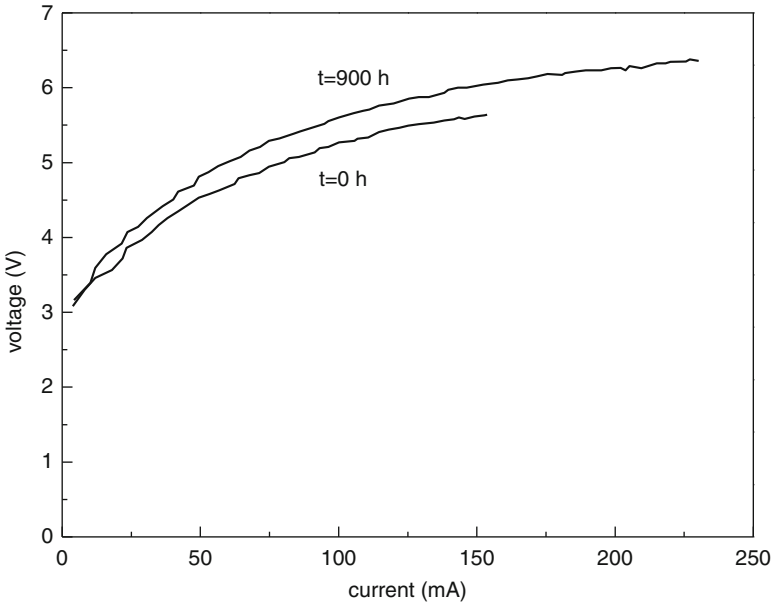



**Fig. 8.10** Comparison between current–voltage characteristics of virgin and stressed laser diode

mentioned before [27]. However, considering the uniqueness of this observation, the discussion about the true importance of this mechanism for nitride laser diode degradation has to be postponed until more facts are gathered.

Another piece of the puzzle which does not fit too well is that the laser diode containing $10^6$ dislocations in square centimeter has only around 10 dislocations within its stripe. If the diffusion is fast along the dislocation, the nonradiative recombination centers should be concentrated close to the dislocations forming dark areas. This effect is up to our knowledge never observed.

### *8.4.3 Contact Degradation: Operation Voltage and the Series Resistance*

Another typical feature of the degradation is an increase of the diode operation voltage under the prolonged operation. Figure 8.10 shows the I-V curves measured for the virgin laser diode and after 900 h of operation at 5 mW optical output power. The increase of 0.5 V of the operation voltage at the current of 150 mA was observed. Gradual degradation of the device series resistance and the operating voltage is most likely related to the deterioration of the metal contact to the *p*-type layer. There is however a lack of broader experimental base to evaluate this phenomenon. Data coming from the transistor reliability studies show no metal diffusion to semiconductor from the hardly stresses metal contacts.

## 8.5 Summary

The degradation of nitride laser diodes seems to be dominated by the increase of the nonradiative recombination in InGaN quantum wells. The increase of nonradiative recombination leads to the increase of threshold current, while the differential efficiency of the device remains roughly constant. This increase can be associated with the generation of point defects in the active layer. The generation of the defects is proportional to the current density and also proportional to the square root of the degradation time. Though diffusion is still considered as a plausible mechanism responsible for degradation, however, it has not been attributed to any dopant or impurity. The role of dislocations though apparently important is totally unclear.

## References

1. M. Osinski, D. Barton, P. Perlin, J. Lee, J. Crys. Grow. **189**, 808 (1998)
2. D. Barton, M. Osinski, P. Perlin, P.G. Eliseev, J. Lee, Microelectron. Reliab. **39**, 1219 (1999)
3. S. Nakamura, M. Senoh, S. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, Y. Sugimoto, H. Kiyoku, Appl. Phys. Lett. **69**, 4056 (1996)
4. S. Nakamura, J. Mater. Res. **14**, 2716 (1999)
5. C. Skierbiszewski, P. Wiśniewski, M. Siekacz, P. Perlin, A. Feduniewicz-Żmuda, G. Nowak, I. Grzegory, M. Leszczyński, S. Porowski, Appl. Phys. Lett. **88**, 221108 (2006)
6. K. Motoki, T. Okahisa, N. Matsumoto, M. Matsushima, H. Kimura, H. Kasai, K. Takemoto, K. Uematsu, T. Hirano, M. Nakayama, S. Nakahata, M. Ueno, D. Hara, Y. Kumagai, A. Koukitu, H. Seki, Jap. J. Appl. Phys. **40**, 140 (2001)
7. R. Dwiliński, R. Doradziński, J. Garczyński, L.P. Sierzputowski, A. Puchalski, Y. Kanbara, K. Yagi, H. Minakuchi, H. Hayashi, J. Cryst. Grow. **310**, 3911 (2008)
8. I. Grzegory, J. Phys. Condens. Matter **13**, 1 (2001)
9. M. Fukuda, *Reliability and Degradation of Semiconductors Lasers and LEDs* (Artech House, Boston, 1991)
10. M. Takeya, T. Hashizu, M. Ikeda, Proc. SPIE **5738**, 63 (2005)
11. M. Ikeda, T. Mizuno, M. Takeya, S. Goto, S. Ikeda, T. Fujimoto, Y. Ohfuji, T. Hashizu, Phys. Stat. Sol. (c) **1**, 1467 (2004)
12. J.A. Sharps, *Proceedings of the 27th Annual Boulder Damage Symposium on Laser-Induced Damage in Optical Materials, 1995* (SPIE, Bellingham, 1996), p. 676
13. Jongwoo Park, D.-S. Shinb, Mater. Chem. Phys. **88**, 410 (2004)
14. T. Schoedl, U.T. Schwarz, V. Kummler, M. Furitsch, A. Leber, A. Miller, A. Lell, V. Harle, J. Appl. Phys. **97**, 123102 (2005)
15. C.C. Kim, Y. Choi, Y.H. Jang, M.K. Kang, M. Joo, M.S. Noh, Proc. SPIE **6894**, 689400–1 (2008)
16. L. Marona, P. Wisniewski, P. Prystawko, I. Grzegory, T. Suski, S. Porowski, P. Perlin, M. Leszczynski, R. Czernecki, Appl. Phys. Lett. **88**, 201111 (2006)
17. M. Meneghini, N. Trivellin, K. Orita, S. Takigawa, T. Tanaka, D. Ueda, G. Meneghesso, E. Zanoni, Appl. Phys. Lett. **97**, 263501 (2010)
18. M. Takeya, T. Mizuno, T. Sasaki, S. Ikeda, T. Fujimoto, Y. Ohfuji, K. Oikawa, Y. Yabuki, S. Uchida, M. Ikeda, Phys. Stat. Sol. (c) **0**, 2292 (2003)
19. O.H. Nam, K.H. Ha, J.S. Kwak, S.N. Lee, K.K. Choi, T.H. Chang, S.H. Chae, W.S. Lee, Y.J. Sung, H.S. Paek, J.H. Chae, T. Sakong, J.K. Son, H.Y. Ryu, Y.H. Kim, Y. Park, Phys. Stat. Sol. (a) **201**, 2717 (2004)

20. L. Marona, P. Perlin, R. Czernecki, M. Leszczyński, M. Boćkowski, R. Jakiela, T. Suski, S.P. Najda, Appl. Phys. Lett. **98**, 241115 (2011)
21. M. Kneissl, D. Bour, L. Romano, Ch Van de Walle, J. Northrup, W. Wong, D. Treat, M. Teepe, T. Schmidt, N. Johnson, Appl. Phys. Lett. **77**, 1931 (2000)
22. L.L. Goddard, M. Kneissl, D.P. Bour, N.M. Johnson, J. Appl. Phys. **88**, 3829 (2000)
23. L.A. Coldren, S.W. Corzine, *Laser Diodes and Photonic Integrated Circuits*, 1st edn. (Wiley-Interscience, New York, 1995)
24. H.Y. Ryu, K.H. Ha, J.H. Chae, K.S. Kim, J.K. Son, O.H. Nam, Y.J. Park, J.I. Shim, Appl. Phys. Lett. **89**, 171106 (2006)
25. M. Meneghini, C. de Santi, N. Trivellin, K. Orita, S. Takigawa, T. Tanaka, D. Ueda, G. Meneghesso, E. Zanoni, Appl. Phys. Lett. **99**, 093506 (2011)
26. H.K. Cho, F.A. Khan, I. Adesida, Z.-Q. Fang, D.C. Look, J. Phys. D: Appl. Phys. **41**, 155314 (2008)
27. S. Tomiya, S. Goto, M. Takeya, M. Ikeda, Phys. Stat. Sol. (a) **200**, 139 (2003)
28. M. Rossetti, T.M. Smeeton, W.S. Tan, M. Kauer, S.E. Hooper, J. Heffernan, H. Xiu, C.J. Humphreys, Appl. Phys. Lett. **92**, 151110–1 (2008)
29. H. Xiu, E.J. Thrush, M. Kauer, T.M. Smeeton, S.E. Hooper, J. Heffernan, C.J. Humphreys, Phys. Stat. Sol. (c) **5**, 2204 (2008)

# Chapter 9
# Radiation-Enhanced Dislocation Glide: The Current Status of Research

**Koji Maeda**

**Abstract** Dislocations in most semiconductors exhibit radiation-enhanced dislocation glide (REDG) which is a cause of serious degradation in bipolar devices under forward biasing. Phenomenologically, REDG is characterized by the remarkable reduction of activation energy for glide velocity and the pre-exponential factor controlled by radiation intensity. Although the REDG shares features common with similar effects in point defects known as the recombination-enhanced defect reaction (REDR) that is well interpreted in terms of the phonon kick mechanism, it is not as yet established whether the REDG and the REDR share or not the same microscopic mechanism due to the lack of knowledge concerning electronic levels associated with the dislocations showing the REDG. Recent progress, however, has been achieved by studies of REDG in 4H-silicon carbide (4H-SiC) in which the dislocation component that exhibits the REDG is identified as 30°-Si(g) partial dislocation. Nevertheless, the driving force for the REDG in 4H-SiC is not the mechanical stress as in usual cases but an anomalous reversal of the sign of the Shockley stacking-fault (SSF) energy depending on the radiation intensity. An attempt was made in our recent experiments of photoinduced REDG to separate the intensity-dependent REDG effect and also the intensity-dependent driving force by simultaneous measurements of SSF expansion velocity (hence, the glide velocity of 30°-Si(g) partials) and the intensity of photoluminescence from the same SSFs (the latter allows us to deduce the driving force for the SSF expansion and hence the REDG). A comment, inspired by the studies of REDG in 4H-SiC, is given also to the mechanism of radiation-enhanced dislocation climb, another cause of degradation induced by minority-carrier injection.

K. Maeda (✉)
Department of Applied Physics, School of Engineering, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo, Japan
e-mail: maedakoji@gakushikai.jp

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_9,

© Springer Science+Business Media New York 2013

## 9.1 Introduction

Dislocations are in most cases detrimental defects in semiconductor devices. Particularly in bipolar optoelectronic devices, they act as vicious killer centers suppressing the radiative recombination of injected carriers that is essential for the device function. In the early stage of development of light-emitting diodes, the removal of dislocations from the device materials was requisite to improve the performance of the devices. A more serious problem was encountered in the development of laser diodes in which the lifetime of the devices was fatally limited by rapid multiplication of dislocations induced by forward biasing [1]. Soon, it was found that the dislocation multiplication is caused in two different modes of dislocation motion, dislocation climb and dislocation glide [2], which are completely different in the atomic processes involved. Technologically, the problem was solved by removing grown-in dislocations from the active region of the devices, but fundamentally, the mechanism has not been completely understood in spite of many experimental studies. Essentially, the same phenomena were found to be induced also by sample irradiation with electrons and photons, both of which generate excess electron-hole pairs in the crystal sharing a common feature with forward biasing in p-n junctions. Recently, the same effect was found to be the cause of degradation in SiC-based p-i-n power diodes which are subject to forward biasing during device operation [3]. This chapter reviews the current status of research and knowledge on the dislocation motion in semiconductors that is enhanced by such electronic stimulations, as an updated supplement to the author's previous review [4] on the same subject.

## 9.2 General Features of REDG

REDG effects are induced by various sources of electronic stimuli such as electron irradiation, photon illumination, and injection of minority carriers. All these stimulations cause recombination of electron-hole pairs so that most people use the abbreviation of REDG to mean the "recombination"-enhanced dislocation glide. However, the author prefers to use "radiation" rather than "recombination" for R in REDG because the term "recombination-enhanced" implies that recombination is essential in the operating mechanism.

Figure 9.1 shows experimental results obtained for dislocations of so-called As (g) type (for the nomenclature, see the next paragraph) in GaAs representing general features of REDG [5]. The filled marks indicate the dislocation glide velocity under the normal condition in darkness, whereas the open marks that obtained when the specimen is kept under radiation with an electron beam of 30 keV in energy. The behavior of dislocation velocity is characterized by the Arrhenius-type temperature dependence with the reduced activation energy and the pre-exponential factor that is also reduced under radiation. The enhanced

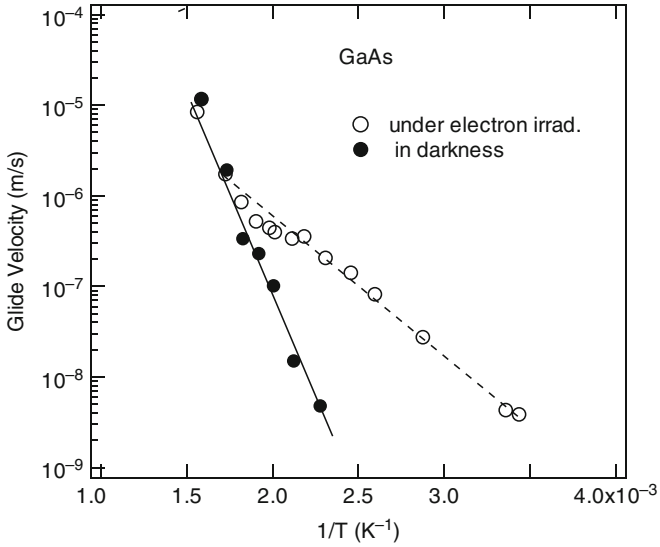



**Fig. 9.1** A typical REDG effect in As(g) dislocations in GaAs [4]. The *filled circles* and *open circles* indicate glide velocity in darkness and under electron beam irradiation, respectively

velocity increases with the radiation intensity (in this case, the electron beam current density) without any change in the activation energy. Thus, the temperature dependence of dislocation velocity $V$ is expressed by an experimental formula written as

$$V = V_{\mathrm{t}} \exp\left(-\frac{E_{\mathrm{t}}}{k_{\mathrm{B}}T}\right) + V_{\mathrm{e}}(I) \exp\left(-\frac{E_{\mathrm{t}} - \Delta E}{k_{\mathrm{B}}T}\right) \tag{9.1}$$

where $k_{\mathrm{B}}$ is Boltzmann constant, $T$ the temperature, $V_{\mathrm{e}}(I)$ the pre-exponential factor of the REDG term that increases with the beam current density or stimulation intensity $I$, and $\Delta E$ the reduction of activation energy independent of $I$ and $T$. Since $V_{\mathrm{t}} \gg V_{\mathrm{e}}$, the REDG effect is observed only at low temperatures at which the second term dominates over the first term of purely thermal origin. Usually $V_{\mathrm{e}}(I) \propto I^m$ with the exponent $m$ being close to unity [5, 6], but in some cases [7], the intensity dependence exhibits a tendency to saturation.

Table 9.1 compiles combinations of crystal and dislocation component for which REDG effects have been observed so far. The materials are arranged in descending order of the bandgap energy $E_{\mathrm{g}}$. The REDG effect is ubiquitously observed over tetrahedrally bonded covalent crystals ranging from elemental to $\mathrm{II_b}$–$\mathrm{IV_b}$ compound semiconductors. The systems for which the experimental values of $E_{\mathrm{t}}$ and $E_{\mathrm{t}} - \Delta E$ are lacking were those studied in qualitative experiments based on direct observations of dislocation motion by transmission electron microscopy.

**Table 9.1** Systems of crystal and dislocation component for which REDG effects have been reported

| Crystal | Dislocation component | $E_g$ (eV) | $E_t$ (eV) | $E_t - \Delta E$ (eV) | $\Delta E$ (eV) | Electronic stimulus | Ref. |
|---|---|---|---|---|---|---|---|
| ZnS | Screw, S(g), Zn(g) | 3.6 | 1.2 | 0.3 | 0.9 | Electron | [7] |
| ZnO | Basal Screw, Pyramidal Edge | 3.4 | – | – | – | Electron Photon | [8, 9] |
| GaN | Basal, Pyramidal | 3.4 | – | – | – | Electron | [10, 11] |
| 4H-SiC | 30°-Si(g) Partial | 3.3 | 1.35 | 0.25 | 1.1 | Carrier injection Photon | [12, 13, 14] |
| 6H-SiC | 90°-Partial | 3.0 | – | – | – | Electron | [15–17] |
| ZnSe | 90°-Se(g) Partial | 2.6 | – | – | – | Electron Photon | [18, 19] |
| CdS | Non-basal | 2.4 | – | – | – | | [20, 21] |
| n-GaP | P(g) | 2.3 | 1.48 | 0.37 | 1.11 | Electron | [6] |
| n-GaAs | Ga(g) | 1.4 | 1.7 | 0.6 | 1.1 | Electron | [5] |
| i-GaAs | Ga(g) | 1.4 | 1.7 | 0.4 | 1.3 | Electron | [22] |
| p-GaAs | Ga(g) | 1.4 | 1.3 | 0.7 | 0.6 | Electron | [23] |
| n-InP | In(g) | 1.3 | 1.6 | 0.7 | 0.9 | Electron | [23] |
| i-Si | 60° | 1.1 | 2.07 | 1.28 | 0.79 | Photon Electron | [24, 25] |
| i-Ge | 60° | 0.67 | 1.7 | 0.7 | 0.7 | | [26] |

As illustrated in Fig. 9.2, dislocations glissile on {111} or {0001} planes are known to be dissociated to Shockley partial dislocations separated by an intrinsic stacking fault (Shockley stacking fault = SSF). The Shockley partials are classified to 90° and 30° dislocations according to the angle between the Burgers vector and the dislocation line that is oriented preferentially along <110> or <11-20>. In polar AB compound semiconductors, edge dislocations with the extra-half plane terminating at A (B) atoms are called A(g) (B(g)) partials, where g in the parenthesis stands for glide set which means that slip takes place between narrowly spaced {111} or {0001} planes.

A common feature to be noted in Table 9.1 is that the reduction of activation energy $\Delta E$ is smaller than the bandgap energy $E_g$ except for the case of Ge. In most cases, the electronic stimuli are electrons or photons with energy exceeding the bandgap energy. Negative photoplastic effects (photo-softening) presumably induced by REDG in CdTe [27], CdS [28] and GaAs [29] showed that the effects become maximum at the bandgap energy of the respective crystals. There have been some attempts to induce REDG effects by illuminating the sample with a subgap light. The glide of 90°-Se(g) partial dislocations in ZnSe is induced by illumination with a threshold of 2.07–2.41 eV, considerably smaller than the bandgap energy $E_g$ of 2.8 eV [19]. Similar REDG induced by a subgap light was reported also for screw dislocations in ZnO [30] and 30°-Si(g) partials in 4H-SiC [12] with photon-energy thresholds both smaller by 0.9 eV than the bandgap energies.

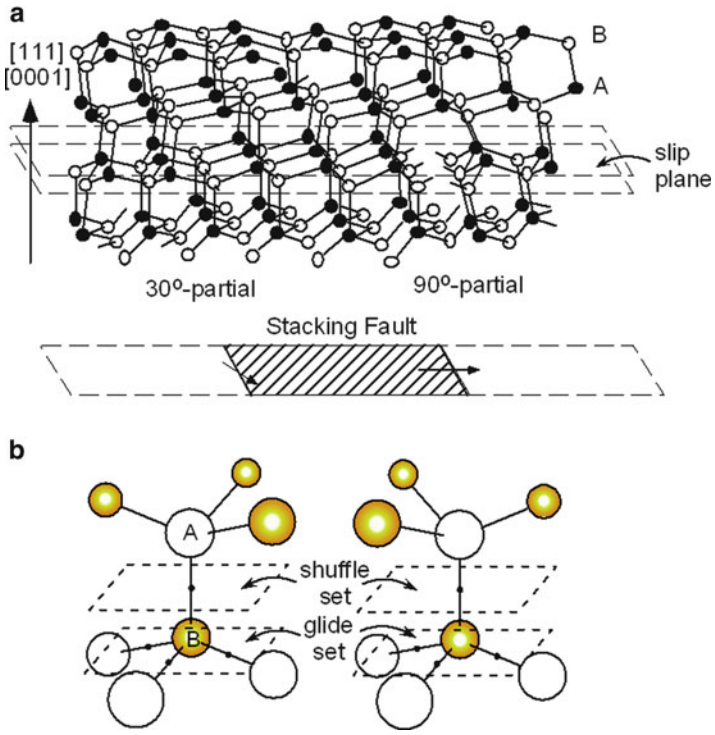



**Fig. 9.2** (**a**) Atomic structure of dislocation core in tetrahedrally bonded semiconductor crystals. The perfect set is dissociated to two Shockley partial dislocations separated by a Shockley stacking fault. (**b**) The slip plane in glide set is located between narrowly spaced {111} or {0001} planes. In AB polar crystals, A(g) and B(g) refer to Shockley partial dislocations in the glide set whose extra-half plane terminates with A atoms and B atoms, respectively

## 9.3 Electronic Activity of Dislocations

Two requisites must be satisfied for the REDG to actually bring about degradation in the optoelectronic devices. The first requisite is the driving force for the multiplication of dislocations, and the second is the electronic activity of the dislocations that degrade, e.g., the optoelectronic properties of the crystal. Since the electronic activity of dislocations may be responsible for the REDG effect to be induced, we should first argue the origin of the electronic levels associated with dislocations in semiconductors. Although the research on this problem has a long history, but experimental knowledge on the dislocation-related electronic levels is still insufficient.

A long time before confronting the serious problem with the rapid degradation of laser diodes, people had noticed the electronic activity of dislocations in semiconductors as a problematic issue already at the beginning of semiconductor devices: It had been known that the presence of dislocations has harmful effects on

the electronic properties of the crystals. Some of them are caused by indirect effects of dislocations such as disturbing uniform dopant distribution as a result of impurity precipitation to dislocation lines or providing the crystal with fast diffusion paths for the dopant atoms. Other problems, such as reduction of carrier mobility due to blocking of carrier transport by charged dislocations, action of dislocations as carrier generation-recombination centers increasing leakage current, have their origins in the structural singularity of the dislocation core.

The most frequently conceived origin of dislocation-related electronic levels is an array of dangling bonds (DBs) which may present along the dislocation cores [31]. However, experimental evidence for this origin is limited, probably because dislocation cores are mostly reconstructed to eliminate DBs, and hence the number of such DB centers is so small that the signal is too weak to be detected experimentally. It is also difficult to discriminate the signal, if observed, from that due to point defects which are generated on glide of dislocations when the crystals are plastically deformed with the intention of increasing dislocation density.

There are reports, more in number, on shallow gap levels associated with dislocations [32]. These shallow levels are considered to originate in the deformation potential formed around a dislocation line. The dislocation levels of this origin, however, are too shallow to act as recombination centers so that they are not the present concern.

Direct experiments attempting to investigate the microscopic origin of electronic centers associated with dislocations are those based on spatially resolved spectroscopic measurements. In many cases, dislocations are imaged in dark contrasts by scanning electron microscopy in cathodoluminescence mode (SEM-CL). In real experiments, however, dislocations are not necessarily free of decoration with impurities and point defects, making the interpretation of the signal origin not straight forward. Nevertheless, in some few cases [33–36], gliding dislocations, which should be free of decoration, are observed emitting photoluminescence at a characteristic wavelength.

Since the dislocations gliding under radiation should be free of decoration, the electronic activity responsible for the REDG effect must be associated with fresh dislocations. In GaAs, gliding As and Ga dislocations are observed in dark contrasts in SEM-CL images [37], which indicates that the carrier recombination at these dislocations is totally nonradiative or radiative with a large Stokes shift reducing the luminescence photon energy below the sensitivity range of the photodetector. It is worth to note that the 30°-Si(g) partial dislocations in 4H-SiC [34–36] and prismatic dislocations in CdS emit a subgap light during their enhanced glide [21].

## 9.4 Theoretical Models for REDG

The experimental features of REDG expressed by Eq. 9.1, the intensity-independent reduction of activation energy and the reduced pre-exponential factor increasing with the stimulation intensity, are quite similar to the features of

recombination-enhanced defect reactions observed for point defects in many semiconductors [38]. The most frequently invoked mechanism for the REDR is the phonon kick model first proposed by Weeks et al. [39]. The phonon kick model assumes that excess carriers generated by electronic stimulation recombine via a gap state associated with the defect nonradiatively releasing an excess energy so as to be utilized to assist the reaction or motion of the defect itself. The theory shows that the defect reaction rate $K$ is given by

$$K = K_t \exp\left(-\frac{Q_{\text{th}}}{k_{\text{B}}T}\right) + \varsigma R \exp\left(-\frac{Q_{\text{th}} - E_{\text{p}}}{k_{\text{B}}T}\right), \tag{9.2}$$

where $R$ is the recombination rate, an efficiency factor, and $E_{\text{p}}$ the energy vibrationally deposited to the defect upon capture of a carrier in the course of recombination. The theory was later reformulated by Sumi [40] who showed that the effective reduction of activation energy $\Delta Q$ in the second term is upper limited by $E_{\text{p}}$.

Since $E_{\text{p}} <$ bandgap energy $E_{\text{g}}$ in the phonon kick mechanism, the reduction of activation energy $\Delta Q$ should be smaller than $E_{\text{g}}$. As seen in Table 9.1, the experimental reduction of activation energy $\Delta E$ in REDG is actually smaller than $E_{\text{g}}$, which looks supporting a hypothesis that the REDG effect is brought about by the phonon kick mechanism. However, in contrast to the case of point defects for which the phonon kick mechanism is supported by the agreement of $E_{\text{p}}$ with the energy depth of defect-related trap level directly measured in a parallel experiment [41], the case of REDG lacks experimental information on the relevant dislocation level and hence on the magnitude of $E_{\text{p}}$. Moreover, the microscopic process of dislocation glide is not simple as reactions of point defects, which makes the interpretation of $\Delta E$ in REDG (Table. 9.1) not straight forward.

To consider the mechanism of REDG, therefore, we have to understand the elementary processes of dislocation glide in semiconductors. The dislocation glide in semiconducting covalent crystals is observed to be smooth and continuous indicating that the glide motion is controlled by a large lattice friction. This is based on the fact that the potential energy of a dislocation in a crystal lattice is minimized when the dislocation is lying between most widely-spaced-atomic columns, and for a dislocation to glide, it must overcome such an intrinsic potential barrier, the so-called Peierls potential. Since the Peierls potential is large in covalent crystals, it is the rate-controlling barrier against dislocation glide.

Surmounting the Peierls barrier is achieved by two successive processes, formation of a kink pair on a straight dislocation line (Fig. 9.3a) and subsequent migration of the generated kinks along the dislocation (Fig. 9.3a). Thus, the dislocation glide in covalent semiconductors consists of two elementary processes, kink-pair formation (KPF) and kink migration (KM). As a consequence, the activation energy governing the dislocation velocity consists of two terms, kink-pair formation energy $F_{\text{kp}}$ and kink-migration barrier $E_{\text{m}}$. If the phonon kick mechanism causes REDG, it can operate in the two elementary processes, which complicates the microscopic mechanism of REDG. A further complication arises from the fact that there are two

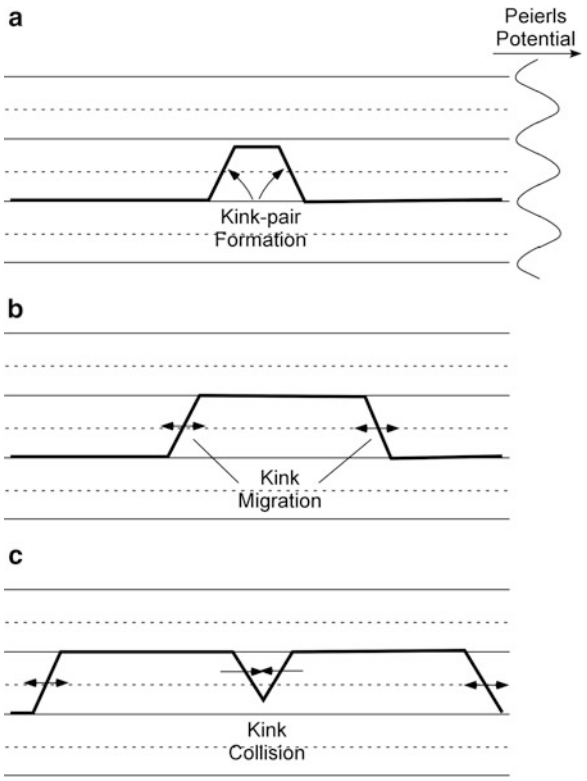



**Fig. 9.3** Elementary processes in the Peierls mechanism. (**a**) A part of a dislocation line surmounts the ridge of the Peierls potential by forming a pair of kinks. (**b**) The kinks generated in (a) diffusively migrate along the dislocation line to advance the dislocation by a period of the Peierls potential. (**c**) Kinks collide with each other before they travel the whole dislocation line

possible cases for kink migration [42], one is the case in which generated kinks migrate all along the whole dislocation line without colliding with other kinks (kink-collisionless case) and another is the case in which kinks collide with each other to annihilate without reaching the end of the dislocation line (kink-collision case illustrated by Fig. 9.3c). An experimentally observable signature which allows us to discriminate between the two cases is the dependence of dislocation velocity $V$ on the dislocation length $L$: $V$ is independent on $L$ in the kink-collision case, but $V$ increases linearly with $L$ in the kink-collisionless case.

Table 9.2 shows features of REDG theoretically predicted by assuming that the phonon kick mechanism operates in one or both of the elementary processes [43]. The cases are assorted by combinations of kink-migration regime and enhanced processes. KPF-(KM) denotes that the enhanced process is kink-pair formation (kink migration), and KPM + KM denotes that both processes are enhanced. $\Delta E_{\mathrm{KPF}}$ and $\Delta E_{\mathrm{KM}}$ are the reduction of activation energy in the kink-pair formation, and that in the diffusive migration of a kink. $R_{\mathrm{KPF}}$ and $R_{\mathrm{KM}}$ indicate the rate enhancing the kink-pair formation and that enhancing the kink migration, respectively. The dependence on radiation intensity $I$ in the parentheses in the column of rate factor $r$ holds when $I$ is low enough for $R_{\mathrm{KPF}}$ and $R_{\mathrm{KM}}$ to be proportional

**Table 9.2** The parameters theoretically expected for two regimes of kink migration and the elementary process(es) enhanced by the phonon kick mechanism [43]

| Kink-migration regime | Enhanced process | Rate factor $r$ | Activation-energy reduction $\Delta E$ |
|---|---|---|---|
| Kink-collision | KPF + KM | $\propto \sqrt{R_{\mathrm{KPF}} R_{\mathrm{KM}}} (\propto I)$ | $\frac{1}{2}(\Delta E_{\mathrm{KPF}} + \Delta E_{\mathrm{KM}})$ |
| | KPF | $\propto \sqrt{R_{\mathrm{KPF}}} (\propto \sqrt{I})$ | $\frac{1}{2}\Delta E_{\mathrm{KPF}}$ |
| | KM | $\propto \sqrt{R_{\mathrm{KM}}} (\propto \sqrt{I})$ | $\frac{1}{2}\Delta E_{\mathrm{KM}}$ |
| Kink-collisionless | KPF + KM | $\propto R_{\mathrm{KPF}} (\propto I)$ | $\Delta E_{\mathrm{KPF}}$ |
| | KPF | $\propto R_{\mathrm{KPF}} (\propto I)$ | $\Delta E_{\mathrm{KPF}}$ |
| | KM | No effect | No effect |

to $I$. If the phonon kick mechanism applies, the linear intensity dependence experimentally observed in some cases [5, 6] suggests that at least kink-pair formation is enhanced, because otherwise the intensity dependence would be proportional to $\sqrt{I}$ or independent of $I$ due to the saturation of rate factor with $I$. In most of the previous experiments, however, since it was not distinguished whether or not the dislocation velocity depended on the dislocation length, it was not possible to judge in which regime the kink migration was. Therfore, in the present state of knowledge, we could not discuss the origin of the activation-energy reduction experimentally measured.

Another uncertainty which makes it difficult to discuss the detailed mechanism of REDG arises from the dissociation of dislocations to Shockley partials. In most semiconductors, the dissociation width is so small that the REDG effect could be observed only for the perfect set of partial dislocations. This means that we could not experimentally resolve which partial is responsible for the observed REDG effect. The electronic properties of partial dislocations should be substantially variable due to the different core structures between 30 and 90° partials and due to the different terminating atom species in polar compounds. Therefore, unless the responsible partial is known, it is not possible to discuss the microscopic mechanism of REDG based on the expected electronic properties of the partial dislocations. Fortunately, however, this problem is absent in 4H-SiC crystals for which recent extensive studies (For comprehensive review, see [44]) have provided us with ample information on the REDG effect at the level of partial dislocations of a specific type. Therefore, the works on 4H-SiC, which were not dealt with in the previous review [4], are explored somewhat in depth in the following.

## 9.5 The REDG Effect in SiC

A peculiar property of SiC is the exceptionally low stacking-fault energy that causes the vast variety of allotropes of SiC in different stacking sequences [44]. 4H denotes a hexagonal stacking with a period of four layers. In this notation, the zincblende SiC is denoted by 3C and the wurtzite SiC by 2H. The REDG effect in SiC was first found in the most common 6H-SiC crystals [15–17]. Mechanically, SiC crystals are so hard that their powder is used as abrasives for grinding. This is primarily due to the high Peierls potential that hinders dislocation glide in this material. In most semiconductors, partial dislocations are strongly bound to the counter partials because widening of stacking fault to cause a separation of the partials from the counter parts needs a large increase of stacking-fault energy. In SiC, however, owing to the small magnitude of $\gamma_{SF}$, the stacking-fault energy per area, glides of relatively mobile partials yield some plastic deformation at ambient temperatures [45], which provides some extent of ductility to the grinding powder which is necessary for practical abrasives usable with a long lifetime.

Electronically, the SiC allotropes are semiconductors whose bandgap energies vary depending on the stacking sequence. Among them, 4H-SiC has the largest bandgap energy of 3.4 eV, which is the reason for this allotrope promising applications to power devices tolerant of high reverse bias voltages. After a couple of works [3, 46] were reported by a Swedish group on degradation of p-i-n power diodes due to extension of stacking faults by enhanced glide of partial dislocations under device operation, intensive studies have been done for the REDG effect in 4H-SiC [44]. The important findings so far obtained are:

(1) The partial dislocations that exhibit a REDG effect are 30°-Si(g) Shockley partials [36].
(2) The temperature dependence of glide velocity, as shown in Fig. 9.4, exhibits features typical in REDG effects, a reduction of activation energy and a reduced pre-exponential factor.
(3) The enhanced glide occurs in long partial dislocations with velocities showing no significant dependence on the dislocation length.
(4) The driving force for the REDG is not a mechanical stress but of an anomalous origin, as discussed in detail in the next section [47].
(5) The 30°-Si(g) partials exhibit photoluminescence at 1.8 eV [35].

The fact (1) provides us with a basis to focus on the specific core structure of partial dislocations that exhibit the REDG effect. The fact (2) suggests that the effect may be caused by the phonon kick mechanism. From the fact (3), we could judge that the glide is in the kink-collision regime. A further discussion has to be suspended until we know experimentally how the dislocation velocity depends on the radiation intensity under the anomalous situation (4).

Usually, REDG is driven by a mechanical shear stress somehow applied on the dislocations. In most cases for which REDG effects are so far observed, stresses are built in the crystal due to a lattice mismatch with the substrate, exerted on the

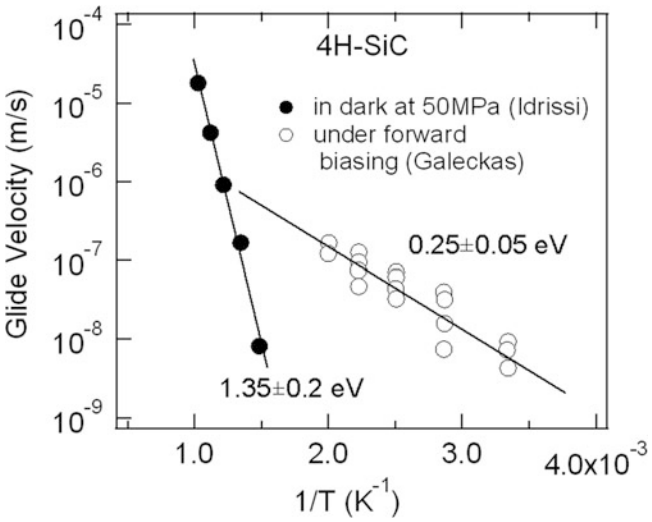



**Fig. 9.4** REDG effect of 30°-Si(g) partial dislocations in 4H-SiC. The *open circles* indicate glide velocity under forward biasing [12] and the *filled circles* velocities in darkness [13]

dislocations by interaction with other dislocations, or intentionally applied by external loading. The stress level is usually around several tens MPa or $10^{-3}$ in unit of shear strain. In dissociated dislocations, the glide of partial dislocations expanding Shockley stacking faults (SSFs) costs energy for increasing the area of the SSFs. The effective shear stress $\tau_r$ impeding the expansion of SSFs via glide of partial dislocations is given by $\gamma_{SF}/b$, where $\gamma_{SF}$ and $b$ denote the formation energy of SSF per unit area and the magnitude of the Burgers vector of the partial dislocation, respectively. In ordinary semiconductors, $\gamma_{SF}$ ranges around several tens mJ/m$^2$ (several tens meV/atom) [48], which corresponds to $\tau_r$ of several hundreds of MPa, one order of magnitude larger than the ordinary stress level in the crystals. This means that expansion of SSFs is not possible under normal conditions, which is the reason why the dislocations in semiconductors normally dissociated could glide only in a perfect set. In 4H-SiC, however, since $\gamma_{SF}$ is as small as $14.7 \pm 2.4$ mJ/m$^2$ (9 meV/atom) [49], expansion of SSFs may be possible if a sufficient stress is present [45, 50].

In the REDG effect of 30°-Si(g) partial dislocations in 4H-SiC, however, the expansion of SSFs cannot be interpreted only by the effect of mechanical stress. Instead, we have to accept that the formation energy of SSF becomes effectively negative only during electronic stimulation [47]. This is evidenced by the facts that the expansion of SSFs under electronic stimulation is reversed in darkness with no stimulation [51], which means that the stacking-fault energy in thermal equilibrium is positive as normally expected.

For this anomalous sign reversal of SSF energy, Miao and his coworkers [52] proposed that excess electrons injected into crystals are captured by an electronic quantum well associated with an SSF (Fig. 9.5a), thereby lowering the energy of the

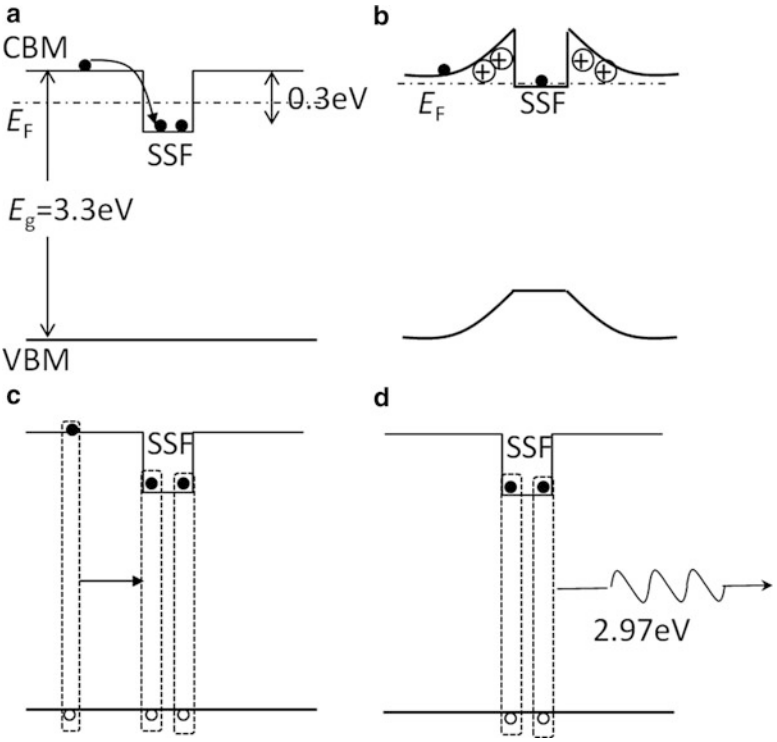



**Fig. 9.5** (**a**) Energy diagram of 4 H-SiC with a Shockley stacking fault (*SSF*) which forms a well of 0.3 eV in depth and could trap electrons in the conduction band. (**b**) Charge up of SSF expected when only electrons are trapped. (**c**) Efficient reduction of excess electronic energy by trapping electrons and holes in pairs. (**d**) The trapped electron-hole pairs recombine to emit photoluminescence of 2.97 eV in energy

excess electrons and as a consequence reducing the effective formation energy of the SSF. The presence of such SSF-associated electron-trap levels below the bottom of conduction band is experimentally verified by a photoluminescence of 2.97 eV spatially localized at SSFs [53]. Since the depth of the trap level is as large as 0.3 eV whereas $\gamma_{SF}$ in thermal equilibrium is as small as 9 meV/atom [49], they claimed that the effective $\gamma_{SF}$ under electron injection would change to be negative at high injection levels. Kuhr et al. [54], however, pointed out a difficulty of this model that only a marginal reduction of $\gamma_{SF}$ is predicted due to an upward shift of electronic bands (Fig. 9.5b) compensating the electric charge-up of SSFs that would occur on electron trapping. Lambrecht and Miao [55] rebutted that objection invoking a screening by free carriers. Otherwise, in author's opinion, the difficulty of the electric charge-up could be avoided by considering trapping of holes together with electrons, as illustrated by Fig. 9.5c. We will assume this model subsequently.

Thus exceptionally, the driving force of REDG in 4H-SiC is not a mechanical stress but of electronic origin. The effective formation energy $\gamma_{SF}^{*}$ of SSF per unit area under stimulation is hence considered to be given by

$$\gamma_{SF}^{*} = \gamma_{SF} - \alpha\rho, \tag{9.3}$$

where $\rho$ denotes the areal density of electron-hole pairs (excitons) trapped at the SSF and $\alpha$ a constant representing the efficiency of energy lowering by the carrier trapping. Since the stress dependence of dislocation glide velocity in semiconductors is almost linear [5], enhanced glide velocity should depend linearly on the effective driving stress $\gamma_{SF}^{*}/b$. Therefore, the phonon kick mechanism would predict that the enhanced glide velocity is to be given by

$$V \propto [\alpha\rho(I) - \gamma_{SF}] \cdot r(I) \exp\left(-\frac{E_{th} - \Delta E}{k_B T}\right). \tag{9.4}$$

Here, $I$ denotes the stimulation intensity, $r(I)$ the rate factor which should be compared with the $I$-dependences listed in Table 9.2, $E_{th}$ the glide activation energy in thermal equilibrium, and $\Delta E$ the reduction of activation energy. Thus, the pre-exponential factor in Eq. 1 is composed of two factors both dependent of $I$, $r(I)$ arising from the genuine REDG effect and the driving force embraced with the parenthesis in front of $r(I)$. The problem now is how to assess the $I$-dependence of the driving force which is absent in the ordinary REDG effects.

As mentioned above, SSFs in 4H-SiC yield a characteristic photoluminescence (PL) when the crystal is illuminated with a supragap light [53]. If we assume that the PL intensity is proportional to $\rho(I)$, the density of electrons trapped at the SSFs pairing with holes, the driving force can be assessed experimentally by measuring the PL intensity. In a very recent experiment, the present author and his coworkers [56] conducted simultaneous measurements of SSF expansion velocity (hence the glide velocity of 30°-Si(g) partials) and the intensity of PL from the same SSFs. A remarkable fact is that the velocity of SSF expansion or 30°-Si(g) partial glide under illumination with a light of 4.66 eV (greater than the bandgap of 3.3 eV) is obviously supralinear. Figure 9.6 shows a result in which the abscissa indicates the light intensity and the ordinate, the glide velocity divided by PL intensity in order to evaluate the relative rate factor $r(I)$ on the assumption of $\alpha\rho(I) \gg \gamma_{SF}$. The rate factor thus experimentally measured depends linearly on $I$ rather than on $\sqrt{I}$, which favors that the enhanced processes in Table 9.2 are both kink-pair formation and kink migration.

For electronic gap levels associated with dislocations in 4H-SiC, Blumeneau et al. [57] conducted first-principles calculations on reconstructed perfect cores and various structures of kinks in 30°-Si(g) partials as well as on those in 30°-C(g) partials. These calculations show that no deep states are present in the energy gap except for one type of kink in 30°-C(g) partials. One may feel it puzzling that the experimentally inferred enhancement of both kink-pair formation and kink

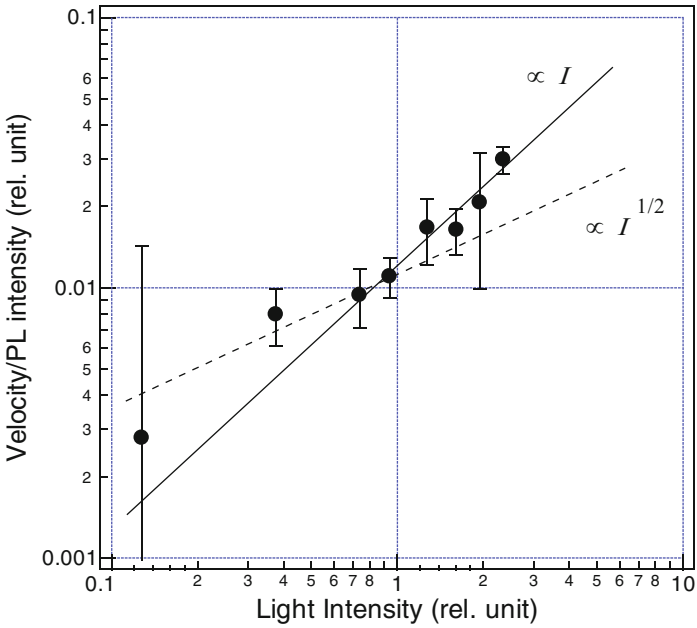



**Fig. 9.6** The intensity dependence of the rate factor $r(I)$ in Eq. 9.4 experimentally deduced from simultaneous measurements of dislocation velocity and photoluminescence intensity representing the driving force [56]

migration appears inconsistent with perfect cores and kinks being electronically inactive. For perfect cores, Savini et al. [58] showed that unreconstructed cores would have gap states originating in dangling bonds. It was proposed [59] that the unreconstructed cores, although energetically higher than reconstructed cores, are favored to become responsible for the REDG effect by a dynamical effect of electron-hole plasma under radiation. Such models that kink-pair formation is induced by radiation on straight dislocations, whichever the case, reconstructed or not, must have some mechanism for the excited carriers extended in space to get localized at a point on the uniform dislocation line at which the kink-pair formation actually takes place. This difficulty as well as the difficulty with the lack of gap levels associated with straight reconstructed cores in 30°-Si(g) partials may be removed by considering some imperfections on straight dislocations. As proposed by Galeckas et al. [12], a strong candidate of such a defect is an antiphase defect in reconstruction cores (the so-called soliton) which possesses a dangling bond.

For the enhancement of kink migration, we may have to consider kink structures different from those investigated by the previous calculations [57, 58]. If we resort to the soliton model for the mechanism of enhanced kink-pair formation, a possible structure of electronically active kinks is a complex of kink and soliton (kink with a dangling bond).

The origin of 1.8 eV luminescence from 30°-Si(g) partial dislocation lines is not yet established. An interesting fact which may be worth to note is that the 1.8 eV

luminescence diminishes anomalously at high excitation intensities [35]. If we adopt a model by Fujimoto and his coworkers [60] who proposed that the luminescence originates in impurity (nitrogen) atoms decorating dislocation cores, we have to consider that the signal decrease with increasing excitation should be due to detachment of the atoms from the dislocations whose glides are enhanced by the radiation. This model is worth to be disputed because it might explain faint luminescence from 30°-C(g) partials at the same energy of 1.8 eV as from 30°-Si(g) partials [35], but it is quite challenging to the common sense that the gliding dislocations should be free of decoration, as assumed earlier in this chapter. Another possible model would be that the 1.8 eV luminescence originates in the dislocation line itself, but it is suppressed by an increase of kinks (kink-soliton complexes) acting as nonradiative recombination centers on the dislocation line due to enhanced kink-pair formation. Further studies, however, are needed to test this hypothesis.

## 9.6 Implications for Radiation-Enhanced Dislocation Climb

Radiation-enhanced dislocation climb (REDC) is another important cause of degradation in optoelectronic devices, but a consensus about the mechanism is not yet reached. Although the REDC is not the main subject of this chapter, one may suspect that the REDC is brought about by a mechanism similar to that of REDC.

For dislocation climb to occur in compound AB semiconductors, an equal number of two species A and B have to be desorbed to or emitted from the extra-half plane of edge dislocation in various processes, as illustrated in Fig. 9.7. Usually, the driving force for dislocation climb is a thermodynamical force and/or a mechanical force. The thermodynamical force originates in supersaturated point defects, the annihilation of which would relieve the defect formation energy. The mechanical force arises from the normal stress applied on the extra-half plane of the edge dislocation, the climb motion of which would lower the elastic energy of the crystal. In cases (d) and (h) in Fig. 9.7 where the climb motion is achieved by emission of point defect pairs, such a mechanical stress, if present, could drive the dislocation climb, but it has to compensate the rather large formation energy of the emitted defect pairs and therefore needs to exceed a high stress level (around 100 GPa) unrealistic in crystals. In cases (b), (c), (e), and (f), nevertheless, the thermodynamical force due to desorption of one type of supersaturated defects may eliminate this difficulty if the annihilation of the excessive point defects will compensate the formation energy of emitted counterparts [61]. In whichever the case except for (d) and (h), we have to assume a high supersaturation of point defects in at least one of the sublattices, which was unlikely in the investigated crystals. A fact that the REDC in light-emitting diodes proceeds with the growth of interstitial dislocation loops [1, 62] can be well understood by considering the case (h) in Fig. 9.7, as was first proposed by O'Hara and his coworkers [63].

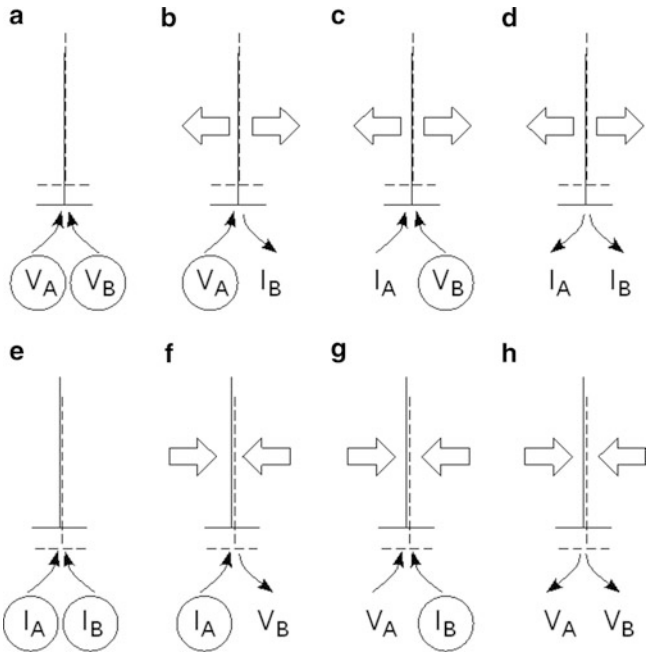



**Fig. 9.7** Various processes of dislocation climb in AB compound crystals. The thick and dashed lines depict the extra-half planes of edge dislocations before and after climb. $V_A$, $V_B$, $I_A$, and $I_B$ indicate A vacancy, B vacancy, A interstitial, and B interstitial, respectively. The defects encircled are excessive defects preexisting in the crystal. The *arrows* show the normal stress applied on the extra-half plane that would drive the climb motion

As pointed out above, we need a reason other than a mechanical driving force for dislocations to climb in the process (h) that must be followed by electronically enhanced diffusion of the vacancies. The driving force here in our concern is one that induces the vacancy emission. The emission of vacancies which costs energy apparently resembles the anomalous expansion of SSFs in 4H-SiC. However, the REDG and REDC are different in the point that the dislocation glide in REDG is bidirectional (the kink-pair formation can occur in opposite directions according to the polarity of the driving force), while the vacancy emission in REDC occurs in such one way as to grow the interstitial dislocation loops. In other words, the dislocation glide in REDG is directed by the driving force originating in a cause (sign reversal of SSF energy) independent of the cause of REDG (enhanced kink-pair formation and kink migration), whereas the REDC is directed by a single mechanism of vacancy emission at the expense of energy. A plausible mechanism for dislocation climb acquiring an extra energy is a phonon kick which occurs on recombination (trapping) of excess carriers at centers being present along the dislocations. For the REDC to persist, the recombination centers should be jogs and therefore different from those centers responsible for the REDG

effect. This means that REDG and REDC are not necessarily correlated in spite of the similarity in the resultant multiplication of dislocation networks.

## 9.7 Summary

Bipolar devices based on p-n junction diodes are apt to degrade in the mode of radiation-enhanced dislocation glide (REDG). REDG is a phenomenon commonly observed in many semiconductors subjected to minority-carrier-injecting radiation with light and electrons and forward biasing p-n diode junctions. The enhanced dislocation velocity is characterized by the reduced activation energy and the reduced pre-exponential factor, which makes the effect more significant at low temperatures. Recent progress in studies on REDG in 4H-SiC triggered by degradation of p-i-n power devices has added more detailed knowledge of the partial dislocation structures exhibiting REDG and of the glide regime that has a fundamental importance in discriminating the elementary process that is enhanced in the Peierls mechanism of dislocation glide in covalent crystals. Simultaneous measurements of radiation-intensity dependence of the glide velocity and of the driving force for the REDG of unusual electronic origin in 4H-SiC have revealed that the REDG effect is brought about by enhanced kink migration as well as enhanced kink-pair formation. Some comments were made on radiation-enhanced dislocation climb which has some similarity to REDG but is fundamentally different in the mechanism [64].

## References

1. P. Petroff, R.L. Hartman, Appl. Phys. Lett. **23**, 469 (1973)
2. S. Kishino, N. Chinone, H. Nakashima, R. Ito, Appl. Phys. Lett. **29**, 469 (1977)
3. J.P. Bergman, L. Lendenmann, P.Å. Nilsson, U. Lindelfelt, P. Skytt, Mater. Sci. Forum **353–356**, 299 (2001)
4. K. Maeda, S. Takeuchi, in *Dislocation in Solids*, ed. by F.R.N. Nabarro, M.S. Duesbery, vol. 10 (North-Holland, Amsterdam, 1996), p. 435
5. K. Maeda, M. Sato, A. Kubo, S. Takeuchi, J. Appl. Phys. **54**, 161 (1983)
6. N. Maeda, S. Takeuchi, Jpn. J. Appl. Phys. **29**, 1151 (1990)
7. C. Levade, A. Faress, G. Vanderschaeve, Philos. Mag. A **69**, 855 (1994)
8. K. Maeda, K. Suzuki, Y. Yamashita, Y. Mera, J. Phys. Cond. Matt. **12**, 10079 (2000)
9. Y. Ohno, T. Taishi, I. Yonenaga, Phys. Stat. Sol. A **206**, 1904 (2009)
10. S. Tomiya, T. Hino, S. Goto, M. Takeya, M. Ikeda, IEEE J. Sel. Topics Quantum Electron. **10**, 1277 (2004)
11. K. Maeda, K. Suzuki, M. Ichihara, S. Nishiguchi, K. Ono, Y. Mera, S. Takeuchi, Physica B **273–274**, 134 (1999)
12. A. Galeckas, J. Linnros, P. Pirouz, Phys. Rev. Lett. **96**, 025502 (2006)
13. H. Idrissi, G. Regula, M. Lancin, J. Douin, B. Pichaud, Phys. Stat. Sol. C **2**, 1998 (2005)
14. J. Caldwell, O.J. Glembocki, R.E. Stahlbush, K.D. Hobart, J. Elec. Mater. **37**, 699 (2008)
15. A.N. Pilyankevich, V.F. Britun, Phys. Stat. Sol. A **82**, 449 (1984)

16. K. Maeda, K. Suzuki, M. Ichihara, Microsc. Microanal. Microstruct. **4**, 211 (1993)
17. J.W. Yang, X.J. Ning, P. Pirouz, *Japan-US Workshop on Functional Fronts in Advanced Ceramics* (Ceramic Society of Japan, Tokyo, 1994), p. 55
18. C. Levade, G. Vanderschaeve, J. Cryst. Grow. **197**, 565 (1999)
19. Y. Ohno, Appl. Phys. Lett. **87**, 181909 (2005)
20. V.I. Roshetov, A.É. Armeeva, G.V. Bushueva, O.N. Talenskii, A.A. Pendyur, A.N. Pechennov, N.A. Tyapunina, Sov. Phys. Solid State **29**, 689 (1987)
21. V.D. Negrii, Yu.A. Osipyan, N.V. Lomak, Phys. Stat. Sol. A **126**, 49 (1991)
22. K. Maeda, S. Takeuchi, *Dislocations in Solids* (University of Tokyo Press, Tokyo, 1984), p. 433
23. K. Maeda, S. Takeuchi, Appl. Phys. Lett. **42**, 664 (1983)
24. K. Küsters, H. Alexander, Physica B **116**, 1151 (1983)
25. N. Maeda, K. Kimura, S. Takeuchi, Bull. Acad. Sci. USSR Phys. Ser. **51**, 93 (1987)
26. I. Yonenaga, M. Werner, M. Bartch, U. Messerschmidt, E.R. Weber, Phys. Stat. Sol. A **171**, 35 (1999)
27. E.Y. Gutmanas, N. Travitzky, P. Haasen, Phys. Stat. Sol. A **51**, 922 (1979)
28. M.S. Shikhsaidov, Sov. Phys. Solid State **23**, 968 (1981)
29. B.E. Midvanyan, M.S. Shikhsaidov, Phys. Stat. Sol. A **107**, 131 (1988)
30. Y. Ohno, T. Taishi, I. Yonenaga, Phys. Stat. Sol. A **209**, 1904 (2009)
31. D.B. Holt, B.G. Yacobi, *Extended Defects in Semiconductors* (Cambridge University Press, Cambridge, 2007)
32. R. Jones, Mater. Sci. Eng. B **71**, 24 (2000)
33. V.D. Negrii, Osipyan YuA, N.V. Lomak, Phys. Stat. Sol. A **126**, 49 (1991)
34. R.E. Stahlbush, M. Fatemi, J.B. Fedion, S.D. Arthur, L.B. Rowland, S. Wang, J. Elect. Mater. **31**, 370 (2002)
35. A. Galeckas, J. Linnros, P. Pirouz, Appl. Phys. Lett. **81**, 883 (2002)
36. S. Ha, M. Benamara, M. Skowronski, H. Lendermann, Appl. Phys. Lett. **83**, 4957 (2003)
37. K. Maeda, S. Takeuchi, J. Phys. (Paris) **44**, C4–375 (1983)
38. L.C. Kimerling, Solid State Electron. **21**, 1392 (1978)
39. J.D. Weeks, J.C. Tully, L.C. Kimerling, Phys. Rev. B **12**, 3286 (1975)
40. H. Sumi, Physica B **117–118**, 197 (1983)
41. L.C. Kimerling, D.V. Lang, Inst. Phys. Conf. Ser. **23**, 589 (1975)
42. J.P. Hirth, J. Lothe, *Theory of Dislocations* (McGraw-Hill, New York, 1968), p. 484
43. N. Maeda, S. Takeuchi, Inst. Phys. Conf. Ser. **104**, 303 (1989)
44. M. Skowronski, S. Ha, J. Appl. Phys. **99**, 011101 (2006)
45. K. Maeda, K. Suzuki, S. Fujita, M. Ichihara, S. Hyodo, Philos. Mag. A **57**, 573 (1988)
46. H. Lendenmann, F. Dahkquist, N. Johansson, R. Söderholm, P.A. Nilsson, J.P. Bergman, P. Skytt, Mater. Sci. Forum **353–356**, 727 (2001)
47. S. Ha, M. Slowronski, J.J. Sumakeris, M.J. Peisley, M.K. Das, Phys. Rev. Lett. **92**, 175504 (2004)
48. S. Takeuchi, K. Suzuki, K. Maeda, H. Iwanaga, Philos. Mag. A **50**, 171 (1984)
49. M.H. Hong, A.V. Samant, P. Pirouz, Philos. Mag. A **80**, 919 (2000)
50. P. Pirouz, J.L. Demenet, M.H. Hong, Philos. Mag. A **81**, 1207 (2001)
51. T. Miyanagi, H. Tsuchida, I. Kamata, T. Nakamura, K. Nakayama, R. Ishii, Y. Sugawara, Appl. Phys. Lett. **89**, 062104 (2006)
52. M.S. Miao, S. Limpijumnong, W.R.L. Lambrecht, Appl. Phys. Lett. **79**, 4360 (2001)
53. B. Chen, T. Sekiguchi, T. Ohyanagi, H. Matsuhata, A. Kinoshita, H. Okumura, J. Appl. Phys. **106**, 074502 (2009)
54. T.A. Kuhr, J. Lin, H.J. Chung, M. Skowronski, F. Szmulowicz, J. Appl. Phys. **92**, 5863 (2002)
55. W.R.L. Lambrecht, M.S. Miao, Phys. Rev. B **73**, 155312 (2006)
56. R. Hirano, Y. Sato. M. Tajima, K.M. Itoh, K. Maeda, Mater. Sci. Forum. **395**, 717–720 (2012)
57. A.T. Blumenau, C. Fall, R. Jones, S. Öberg, T. Frauenheim, P.R. Briddon, Phys. Rev. B **68**, 174108 (2003)

58. G. Savini, G. Savini, A. Marocchic, I. Suarez-Martineza, G. Haffendena, M. Heggie, S. Öberg, Physica B **401–402**, 62 (2007)
59. G. Savini, Phys. Stat. Sol. C. **4**, 2883 (2007)
60. T. Fujimoto, T. Aigo, M. Nakabayashi, S. Sato, M. Katsuno, H. Tsuge, H. Yashiro, H. Hirano, T. Hoshino, W. Ohashi, Mater. Sci. Forum **645–648**, 319 (2010)
61. P. Petroff, L.C. Kimerling, Appl. Phys. Lett. **29**, 461 (1976)
62. O. Ueda, S. Isozumi, S. Yamakoshi, T. Kotani, J. Appl. Phys. **50**, 765 (1979)
63. S. O'Hara, P.W. Hutchinson, P.S. Dobson, Appl. Phys. Lett. **30**, 368 (1977)
64. R. Hirano, Y. Sato, H. Tsuchida, M. Tajima, K.M. Itoh, K. Maeda, App. Phys. Exp. **5** (2012) in press

# Chapter 10
# Mechanism of Defect Reactions in Semiconductors

**Yuzo Shinozuka**

**Abstract** We discuss mechanisms of defect reactions (defect creation, annihilation, multiplication, reconstruction, impurity diffusion, etc.) in semiconductors starting with the origin of the electron-lattice interactions. The key mechanism of extrinsic self-trapping is introduced to understand strong carrier localization and large accompanied lattice distortion. Symmetry-breaking structural instability is explained in connection with the Jahn-Teller effect. It is found that the tetrahedral coordination is not so rigid as it is considered, especially for hole localization. Next, we explain the true meaning of the configuration coordinate diagram (CCD), which is usually misused in literature. The multiphonon carrier capture and the following induced lattice relaxation processes are discussed using proper CCD for deep-level defects. Finally various mechanisms of defect reactions are discussed including instability mechanism and phonon-kick mechanism. Energy released by a carrier capture can enhance the next capture, and then for high carrier density, a positive feedback may occur to create a rapid increase of lattice vibrations, which then causes defect reactions.

## 10.1 Introduction

In this chapter, we discuss the mechanisms of defect reactions in semiconductors. Since defect reactions are nothing but atomic movements around defect(s), they are enhanced by thermal excitations overcome the potential barrier $\Delta U$, showing the Arrhenius-type probability ~exp $(-\Delta U/k_{\mathrm{B}}T)$. On the other hand, it has been shown that defect reactions are enhanced by electronic excitations such as irradiation with photon, X-ray, $\gamma$-ray, and electron, and also carrier injection. A carrier trapping by a

Y. Shinozuka (✉)
Department of Materials Science and Chemistry, Faculty of Systems Engineering, Wakayama University, Sakaedani 930, Wakayama 640-8510, Japan
e-mail: yuzo@sys.wakayama-u.ac.jp

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_10,
© Springer Science+Business Media New York 2013

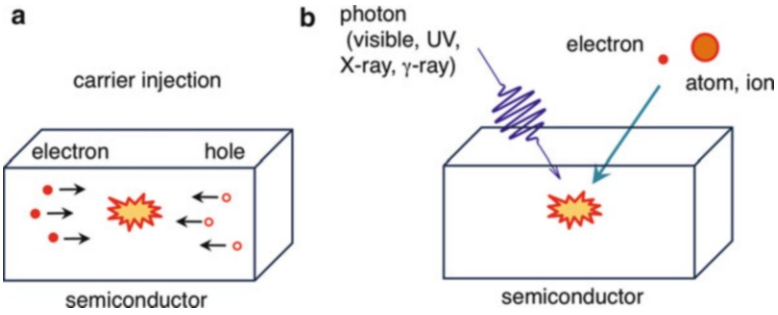



**Fig. 10.1** (**a**) Intrinsic and (**b**) extrinsic triggers for defect reactions in semiconductors

defect sometimes induces a large atomic displacement around it, resulting structural change of a defect, bond breaking, creation of another defect, desorption of an atom from a surface, and so on [1]. Some substitutional impurities in III–V semiconductors are found to show the off-center instability [2, 3], which is induced by a change in the impurity charge, and invalidate the traditional common sense of the valence control using shallow impurity doping. Degradations of light-emitting semiconductor diodes and lasers are caused by the climb and glide motions of dislocations [4, 5]. Maeda et al. found that dislocation movements are enhanced by carrier injection [6]. Control of defect reactions in electronic devices in operation is crucial to achieve their high efficiency and reliability. Above these reasons, we will mainly discuss mechanisms of electronically enhanced defect reactions.

Electronically enhanced defect reactions in semiconductor devices are phenomenologically classified into two categories: (a) intrinsic and (b) extrinsic (Fig. 10.1). The former is originated from carriers (conduction electron and valence hole) in devices, and the latter is originated from irradiations from outside.

### *10.1.1 Intrinsic*

This is crucial for light-emitting devices such as semiconductor laser diodes, where carrier concentration is very high. Captures and recombinations of carriers at point defects, dislocation, interfaces, etc., are considered as a trigger of defect reactions [4, 5]. This chapter will be mainly concerned with these intrinsic origins.

### *10.1.2 Extrinsic*

This is crucial for electronic devices in space satellites and light-harvesting devices such as solar cells. When a material is irradiated with either light (visible, UV, X-ray, γ-ray), electron beam, or ion beam, energy transfer and momentum transfer to the material take place, which sometimes result in degradation. Degradations

generated by extrinsic origins can be partially prevented by shielding over the devices; however, it cannot be done for light-harvesting devices. The mechanism of these extrinsic origins will be discussed in Sect. 10.5.

As will be explained later, both intrinsic and extrinsic origins are closely related with each other because they both generate electronic excitations (transitions) and hence induce strong electron-lattice interactions in materials. So we will start this chapter with a review on the electron-lattice interaction in Sect. 10.2. The key mechanism of extrinsic self-trapping of a carrier is introduced to explain why strong electron-lattice interaction accompanied by a strong localization of a carrier takes place in semiconductors. Symmetry-breaking structural changes at point defects are discussed, recalling the Jahn-Teller effect. Because of strong localization of electrons at deep-level defects, the effects of the electron-electron interaction become important as well as the electron-lattice interaction. The cooperation and competition of two interactions at point defects are discussed using a tetrahedral T-U-S model. Section 10.3 starts with the key mechanism of adiabatic instability using configuration coordinate diagram (CCD), then we discuss the lattice relaxation dynamics: how the energy transfers from electronic to atomic. With a special remark on CCD for deep-level defects, the nonradiative multiphonon recombination mechanism is discussed. In Sect. 10.4, proposed mechanisms so far on the electronically enhanced (induced) defect reactions are reexamined including the structural instability and the recombination enhanced. The mechanism of the energy conversion during the reaction is discussed paying attention to the relation between the lattice relaxation mode and the symmetry-breaking reaction coordinate. In Sect. 10.5, we briefly discuss the extrinsic origins for the defect reactions, and Sect. 10.6 is a summary of this chapter.

## 10.2 Electron-Lattice Interaction in Semiconductors

The present semiconductor technology is based on a physical hypothesis where carriers (conduction electron and valence hole) move in an almost static lattice and sometimes interact with crystal imperfections: static (vacancy, impurity, etc.) and dynamical (lattice vibration). It has long been considered that the electron-lattice interaction in semiconductors can be treated as a small perturbation. Recently, however, there have been found various phenomena that break this hypothesis, as shown above. We are going to explain why the electron-lattice interaction can be treated as a small perturbation in some cases and cannot in other cases. In order to explain the reason, we have to return the origin of the electron-lattice interaction in semiconductors.

### *10.2.1 Many-Electron Scheme for Condensed Matter Physics*

The atomic motion in condensed matter is strongly affected by the electronic state and vice versa. In order to discuss this interaction, we usually employ one of the following three schemes:

(A) All electrons + all nuclei
(B) All valence electrons + all ions
(C) Several conduction electrons and valence holes + lattice

Although the scheme (C) is popular in semiconductor physics, we will employ scheme (B) hereafter in order to clarify the scheme (C), with which we sometimes fall in incorrect results [7].

There are many (~$10^{23}$) valence electrons and many (~$10^{23}$) ions in a material, and they interact with each other. The Hamiltonian can be formally written as

$$H = H_{\mathrm{e}}(r_1, r_2, \ldots) + V_{\mathrm{eI}}(r_1, r_2, \ldots, R_1, R_2, \ldots) + H_{\mathrm{I}}(R_1, R_2, \ldots), \tag{10.1}$$

where

$$H_{\mathrm{e}}(r_1, r_2, \ldots) = \sum_i \frac{p_i^2}{2m} + V_{\mathrm{e}}(r_1, r_2, \ldots) \tag{10.1a}$$

$$H_{\mathrm{I}}(r_1, r_2, \ldots) = \sum_j \frac{P_j^2}{2M_j} + V_{\mathrm{I}}(R_1, R_2, \ldots). \tag{10.1b}$$

Here, $r_i$ and $p_i$ are, respectively, the position and the momentum of the $i$th electron with mass $m$. $R_j$ and $P_j$ are, respectively, the position and the momentum of the $j$th ion with mass $M_j$. The terms $V_{\mathrm{e}}$, $V_{\mathrm{I}}$, and $V_{\mathrm{eI}}$ denote, respectively, the interactions among electron-electron, ion-ion, and electron-ion. Because $M_j$ is much heavier than $m$, the adiabatic approximation can be applied to separate the electronic motion and the ionic motion. First, we fix the ionic positions $\{R\} \equiv (R_1, R_2, \ldots)$ and solve the Schrödinger equation for the electrons:

$$\begin{aligned}&[H_{\mathrm{e}}(r_1, r_2, \ldots) + V_{\mathrm{eI}}(r_1, r_2, \ldots, \{R\})]\Psi_{l,\{R\}}(r_1, r_2, \ldots) \\ &\quad = E_l(\{R\})\Psi_{l,\{R\}}(r_1, r_2, \ldots).\end{aligned} \tag{10.2}$$

Here, $\Psi_{l,\{R\}}(r_1, r_2, \ldots)$ is the $l$th eigenstate with the eigen-energy $E_l(\{R\})$ of the electrons for the fixed ionic configuration $\{R\}$. The adiabatic potential of the ions for the $l$th eigenstate of the electrons is given by

$$U_l(\{R\}) = E_l(\{R\}) + V_I(\{R\}), \tag{10.3}$$

which is schematically shown in Fig. 10.2. It should be noted that along the same $l$th branch of $U_l(\{R\})$, the electronic wave function $\Psi_{l,\{R\}}(r_1, r_2, \ldots)$ changes continuously with the change of the ionic configuration $\{R\}$.

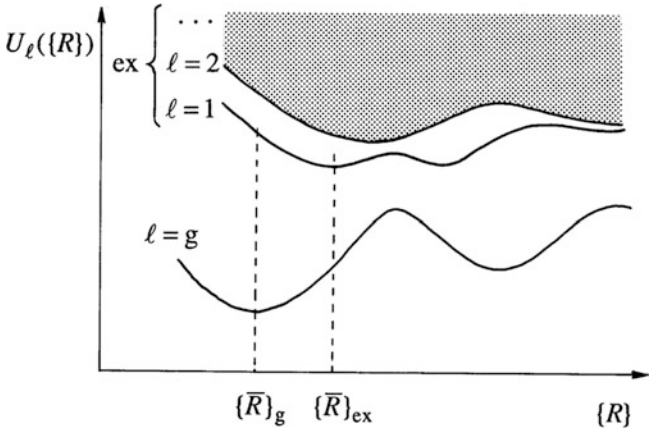



**Fig. 10.2** Schematic representation of the adiabatic potentials for condensed matter

#### 10.2.1.1 Perfect Crystal

For a macroscopic and electrically neutral system, the minimum $\{\bar{R}\}$ of $U_g(\{R\})$ for the ground state $\Psi_{g,\{R\}}(r_1, r_2, \ldots)$ gives a crystal structure that has a translational symmetry. The type of the realized crystal structure depends on the constituent of the material. It is the diamond structure for silicon, the zincblende for GaAs, etc. For carbon, besides the ground graphite structure, there are several metastable structures: diamond, $C_{60}$, nanotube, etc. The one-electron eigenstates in the crystal are characterized by a wave vector as a result of the Bloch theorem and are classified into the conduction band and the valence band. In semiconductors and insulators, the former is empty and the latter is fully occupied by electrons in the ground state.

For a given crystal structure$\{\bar{R}\}_g$, each ion can vibrate around each equilibrium position $\bar{R}_j$ due to the kinetic energy of ions. For a small ionic displacement, $\Delta R_j \equiv R_j - \bar{R}_j$, much less than the lattice constant $d$, one can expand the adiabatic potential $U_g(\{R\})$ in a power series as

$$U_g(\{\Delta R\}) = U_g(\{\bar{R}\}) + \sum_j \frac{\partial}{\partial \Delta R_j} U_g(\{R\})\bigg|_{\{\bar{R}\}_g} \Delta R_j + \sum_{j'} \sum_j \frac{\partial^2}{\partial \Delta R_{j'} \partial \Delta R_j} U_g(\{R\})\bigg|_{\{\bar{R}\}_g} \Delta R_{j'} \Delta R_j + \cdots \tag{10.4}$$

Since the coefficient of the second term is zero at the stable position, the harmonic approximation can be applied, and the quadratic terms can be

diagonalized by transforming $\{\Delta R\}$ into the normal modes $\{Q\} \equiv Q_{\sigma,k}$ of the lattice vibrations, as follows:

$$\Delta R_j \propto \sum_k Q_{\sigma,k} \exp(i\, k \cdot \bar{R}_j), \tag{10.5}$$

$$U_{\mathrm{g}}(\{\Delta R\}) = U_{\mathrm{g}}(\{Q\}) = \text{const.}(=0) + \frac{1}{2}\sum_{\sigma,k} \omega_{\sigma,k}^2 Q_{\sigma,k}^2, \tag{10.6}$$

$$\begin{aligned}\left\langle \Psi_{\mathrm{g},\{R\}} \middle| H \middle| \Psi_{\mathrm{g},\{R\}} \right\rangle &= \sum_j \frac{P_j^2}{2M_j} + U_{\mathrm{g}}(\{\Delta R\}) = \frac{1}{2}\sum_{\sigma,k}\left(P_{\sigma,k}^2 + \omega_{\sigma,k}^2 Q_{\sigma,k}^2\right) \\ &\equiv H_{\mathrm{L}}. \end{aligned} \tag{10.7}$$

Here, $k$ represents the wave vector and $\sigma$ the branch index and $\omega_{\sigma,k}$ the angular frequency. The normal modes are typically classified into the acoustic mode ($\sigma =$ ac) and the optical modes ($\sigma =$ op). Thus, in the ground state, the interaction between many valence electrons and ions in scheme (B) results in the lattice vibration $H_{\mathrm{L}}$. There is no electron-lattice interaction in the ground state.

At finite temperature, the lattice vibrations are excited thermally. When temperature is high, atoms happen to overcome the potential barrier for reaction, which have been neglected in the harmonic approximation (10.6). For crystal structures, the thermal melting takes place when the average atomic displacement is roughly equal to the lattice constant, or the thermal energy $k_{\mathrm{B}}T$ is equal to the potential barrier height $\Delta U$. If there is a defect in the crystal, $\Delta U$ around the defect is different from that of the host, then an atomic rearrangement occurs at lower than melting temperature: for example, for interstitial atoms.

## *10.2.2 Interaction Hamiltonian*

If a carrier (conduction electron, valence hole) is created by a band-to-band photoexcitation or a carrier injection in a semiconductor, the balance of the interatomic (ionic) forces mediated by valence electrons is disturbed, and then the atomic configuration is rearranged. This is the origin of the electron-lattice interaction [7].

### 10.2.2.1 Conduction Electron

Let us introduce an extra electron into the conduction band. This is an $N + 1$ electron problem. Denoting $r_{\mathrm{e}}$ as the position of an extra electron, one can approximate the total electronic state as

$$\left|\Psi_{l,\{R\}}^{N+1}\right\rangle \cong \left|\phi_{l,\{R\}}(r_{\mathrm{e}})\right\rangle \left|\Psi_{\mathrm{g},\{R\}}^{N}\right\rangle. \tag{10.8}$$

Instead of the one-electron eigenstate $\phi_{l,\{R\}}(r_e)$, it is usual to employ a Wannier state $\phi_{\bar{R}_j}(r_e)$ localized at the position $\bar{R}_j$, which is a linear combination of the Bloch states of the conduction band. The effective Hamiltonian of this situation is then given by

$$H = h_e + V_{eL} + H_L, \tag{10.9}$$

$$h_e = \frac{p_e^2}{2m_e^*}, \quad H_L = \frac{1}{2}\sum_{\sigma,k}(P_{\sigma,k}^2 + \omega_{\sigma,k}^2 Q_{\sigma,k}^2).$$

Here, $m_e^*$ is the effective mass of a conduction electron. For the electron-lattice interaction term, replacing the discrete variable $\bar{R}_j$ by a continuous variable $r_e$, we obtain the following formula:

$$V_{eL} = -\sum_{\sigma,k} A_{\sigma,k}^c Q_{\sigma,k} \exp(ik \cdot r_e). \tag{10.10}$$

The interaction with the acoustic mode is shown to be of the deformation potential type

$$A_{ac,k}^c = \left(\frac{2}{NM}\right)^{1/2} E_d^c k, \tag{10.11}$$

and the interaction with the optical mode in ionic crystals is the Fröhlich type [8]

$$A_{op,k}^c = \left[\frac{2\pi\omega_{op}^2 e^2}{V}\left(\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_0}\right)\right]^{1/2} k^{-1}. \tag{10.12}$$

#### 10.2.2.2 Valence Hole

If we subtract an electron from the valence band, it becomes an $N - 1$ electron system. We can obtain a similar effective Hamiltonian for a hole. For details, see Ref. [7].

### *10.2.3 Polaron (Self-trapping)*

An extra electron (hole) in the conduction (valence) band thus interacts with the lattice vibrations. What is the ground state of this system? This is the so-called polaron problem, which has been intensively studied. It has been shown that the character of a polaron is classified into two types: large polaron and small polaron [8, 9]. Large polaron is realized for weak electron-lattice interaction, where the

wave function of an electron extends over large area and accompanied lattice distortion is weak. Small polaron is realized for strong electron-lattice interaction, where the wave function of an electron is localized within a small area and accompanied lattice distortion is strong.

Let us study the character of a polaron using a continuum model proposed by Toyozawa [10], where the lattice is regarded as a superposition of the elastic medium and the dielectric medium, representing the acoustic phonon field and the optical phonon field, respectively. Suppose an electron is localized around a certain point in a semiconductor with a wave function,

$$\varphi(r) \sim \exp\left(-\pi \frac{r^2}{a^2}\right), \tag{10.13}$$

where $a$ is the extension. The expectation value of the Hamiltonian (10.9) is given by

$$\begin{aligned} &\langle\varphi(r)|\frac{-\hbar^2}{2m_{\mathrm{e}}^*}\Delta|\varphi(r)\rangle \\ &\quad - \sum_{\sigma,k}\left\{\langle\varphi(r)|A_{\sigma,k}^c \exp(ik\cdot r_{\mathrm{e}})|\varphi(r)\rangle Q_{\sigma,k} + \frac{1}{2}\omega_{\sigma,k}^2 Q_{\sigma,k}^2\right\}. \end{aligned} \tag{10.14}$$

After each lattice mode has been displaced from zero to each equilibrium position, $\bar{Q}_{\sigma,k}^c = \langle\varphi(r)|A_{\sigma,k}^c \exp(ik\cdot r_{\mathrm{e}})|\varphi(r)\rangle \Big/ \omega_{\sigma,k}^2$, the resulting total lattice relaxation energy $\sum_{\sigma,k}\omega_{\sigma,k}^2 \bar{Q}_{\sigma,k}^{c\,2}/2$ is given by

$$E_{\mathrm{LR}}^{\mathrm{ac}}(a) = \frac{E_d^2}{2Ca^3}, \quad E_{\mathrm{LR}}^{\mathrm{op}}(a) = \frac{e^2}{\varepsilon\, a}. \tag{10.15}$$

Adding the increase of the electron kinetic energy, the total energy of the system is shown to be

$$E(a) = B\left(\frac{d}{a}\right)^2 - E_{\mathrm{LR}}^{\mathrm{ac}}\left(\frac{d}{a}\right)^3 - E_{\mathrm{LR}}^{\mathrm{op}}\left(\frac{d}{a}\right), \tag{10.16}$$

where $d$ is the lattice constant of the crystal. The parameters here are $B = 3\pi\hbar^2/2m_{\mathrm{e}}^* d^2$, $E_{\mathrm{LR}}^{\mathrm{ac}} = E_{\mathrm{LR}}^{\mathrm{ac}}(d)$, and $E_{\mathrm{LR}}^{\mathrm{op}} = E_{\mathrm{LR}}^{\mathrm{op}}(d)$ where $E_d$ is the deformation potential, $C$ is the elastic constant, and $\varepsilon^{-1} = \varepsilon_\infty^{-1} + \varepsilon_0^{-1}$ with $\varepsilon_0$ and $\varepsilon_\infty$ are the static and high-frequency dielectric constants, respectively. For a typical semiconductor with $E_d = 6$ eV, $C = 4\times 10^{10}\mathrm{N/m}^2$, $\varepsilon = 4$, and $d = 3$A, the (maximum) lattice relaxation energies are $E_{\mathrm{LR}}^{\mathrm{op}}(d) = 1.2$ eV and $E_{\mathrm{LR}}^{\mathrm{ac}}(d) = 2.7$ eV, for example. The degree of localization $d/a$ is allowed to be $0 \le d/a \le 1$ since the wave function (10.13) is an envelope function of Wannier orbitals. The realized ground state, $\bar{a}$, is obtained by a minimization of (10.16) in the range $0 \le d/a \le 1$. The value of $\bar{a}$ depends on the strength of the electron-lattice interactions: $\vec{E}_{\mathrm{LR}}/B = \left(E_{\mathrm{LR}}^{\mathrm{ac}}/B,\ E_{\mathrm{LR}}^{\mathrm{op}}/B\right)$.

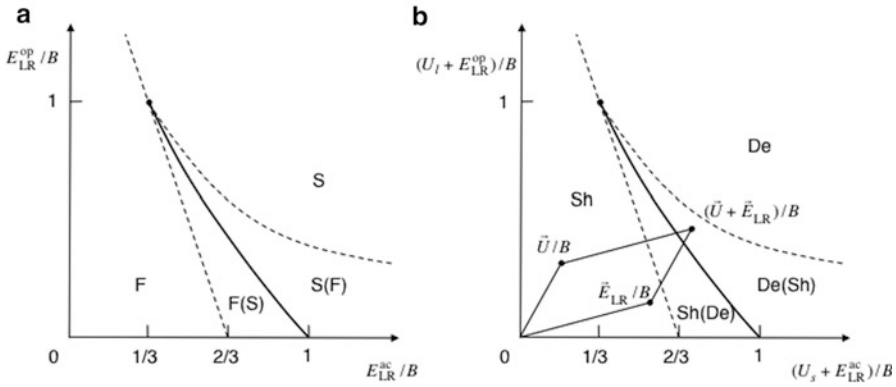



**Fig. 10.3** (**a**) The phase diagram of an electron interacting with the lattice. (**b**) The phase diagram of an electron interacting with the lattice and a point defect

Figure 10.3a shows the phase diagram of an electron in the phonon fields. In the region F, the extension of the electron wave function is large ($\bar{a} \gg d$) with small lattice distortion; thus, large polaron is realized, where an electron almost freely moves through the crystal. In the region S, an electron is strongly localized ($\bar{a} = d$) with large lattice distortion; thus, small polaron is stable, where an electron is trapped by self-induced lattice distortion at some point in the crystal (self-trapped electron). If there is a metastable state, it is indicated in the parenthesis. It should be noted that there is a discontinuous transition line in the phase diagram. If the material parameter crosses the line, the character of a polaron discontinuously changes from F to S or S to F.

### *10.2.4 Extrinsic Self-trapping*

Next, let us consider the effect of e-L interaction to an electron bound to a point defect. Adding a spherical impurity potential $U(r)$ to the Hamiltonian (10.9), the total energy is

$$E(a) = \langle \varphi(r)|H|\varphi(r)\rangle = \langle \varphi(r)|h_e + V_{eL} + H_L + U|\varphi(r)\rangle. \tag{10.17}$$

If the defect potential $U(r) = U_s(r) + U_l(r)$ can be decomposed as

$$\text{the long-range part} \quad U_l(r) = \frac{-U_l d}{r}, \tag{10.18a}$$

$$\text{the short-range part} \quad U_s(r) = \begin{cases} -U_s\ (=\text{const.}) & 0 \le r \le d \\ 0 & d < r \end{cases}, \tag{10.18b}$$

the expectation value of the total energy for the function (10.13) is approximately given by

$$E(a) = B\left(\frac{d}{a}\right)^2 - \left(U_s + E_{\mathrm{LR}}^{\mathrm{ac}}\right)\left(\frac{d}{a}\right)^3 - \left(U_l + E_{\mathrm{LR}}^{\mathrm{op}}\right)\left(\frac{d}{a}\right). \tag{10.19}$$

The stable electron radius $\bar{a}$ now depends on two vector strengths for electron localization: intrinsic $\vec{E}_{\mathrm{LR}}/B = \left(E_{\mathrm{LR}}^{\mathrm{ac}}/B,\ E_{\mathrm{LR}}^{\mathrm{op}}/B\right)$ and extrinsic $\vec{U}/B = (U_s/B,\ U_l/B)$. Figure 10.3b shows the phase diagram of an electron interacting with the lattice and a point defect. In the region Sh, an electron is shallowly localized around a point defect ($\bar{a} \gg d$) with small lattice distortion. In the region De, it is deeply localized ($\bar{a} = d$) with strong lattice distortion. If there is a metastable state, it is indicated in the parenthesis. For a typical donor in nonpolar semiconductors such as P in Si, $E_{\mathrm{LR}}^{\mathrm{op}} \sim 0, U_s \sim 0, \varepsilon \gg 1$, and $m_{\mathrm{e}}^*$ is light, then $\bar{a} \gg d$ and $E_{\mathrm{LR}}^{\mathrm{ac}}(\bar{a}) \sim 0$ in spite of large $E_{\mathrm{LR}}^{\mathrm{ac}}(d)$. As a result, the effect of the electron-lattice interaction can be negligible for shallow donor (acceptor) impurities in Si. On the other hand, it is possible that even if both $\vec{E}_{\mathrm{LR}}/B$ and $\vec{U}/B$ locate in the region Sh, but the vector sum $(\vec{E}_{\mathrm{LR}} + \vec{U})/B$ locates in the region De or De (Sh), as is shown in Fig. 10.3b, then the ground state is a deep state with strong lattice distortion. This mechanism is called the extrinsic self-trapping [11, 12]. Here, a strong electron localization is a trigger and also a result of a strong lattice distortion. A nitrogen impurity in Si [13, 14] might be an example of this extrinsic self-trapping (De (Sh)), which accompanies a metastable shallow state with small lattice distortion. Thus, we have shown here that strong electron-lattice interaction, triggered by the short-range interaction with acoustic phonons (see the $a$ dependence in Eq. 10.15), induces a shallow-deep instability at point defects in semiconductors.

### *10.2.5 Symmetry Breaking*

So far, we have used a continuum model and assumed a spherical symmetry in electron localization. When an electron is strongly localized, where the extension $a$ of the wave function is the order of the lattice constant $d$, it feels the lattice structure and also interacts with other valence electrons. Then we must explicitly consider the discrete lattice structure and its symmetry to discuss the lattice distortion in real space. Several point defects in semiconductors show off-center instability, such as EL2 center in GaAs and DX center in AlGaAs [15]. These instabilities can be understood in terms of the Jahn-Teller (J-T) effect, which has been first proposed to discuss the stability of molecular structures [16]. When the atomic structure of a point defect is symmetric, the localized electronic states are often degenerate. The J-T theory tells us that for degenerate electronic states, the structure is unstable and atoms are always displaced to lower the symmetry.

Let us consider a point defect in a semiconductor, whose point symmetry is characterized by a symmetry group denoted by **G**. The adiabatic Hamiltonian of the system is given by

$$H = H_{\mathrm{e}}(\{r\}) + V(\{r\},\{R\}) + U_{\mathrm{L}}(\{R\}). \tag{10.20}$$

Here, $\{r\}$ denotes the set of the position of electrons $\{r_1, r_2, \ldots\}$, and $\{R\}$ represents the displacement $\{R_0, R_1, \ldots\}$ of atoms around the defect. We must distinguish the relevant electrons and the other valence electrons that are combined with all ions to form the lattice medium [7]. $U_{\mathrm{L}}(\{R\})$ is the elastic energy for the displacements of atoms. The set of displacements $\{R\}$ can be transformed into the set of normal modes $\{Q_\Gamma\}$ in the irreducible representation $\Gamma$ of **G** [17]. The electron-lattice interaction term $V(\{r\}, \{R\}) = V(\{r\}, \{Q_\Gamma\})$ can be expanded as a power series of $R_j$'s or $Q_\Gamma$'s:

$$\begin{aligned} V(\{r\},\{R\}) &= V(\{r\},\{R\}=0) + \sum_j \frac{\partial V}{\partial R_j} R_j + \cdots \\ &= V(\{r\},\{Q_\Gamma\}=0) + \sum_\Gamma \frac{\partial V}{\partial Q_\Gamma} Q_\Gamma + \cdots . \end{aligned} \tag{10.21}$$

In the rigid lattice $\{R\} = 0$, all electronic states $\Psi(\{r\})$ which obey the Schrödinger equation

$$[H_{\mathrm{e}}(\{r\}) + V(\{r\},\{R\}=0)]\Psi(\{r\}) = E\Psi(\{r\}) \tag{10.22}$$

are classified by the irreducible representation $\Gamma_{\mathrm{e}}$ of **G** as $\Psi_{\Gamma \mathrm{e}}(\{r\})$.

Next, we allow the displacements of atoms. The generalized force $F^{\Gamma}_{\Gamma_{\mathrm{e}}}$ with respect to $Q_\Gamma$ for the electronic state $\Psi_{\Gamma \mathrm{e}}(r)$ is given by

$$F^{\Gamma}_{\Gamma_{\mathrm{e}}} = \left\langle \Psi_{\Gamma_{\mathrm{e}}} \left| \frac{\partial V}{\partial Q_\Gamma} \right| \Psi_{\Gamma_{\mathrm{e}}} \right\rangle. \tag{10.23}$$

If $\Psi_{\Gamma \mathrm{e}}(\{r\})$ is degenerate, the symmetric representation $[\Gamma_{\mathrm{e}} \times \Gamma_{\mathrm{e}}]$ always contains at least one nonidentity representation $\Gamma$ [17]. This means $F^{\Gamma}_{\Gamma_{\mathrm{e}}} \neq 0$, and then atoms are always displaced to lower the symmetry. Figure 10.4 shows an example of the J-T distortion at a square molecule of $D_{4h}$. When the electronic state is doubly degenerate $E_g$, it distorts to $D_{2h}$. It should be noted that there are several equivalent stable configurations: in Fig. 10.4, there are two equivalent J-T distortions ($Q_{\mathrm{B1g}} > 0$ and $Q_{\mathrm{B1g}} < 0$); thus, the total symmetry remains conserved in some sense. Table 10.1 shows typical examples of the relation between electronic states and the J-T distortion modes for some point symmetries. A vacancy ($V$) in the diamond structure has the $T_{\mathrm{d}}$ symmetry. In Si crystal, when a vacancy is positively charged $V_{\mathrm{Si}}{}^{+}$, the electronic ground state is triply degenerate ($\Gamma = T$), then the E-type J-T distortion takes place to lower the symmetry to $D_{2d}$, which has triple

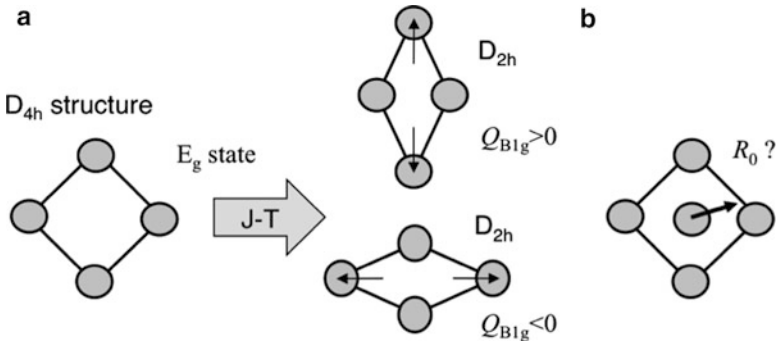



**Fig. 10.4** An example of the Jahn-Teller effects. (**a**) $D_{4h}$ structure is unstable for the $E_g$-type electronic states and then distorts to $D_{2h}$. (**b**) Does the J-T distortion contain the displacement of the central atom?

equivalent minimums. This J-T distortion at $V_{Si}^{+}$ has been observed experimentally by EPR measurements [18], and the potential barrier height between equivalent minimums is estimated as 0.07 eV [19].

So far we have treated the atomic motion classically. As has been stated in the original paper [16], the J-T theorem only gives us the initial instability but does not give the magnitude of the atomic distortion and hence that of $E_{LR}$. When the J-T lattice relaxation energy $E_{LR}$ is smaller than the average phonon energy, $\hbar\omega$, quantum transitions occur rapidly among the equivalent minimums of the atomic configuration. If the mean transition time is shorter than the characteristic time of the experiment, the average of the symmetry lowering is zero (the dynamic J-T effect). This explains why the J-T effect is practically absent at shallow-level impurities. On the other hand, if the electron wave function is localized, as is in deep-level defects, electrons interact strongly with the lattice, then $E_{LR} \gg \hbar\omega$ and the present J-T instability really takes place and lowers the symmetry for degenerate electronic states.

### *10.2.6 Stability of the Position of an Impurity*

Let us next apply the above Jahn-Teller theorem to the stability of positions of an impurity atom in semiconductors, either interstitial or substitutional. In the former case, an impurity atom may occupy a position with larger empty space and hence has a high point symmetry. Let us denote the symmetry of the position as **G**. If the electronic state is degenerate, atoms around the impurity are always displaced to lower the symmetry due to the J-T effect as shown above. The point now is whether these J-T displacements contain the displacement $R_0$ of the impurity atom located at the center (Fig. 10.4b). The electronic states usually couple many atoms, then the transformation of $\{R_j\}$ into $\{Q_\Gamma\}$ is practically impossible. On the other hand, it is

**Table 10.1** Typical examples of the Jahn-Teller distortion modes and the type of the displacement

| $T_d$ | | $D_{3d}$ | | $C_{3v}$ | |
|---|---|---|---|---|---|
| Electronic state $\Psi_{\Gamma e}$ | J-T distortion mode $Q_\Gamma$ | Electronic state $\Psi_{\Gamma e}$ | J-T distortion mode $Q_\Gamma$ | Electronic state $\Psi_{\Gamma e}$ | J-T distortion mode $Q_\Gamma$ |
| E | E | | | | |
| $T_1$ | E, $T_2$ | $E_g$, $E_u$ | $A_{2g}$, $E_g$ | E | E |
| $T_2$ | E, $T_2$ | | | | |
| $R_0$ belongs to $T_2$ | | $R_0$ belongs to $A_{2u}$ and $E_u$ | | $R_0$ belongs to $A_1$ and E | |
| $T_d$ position is unstable for $T_1$ and $T_2$ electronic states | | $D_{3d}$ position does not show the instability caused by J-T effect | | $C_{3v}$ position is unstable for E electronic states | |

$R_0$ of the center atom. E states are doubly degenerate and T states are triply degenerate

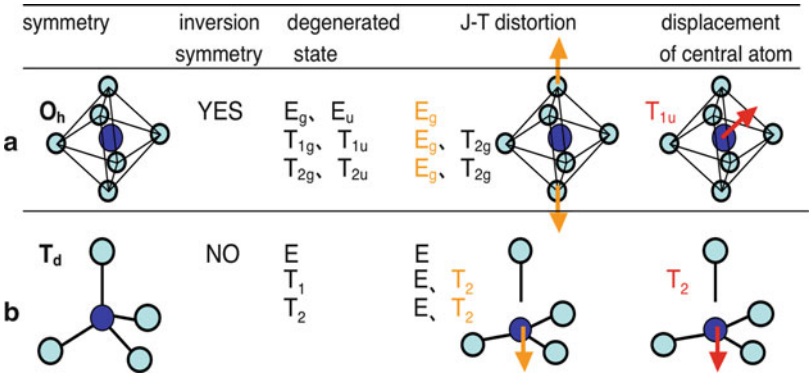



**Fig. 10.5** Defect molecules (**a**) with and (**b**) without inversion symmetry

much easy to find to which representation Γ('s) the central displacement $R_0$ belongs. The criterion of the instability of the positions of an impurity for several point symmetries has been discussed [20], and some of the results are summarized in Table 10.1. Thus, we have a useful information on the stability of an impurity position without any elaborate calculations. It should be noted that an impurity atom at any positions which have the inversion symmetry does not show the instability caused by the J-T effect. This is because the irreducible representations for any system with the inversion symmetry are classified into gerade and ungerade. All the J-T distortion modes are gerade, while the displacement of the central atom always belongs to ungerade modes. Examples are the interstitial hexagonal position ($D_{3d}$) in the diamond structure, the body-centered position ($O_h$), and the face-centered position ($D_{4h}$) in the simple cubic lattice (Fig. 10.5).

If electronic states $\Psi_{\Gamma e}(r)$ contain a spin freedom, the condition of the Jahn-Teller instability depends on the number of bound electrons. The criterion for even number electrons is whether the symmetric representation $[\Gamma_e \times \Gamma_e]$ contains the displacement $R_0$ of the impurity, while that for odd number is whether the

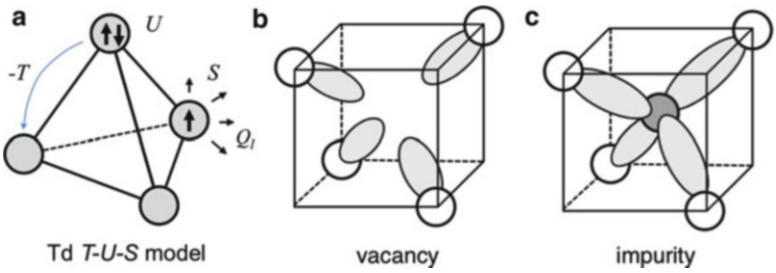



**Fig. 10.6** (**a**) Td *T-U-S* model for (**b**) vacancy and (**c**) impurity in semiconductors

antisymmetric representation $\{\Gamma_e \times \Gamma_e\}$ contains $R_0$ or not [20]. The present J-T instability for substitutional impurities corresponds to off-center instability. Examples are N in Si [13, 14], EL2 in GaAs [2, 3], and DX center in AlGaAs [15, 21]. It is surprising that although the $T_d$ is the most popular symmetry met in the diamond and the zincblende structures, it shows an inherent instability as shown here. Thus, the diamond structure is not so stable as we have been considered with respect to local electronic excitations.

### *10.2.7 $T_d$ T-U-S Model*

In case of strong carrier localization such as deep-level defects in semiconductors, we must take account of the interaction among localized electrons as well as the electron-lattice interaction. The simplest model to discuss these two effects would be the *T-U-S* model [22], which consists of several interacting electrons in a defect molecule. Figure 10.6 shows (a) the tetrahedral *T-U-S* model, which represents (b) a vacancy or (c) a substitutional impurity in the diamond (zincblende) lattice [23–25]. It is characterized by the electronic transfer energy, $-T$, describing the quantum hopping of an electron from site to site; the on-site Coulomb repulsion energy, $U$, between two electrons with opposite spins at the same site; the lattice relaxation energy, $S$, due to the short-range interaction of an electron with the lattice mode, $Q_l$, at the site $l$ (=1 ~ 4); and the number $n$ (=0 ~ 8) of the occupied electrons. The site $l$ represents either atomic or molecular orbital: (b) four dangling bonds for a vacancy and (c) four bonding (or antibonding) orbitals for a substitutional impurity. For case (c), $T$ is equal to the s-p hybridization energy $\varepsilon_p - \varepsilon_s$ of the impurity atom, and $Q_l$ stands for the bond stretching mode between the central impurity and $l$th host atom.

In the rigid lattice ($Q_l = 0$), when the electron-electron interaction is absent ($U = 0$), one-electron eigenstates are one $a_1$ state and three $t_2$ states. If we allow the lattice to deform ($Q_l \neq 0$) and the electron-electron interaction is switched on ($U \neq 0$), eigenstates of $n$ electrons become complicated, and structural changes may take place lowering the symmetry. The stable structures for $n$ (=0 ~ 8)

electrons and each spin state are comprehensively studied by El-Maghraby et al. [24] and Takahashi et al. [25]. The stable structures are found to depend on the relative strength among $T$, $U$, and $S$. Some of the results are as follows:

#### 10.2.7.1 Vacancy

The one-electron energy at a dangling bond is located in the middle of the band gap, and then it is a good approximation to separate electrons at dangling bonds with other host valence electrons. The lattice distortion $Q_l$ here represents a linear combination of the displacements of three host atoms next to the host atom at $l$ (Fig. 10.6b). The vacancy ($V$) in Si has been found to have different stable structures depending on its charge [18, 19]: $D_{2d}$ for $V^0$ and $V^+$, $C_{2v}$ for $V^-$, and $T_d$ for $V^{++}$. The neutral $V^0$ corresponds to $n = 4$ in the present model; the parameters for $V_{Si}$ may fall into the third case ($U/T = 1$, $S/T = 2.4$). Since only the diagonal-type electron-lattice interaction has been taken into account, the $D_{2d}$ type structure does not appear here. It is stabilized by the off-diagonal electron-lattice interaction: modulation of the electron transfer $T(Q_l)$.

#### 10.2.7.2 Substitutional Impurity

Figure 10.7 shows the phase diagrams of the stable structures for substitutional impurities, displayed in the $T$-$U$-$S$ triangle. The coordinates ($T$, $U$, $S$) of any point in this triangle are proportional to the lengths of perpendiculars therefrom to $US$, $ST$, and $TU$ sides. A singly ionized donor ($D^+$) and acceptor ($A^-$) correspond to fully occupied four bonding orbitals and empty four antibonding orbitals. Then, both $D^+$ and $A^-$ correspond to $n = 8$ or $n = 0$ in the present model. It is found that the $T_d$ structure is found to be always stable there. But when the electronic occupation is changed, a structural instability may take place, which is different for donor and acceptor as can be seen in Fig. 10.7. The displacements of the central impurity atom in real space are also illustrated. Neutral donor $D^0$ corresponds to $n = 1$, where fully occupied bonding orbitals have been included into the elastic medium. It is found that $T_d$ structure is stable for $S/T < 6$ and $C_{3v}(1)$ is stable for $S/T > 6$. Nitrogen impurity in Si [13, 14], which is found to show an off-center configuration for a neutral state, may correspond to the latter case. For $D^-$ donor, $C_{3v}(1)$-type off-center configuration is more stable when $S/T > 3$ and $S > U$. DX center in $Al_{1-x}Ga_xAs$ [15, 21] may correspond to this case: $T_d$ structure is stable for neutral donor and $C_{3v}(1)$ off-center structure for $DX^-$ state.

For neutral acceptor $A^0$ (corresponds to $n = 7$) on the other hand, $T_d$ is always unstable and $C_{3v}(2)$ type off-center structure is always stable. Then we can conclude that the on-center-type neutral deep-acceptor $A^0$ is absent in tetrahedrally bonded semiconductors. This is not the case for a shallow acceptor $A^0$, which should be regarded as $A^-$ + one shallowly bound hole in the present approximation. It should be reminded that we have used a crucial approximation where a "defect molecule" has been cut out of the crystal.

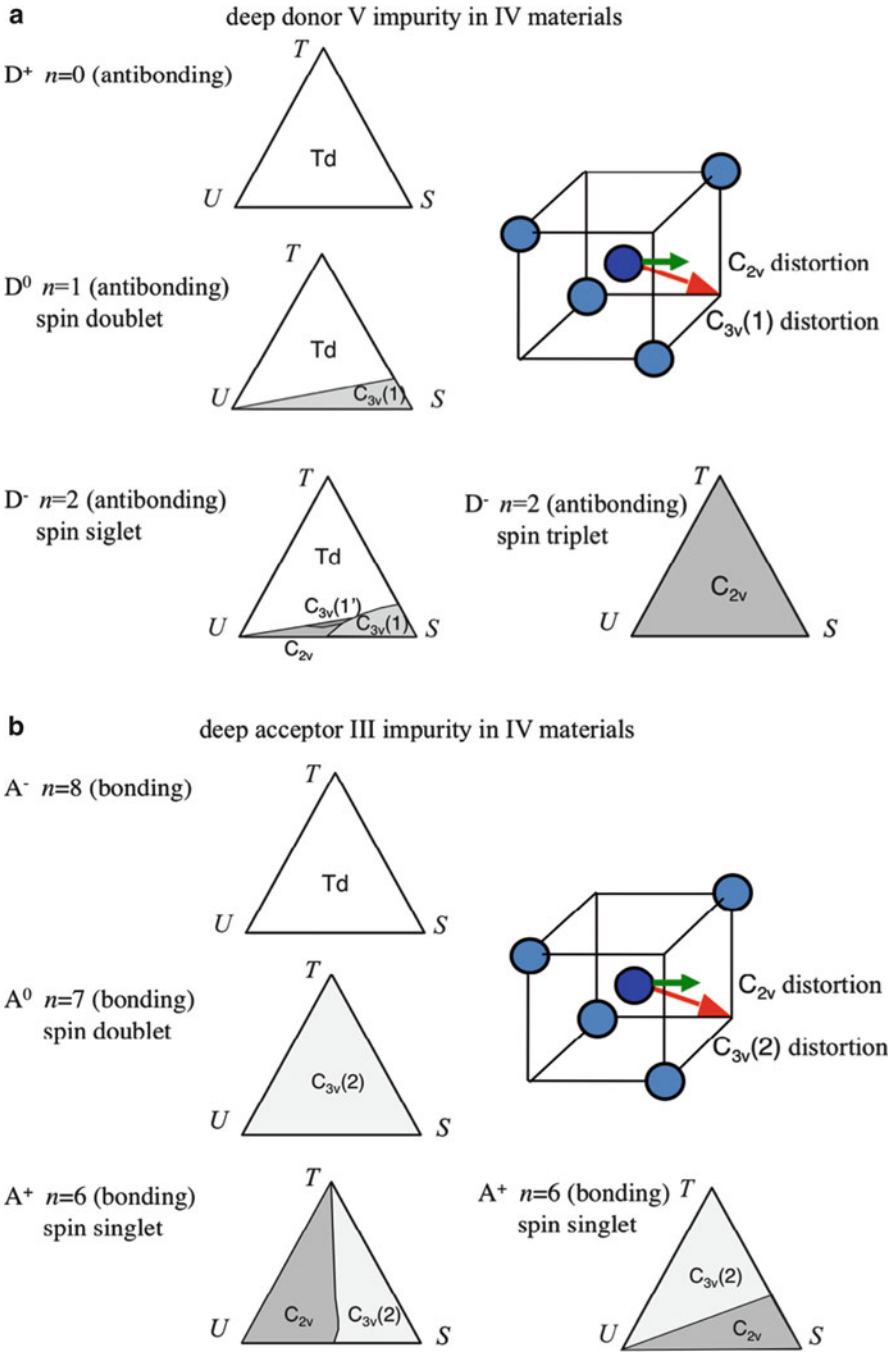



**Fig. 10.7** Phase diagrams of the ground-state structure for a substitutional deep (**a**) donor or (**b**) acceptor impurity in semiconductors [25]

## 10.3 Configuration Coordinate Diagram

### *10.3.1 Adiabatic Instability*

The configuration coordinate diagram (CCD) is a useful tool in discussing various electronic-atomic processes in condensed matter. In the CCD, the adiabatic potentials (=the total energy minus the atomic kinetic energy) are illustrated as a function of the configuration coordinates $\{Q_i\} \equiv Q$, which represents the positions of all the constituent atoms, hence multidimensional variables, basically ~$10^{23}$. In Fig. 10.8, we illustrate a CCD for a two-atom molecule, which is the most simple example of condensed matter. The lower branch of the adiabatic potential $U_g(Q)$ is for the electronic ground state $\Psi_g$, and the upper one $U_{ex}(Q)$ is for the excited state $\Psi_{ex}$ as a function of $Q$, the atomic distance. The minimum position $\bar{Q}$ of $U_g(Q)$ corresponds to the stable configuration for the total ground state, where the effective force $F_i^g$ for the $i$th ion is zero. On the other hand in the excited state, the effective force $F_i^{ex}$ at $\bar{Q}$ is nonzero because the charge density is different. Atoms feel induced force $F_i^{ex}$ and then move toward to a new stable position (adiabatic instability).

### *10.3.2 CCD for Two-Level System*

Next, let us discuss the lattice relaxation dynamics induced by an electronic transition. Figure 10.9a illustrates two-level system: for example, a bound electron in a semiconductor interacts with lattice vibrations. When an electron is in the

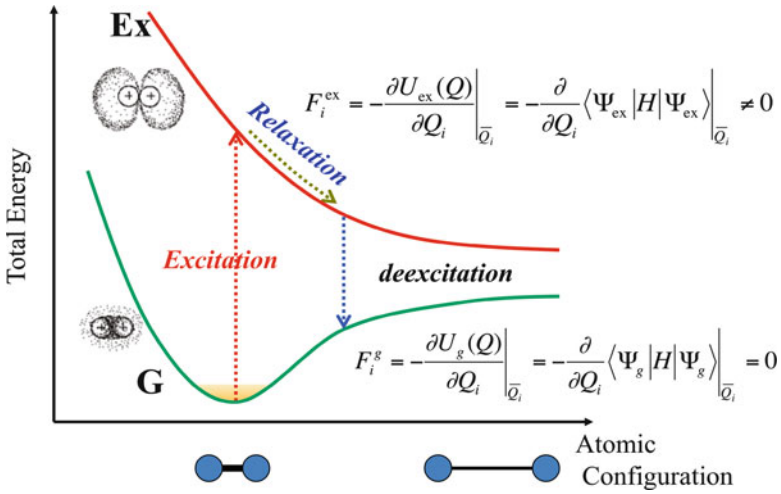



**Fig. 10.8** Typical CCD for a two-atom molecule

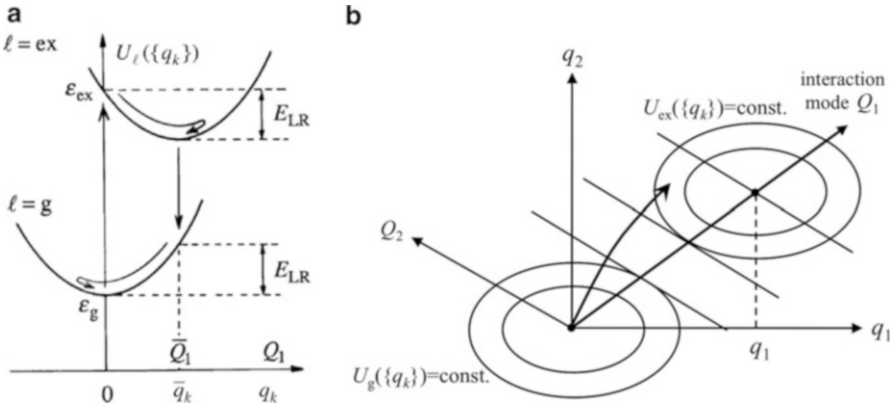



**Fig. 10.9** (**a**) The adiabatic potentials for two-level system. (**b**) Contour map of the adiabatic potentials

ground state $\Psi_g$ with electronic energy $\varepsilon_g$, the adiabatic potential is assumed to be given by

$$U_g(\{q_k\}) = \varepsilon_g + \frac{1}{2}\sum_k \omega_k^2 q_k^2. \tag{10.24}$$

Here, $q_k$ is the normal mode of the lattice vibration (defined in the electronic ground state) with angular frequency $\omega_k$. The subscript $k$ stands all for the wave vector, mode index, and localized mode if any. The adiabatic potential for the electronic excited state $\Psi_{ex}$ with $\varepsilon_{ex}$ is assumed to be given by

$$\begin{aligned} U_{ex}(\{q_k\}) &= \varepsilon_{ex} - \sum_k a_k q_k + \frac{1}{2}\sum_k \omega_k^2 q_k^2 \\ &= \varepsilon_{ex} - \frac{1}{2}\sum_k \omega_k^2 \left(q_k - \frac{a_k}{\omega_k^2}\right)^2 - \frac{1}{2}\sum_k \omega_k^2 \left(\frac{a_k}{\omega_k^2}\right)^2. \end{aligned} \tag{10.25}$$

Here, the generalized force for the mode $q_k$ at $q_k = 0$ is $-\partial U_{ex}/\partial q_k = a_k$; then the equilibrium position is shifted to $\bar{q}_k = a_k/\omega_k^2$. After all modes have been shifted to each minimum as $\{q_k : \ 0 \to \bar{q}_k\}$, the total energy is lowered as $E_{LR} = \sum_k a_k^2/\omega_k^2$ (the lattice relaxation energy). It should be noted here that the strength of the induced force $-\partial U_{ex}/\partial q_k = a_k$ and $a_k^2/\omega_k^2$ are infinitesimally small, but the total sum over all modes $E_{LR} = \sum_k a_k^2/\omega_k^2$ has an energy order of eV.

Next, we take an orthogonal transformation of the lattice modes from $\{\omega_k q_k\}$ to $\{Q_l\}$ where one of the modes ($l$=1) is chosen as (Fig. 10.9b).

$$Q_1 = \frac{1}{\bar{Q}_1}\sum_k a_k q_k. \tag{10.26}$$

Then the adiabatic potentials in $\{Q_l\}$ space are

$$U_{\mathrm{g}}(\{Q_l\}) = \varepsilon_{\mathrm{g}} + \frac{1}{2}\sum_l Q_l^2, \tag{10.27}$$

$$U_{\mathrm{ex}}(\{Q_l\}) = \varepsilon_{\mathrm{ex}} - \bar{Q}_1 Q_1 + \frac{1}{2}\sum_l Q_l^2. \tag{10.28}$$

Here, the electron-lattice interaction appears only through the interaction mode $Q_1$. In $\{Q_l\}$ space, the minimum of $U_{\mathrm{ex}}(\{Q_l\})$ is shifted to $\bar{Q}_1$ along the $Q_1$ direction, where the lattice relaxation energy is $\bar{Q}_1{}^2/2$ and the generalized force is $-\partial U_{\mathrm{ex}}/\partial Q_1 = -\bar{Q}_1$. Since the other modes do not affect electronic states, the adiabatic potentials are enough to illustrate only along $Q_1$ mode. In general, the interaction mode is defined so as to connect the minimums of two different electronic states and is not a normal mode in general. For two spherical electronic states (ground and excited), $Q_1$ mode is a kind of breathing mode, for example.

### *10.3.3 Lattice Relaxation Dynamics of Two-Level System*

Assume that before the excitation ($t < 0$), the electron is in the ground state and the lattice does not move at all $\{q_k = 0\}$ ($\{Q_\ell = 0\}$). Suppose at time $t = 0$, the electron is excited by a photon absorption. Since the optical transition time is very short (~fsec), the lattice does not move during the electronic transition (Franck-Condon principle). After the excitation with $h\nu = \varepsilon_{\mathrm{ex}} - \varepsilon_{\mathrm{g}}$, the induced force $F_k^{\mathrm{ex}}$ makes the lattice $q_k$ oscillate from $q_k = 0$ around a new equilibrium $\bar{q}_k$ in $U_{\mathrm{ex}}(\{q_k\})$ space as

$$q_k(t) = \theta(t)\,\bar{q}_k(1 - \cos\omega_k t). \tag{10.29}$$

Here, $\theta(t)$ is a step function. The total motion $\{q_k(t)\}$ is a Lissajous in $\{q_k\}$ space. On the other hand, in $\{Q_\ell\}$ space,

$$Q_1(t) = \frac{1}{\bar{Q}_1}\sum_k \theta(t)\,a_k\bar{q}_k(1 - \cos\omega_k t) \quad \text{and other } Q_l(t) = 0. \tag{10.30}$$

If there is a dispersion in the phonon frequency distribution $\omega_k$, as is likely, whose width is $\Delta\omega$, the phase in each $\cos\omega_k t$ term will be out of phase, resulting a damping oscillation of $Q_1(t)$ around $\bar{Q}_1$ (Figs. 10.10 and 10.11). During the dephasing time $\tau \sim 2\pi/\Delta\omega$, among the initial excess electronic energy $(\varepsilon_{\mathrm{ex}} - \varepsilon_{\mathrm{g}})$, the amount of

$$E_{\mathrm{LR}} = \frac{1}{2}\bar{Q}_1^2 = \frac{1}{2}\sum_k \frac{a_k^2}{\omega_k^2} \tag{10.31}$$

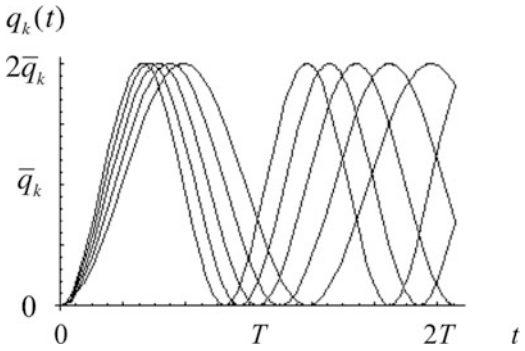



**Fig. 10.10** Dephasing in the normal modes

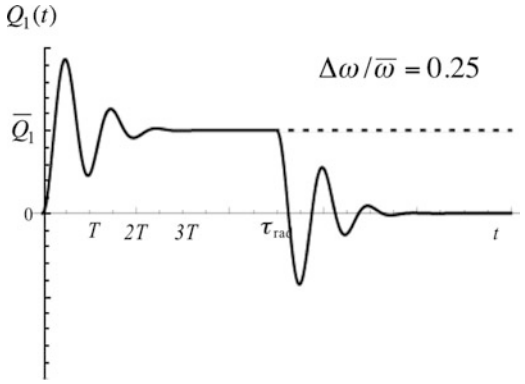



**Fig. 10.11** Transient oscillation in the interaction mode. Optical absorption at $t = 0$, emission at $t = \tau_{\mathrm{rad}}$. $T = 2\pi/\bar{\omega}$

has been transformed into the lattice kinetic energy and dissipated over the crystal. This is the reason that the quantity (10.31) is called the lattice relaxation energy. At the same time, the same amount of the elastic energy $U_{\mathrm{L}}(= E_{\mathrm{LR}})$ has been stored around the defect. It should be noted that in the normal mode space $\{q_k\}$, each $q_k(t)$ simply oscillates, and there is no dissipation at all. The quantity $S = E_{\mathrm{LR}}/\hbar\bar{\omega}_k$, where $\bar{\omega}_k$ is the average frequency, is called the Huang-Rhys factor, denoting the average-induced phonon number generated during the relaxation. For $S \gg 1$, above classical treatment of the lattice can be applied.

If there is a local lattice vibration mode (=normal mode) at the defect and $a_{\mathrm{loc}}$ $\bar{q}_{\mathrm{loc}} \sim 1$, the transient lattice vibration lasts after $t \geq \tau$,

$$Q_1(t) \sim \frac{1}{\bar{Q}_1} a_{\mathrm{loc}} \bar{q}_{\mathrm{loc}} (1 - \cos \omega_{\mathrm{loc}} t). \tag{10.32}$$

Except for the local mode $q_{\mathrm{loc}}$, the lattice relaxation energy $E'_{\mathrm{LR}} = (1/2) \sum_{k \neq \mathrm{loc}} a_k^2 / \omega_k^2$ is dissipated over the crystal during $0 \leq t \leq \tau$, and the rest energy $E_{\mathrm{LR}} - E'_{\mathrm{LR}}$ is gradually dissipated through the unharmonic terms neglected in Eq. 10.25.

The radiative transition time $\tau_{rad}$ is usually much larger than the lattice relaxation time $\tau$. Let us consider a radiative transition occurs from the relaxed $\Psi_{ex}$ to $\Psi_g$ at $t = \tau_{rad}$. According to the Franck-Condon principle as same above, a photon is emitted with an energy $\varepsilon_{ex} - \varepsilon_g - (E_{LR} - U_L)$. After the emission of a photon, the lattice now moves in the $U_g(Q)$ space. The time evolution of the lattice is shown to be a sum of

$$Q_1(t) = \frac{1}{\bar{Q}_1} \sum_k \theta(t)\, a_k \bar{q}_k (1 - \cos \omega_k t) + \frac{1}{\bar{Q}_1} \sum_k \theta(t - \tau_{rad})\, a_k \bar{q}_k \cos \omega_k (t - \tau_{rad}). \tag{10.33}$$

The first term has been generated by an optical absorption at $t = 0$ and hence has damped out for $t > \tau_{rad}$. Then, the second term generated by an optical emission at $t = \tau_{rad}$ is the main term, which shows a damping oscillation around $Q_1 = 0$ (Fig. 10.11). During the damping oscillation in a period $\tau \sim 2\pi/\Delta\omega$, the restored elastic energy $U_L(= E_{LR})$ is dissipated as a kinetic energy of the atomic motion. After all, among the initial excess electronic energy $\varepsilon_{ex} - \varepsilon_g$, the amount $E_{LR}$ is dissipated after the first transition (absorption), and the amount $U_L(= E_{LR})$ is dissipated after the second transition (emission) during each transient lattice vibration.

### *10.3.4 Nonradiative Transition*

If the radiative lifetime $\tau_{rad}$ is short and comparable to the relaxation time $\tau$, hot luminescence may occur during the transient lattice vibration $Q_1(t)$ with a photon energy $h\nu = U_{ex}(Q_1(t)) - U_g(Q_1(t))$ due to the Franck-Condon principle. In the excited state, the lattice $Q_1(t)$ vibrates around $\bar{Q}_1$ and can have chances to cross or reach the crossing point $Q_1^c$ where $U_{ex}(Q_1^c) - U_g(Q_1^c) = 0$, that is, the Franck-Condon energy equals to zero [26]. In this case, the nonradiative transition occurs dynamically without activation energy, and the initial excess electronic energy $\varepsilon_{ex} - \varepsilon_g$ is dissipated in a very short period $\tau \sim 2\pi/\Delta\omega$. The probability depends on the relative positions of $\bar{Q}_1$, $Q_1^c$, and $\tau$ (Fig. 10.12a case 2). On the other hand, after the lattice relaxation in the excited state has completed ($t \gg \tau$), the lattice can be thermally excited around $Q_1 = \bar{Q}_1$. In this case, the nonradiative transition needs an activation energy $(Q_1 - Q_1^c)^2/2$; hence, the probability shows an activation type.

When the lattice relaxation energy is large as $\varepsilon_{ex} - \varepsilon_g - 2E_{LR} < 0$, radiative transitions from the relaxed excited state ($\bar{Q}_1 = \{\bar{q}_k\}$) become prohibited, then the relaxed excited state has a long lifetime (Fig. 10.12b case 3).

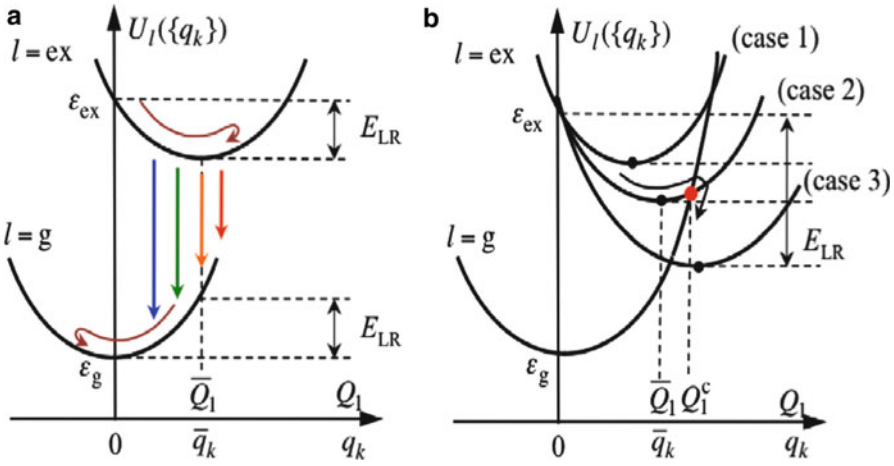



**Fig. 10.12** (**a**) Hot luminescence during the relaxation. (**b**) Nonradiative transition occurs in cases 2 and 3

### *10.3.5 CCD for a Deep-Level Defect*

As have been found experimentally, some deep-level defects show strong electron-lattice interaction. Some of them are considered to behave as a nonradiative recombination center by multiphonon process, where the electronic energy ~$E_{\mathrm{g}}$ is transformed into the lattice vibrational energy [27]. In order to discuss such processes, the configuration coordinate diagram (CCD) is also useful; however, the correct use of the CCD for deep-level defects is rarely seen in the literature. Only when we concentrate on single particular capture process, the adiabatic potential obtained by $U_\alpha(Q) = \varepsilon_\alpha(Q) + CQ^2/2$ may work to some extent. However, if we discuss recombination(s) that consists of series of captures, carrier numbers are not constant. Then the adiabatic potential cannot be calculated as a sum of the electron energy and the hole energy plus the lattice elastic energy: $U_{\alpha,\beta}(Q) = \sum_\alpha \varepsilon_\alpha(Q) - \sum_\beta \varepsilon_\beta(Q) + CQ^2/2$.

The correct meaning of CCD for deep-level defects and how to figure out are precisely given in Refs. [7, 28, 29]. We must recall here the many-electron representation in Sect. 10.2 for a valence hole state if a localized level moves throughout the band gap and couples both with the conduction and the valence bands. Figure 10.13 is schematically presented for the better understanding on the relation among the CCD, the atomic position, and the carrier occupation in the conduction and valence bands [29]. The atomic position illustrated in Fig. 10.13b is an example only a guide for eyes. The numbers indicated in the CCD correspond respectively with the number in the figure for carrier occupation. It should be noticed that at the positions 5, 11, and 17, a trapped electron resonated with the valence band states, but the total charge remains localized as exactly the same as in 4, 10, and 16, respectively. The same is true for a hole localization between 2, 8, and

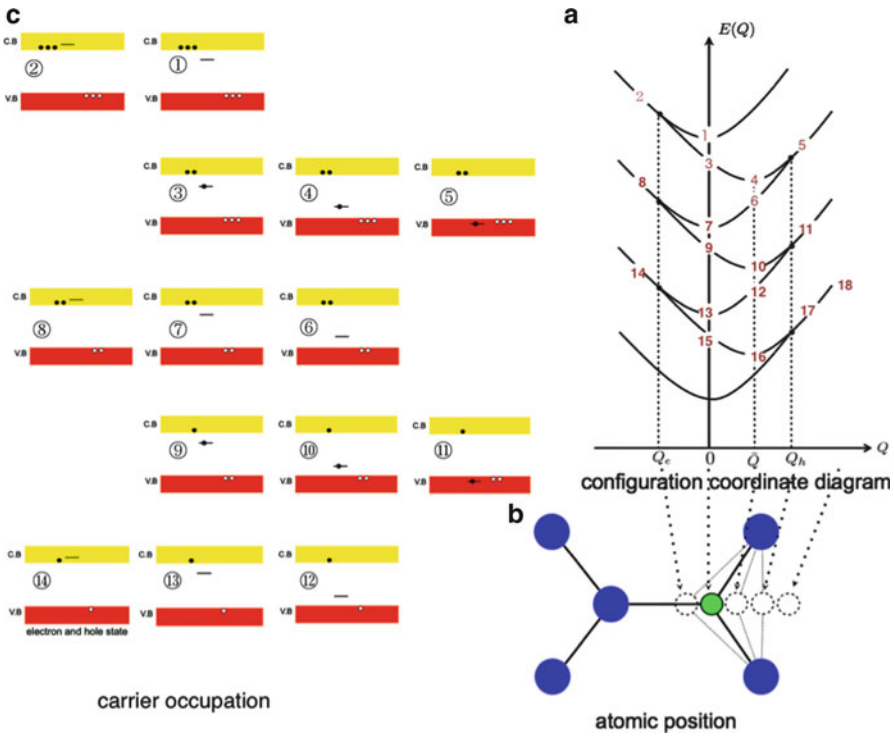



**Fig. 10.13** Schematic relation among (**a**) CCD, (**b**) an example of the atomic position (guide for eyes), and (**c**) the carrier occupation. The numbers indicated in (**a**) CCD correspond, respectively, with the number in (**c**) the carrier occupation [29]

14 and 1, 7, and 13, respectively, if the neutral defect is considered to be a hole localized state.

Figure 10.14 shows a typical example of the CCD for a deep-level defect in a semiconductor with $n_e$ electrons in the conduction band, $n_h$ holes in the valence band. For simplicity, the defect is assumed to have $D^+$ and $D^0$ charged states, for example. Following remarks should be addressed for nonradiative recombination (multiphonon process).

(a) The ordinate $Q_1$ is so called the interaction mode whose direction is chosen between the stable configurations for two electronic states, for example, $D^+$ and $D^0$. The origin $Q_1 = 0$ is taken for $D^0$ for simplicity. $Q_1$ is usually a linear combination of the atomic displacements around the defect and is not a normal mode in general.

(b) In the CCD, each single point represents the total electronic states and the energy of bound electrons and the kinetic energies of free carriers, and the lattice elastic energy stored around the defect as a function of the lattice distortion $Q_1$. The direction of increasing energy is always upward for electrons, holes, and the lattice elastic energy. The adiabatic potential for ($D^+$ + $n_e$ free electron + $n_h$

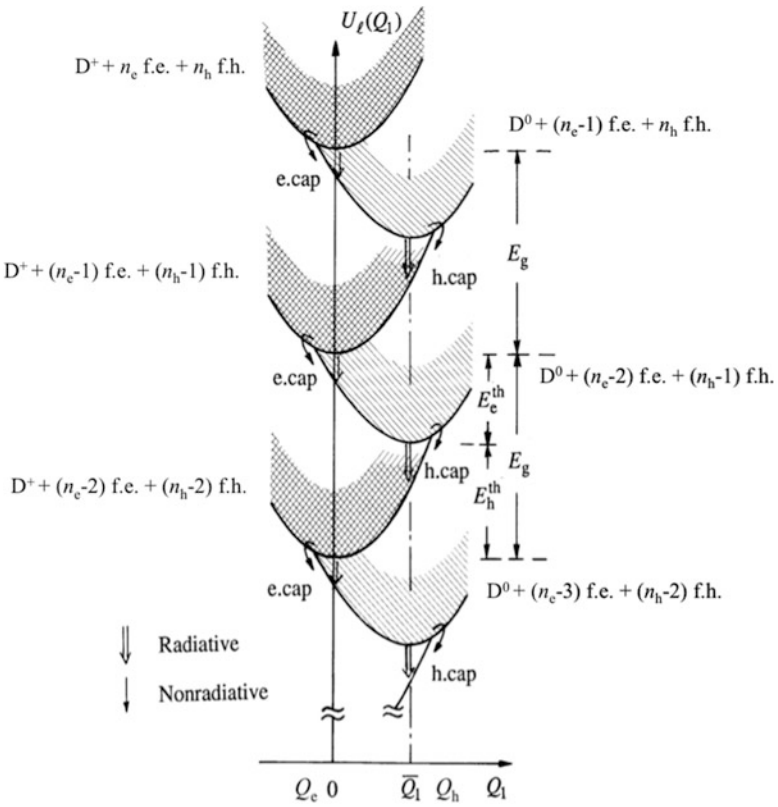



**Fig. 10.14** Configuration coordinate diagram for a deep-level defect with many carriers

free hole) state is exactly the same but shifted $E_g$ higher that for ($D^+ + n_e - 1$ free electron + $n_h - 1$ free hole) state. For example, $n_e = 1$ and $n_h = 1$ state is a double continuum state, owing to the free motions of an electron and a hole.

(c) The relative positions between the adiabatic potentials for ($D^+ + n_e$ free electron + $n_h$ free hole) state and ($D^0 + n_e - 1$ free electron + $n_h$ free hole) state depend on the nature of the defect: $E_e^{th}$, $E_h^{th}$, $Q_e$, $Q_h$, and $\bar{Q}_l$.

(d) At the crossing point $Q_e$ ($Q_h$) of the two adiabatic potentials where the (transient) optical depth of a trapped electron (hole) is zero, a nonradiative capture of an electron (hole) is possible. This electron (hole) capture is a transition from free-electron (hole) continuum states to a single localized state.

(e) During the lattice relaxation after an electron nonradiative capture at $Q_e$, an electronic energy equal to the thermal depth $E_e^{th}$ of a trapped electron is converted to the lattice vibration energy. The average number of created phonons is $E_e^{th}/\hbar\bar{\omega}_k$ (multiphonon process).

(f) At the stable configuration $Q_1 = \bar{Q}_1$ for $D^0$, where the defect is occupied by an electron, the lattice elastic energy stored around the defect is given by $C\bar{Q}_1^{\,2}/2$.
(g) During the lattice relaxation after a hole nonradiative capture at $Q = Q_h$, an electronic energy equal to the thermal depth $E_h^{\,th} = E_g - E_e^{\,th}$ of a trapped hole is converted to the lattice vibration energy. Here, a hole is considered to be trapped by $D^0$ since the transition occurs from free-hole continuum states. Thus, the electronic energy $E_g = E_e^{\,th} + E_h^{\,th}$ is converted to phonons during a pair of e and h captures: a recombination is completed.
(h) Optical electronic transitions take place vertically in the CCD according to the Franck-Condon principle, then the optical depth $E_e^{\,op}$ of a trapped electron at the stable configuration $Q_1 = \bar{Q}_1$ is given by $E_e^{\,th} + C\bar{Q}_1^{\,2}/2$, and the optical depth $E_h^{\,op}$ of a trapped hole at the stable configuration $Q_1 = 0$ is given by $E_h^{\,th} + C\bar{Q}_1^{\,2}/2$.

The dynamics of nonradiative multiphonon recombinations will be discussed in Sect. 10.4.3.

## 10.4 Mechanisms of Defect Reactions

### *10.4.1 Defect Reactions*

So far, we have seen that the stable atomic configuration of defects will change when the electronic state has changed. Then, a carrier capture or an electronic excitation can induce atomic motions around a defect toward a new stable structure. During the lattice relaxation, an amount of the electronic energy is converted to the kinetic energy of atoms, which may induce a defect reaction. In this section, we will discuss the dynamics of defect reactions.

Figure 10.15 schematically shows typical examples of defect reactions [30]: impurity diffusion, structural change of a defect (creation, annihilation, and multiplication), and climb and glide motions of dislocation. It has been shown that defect reactions are induced or enhanced by photoexcitation, electron beam excitation, and carrier injection [31]. Hereafter, the reactions are indicated symbolically as $D_i \rightarrow D_f$ .

Several models have been proposed so far, which can be classified into two categories as schematically shown in Fig. 10.16. One is the structural instability mechanism [32, 33] (Fig. 10.16a) where a particular electronic state of a defect induces a symmetry-breaking atomic motion such as the off-center distortion. For example, if the relaxation mode $Q_1$ is nothing but the reaction coordinate $Q_R$ ($D_i \rightarrow D_f$) itself, an electronic excitation to $ex_1$ promptly induces the reaction or generates a precursor of $D_f$. An excitation to $ex_2$ decreases the potential barrier height ($E_t \rightarrow E_t^{\,*}$) of the reaction. The activation energy is decreased for the excited state whose lifetime $\tau$ is much larger than the inverse of the reaction rate $1/P_{reac}$.

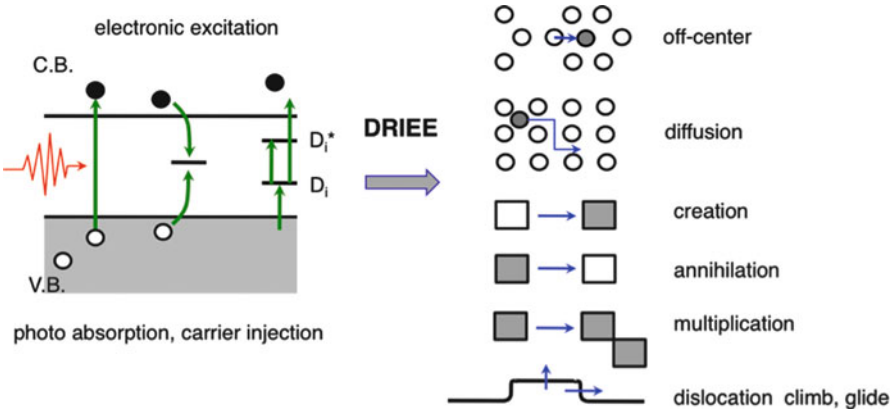



**Fig. 10.15** Defect reactions induced by electronic excitations

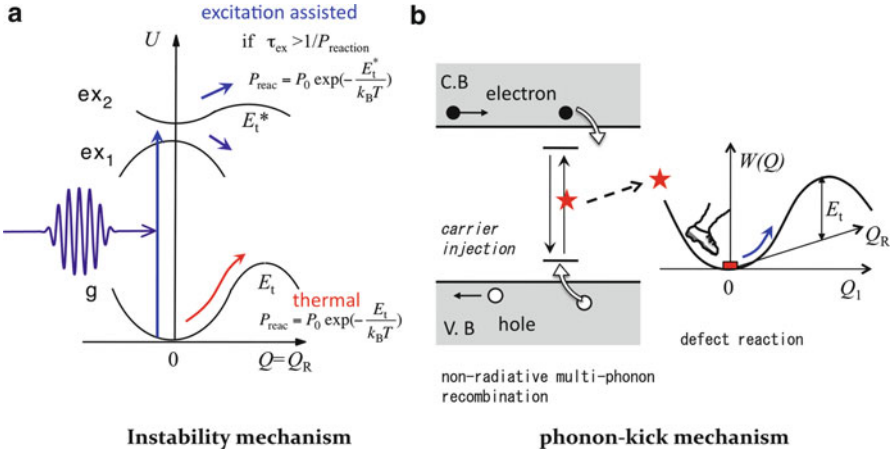



**Fig. 10.16** Two categories for mechanisms of defect reactions

The velocity of b-dislocation in p-GaAs is enhanced under illumination, and then the mechanism is considered to be this type [31].

The other mechanism is the recombination-enhanced (or phonon-kick) mechanism (Fig. 10.16b), where recombination(s) of carriers at a defect induces a transient atomic motion in the reaction coordinate $Q_R$ [34, 35]. The first mechanism (a) stresses why and how the instability takes place toward a "defect" configuration. The second mechanism (b) stresses how the electronic energy converts to the lattice vibration energy and the motion in the reaction coordinate to overcome the potential barrier. Though two mechanisms look different, they can be treated in a unified scheme as follows.

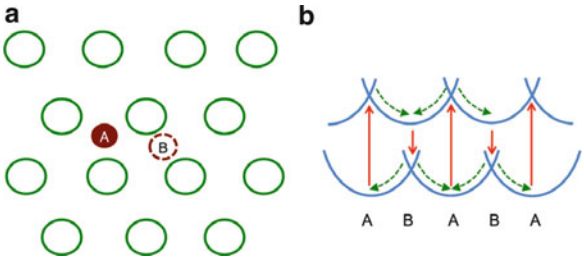



**Fig. 10.17** (**a**) Interstitial positions in a lattice. (**b**) Electronically enhanced impurity diffusion

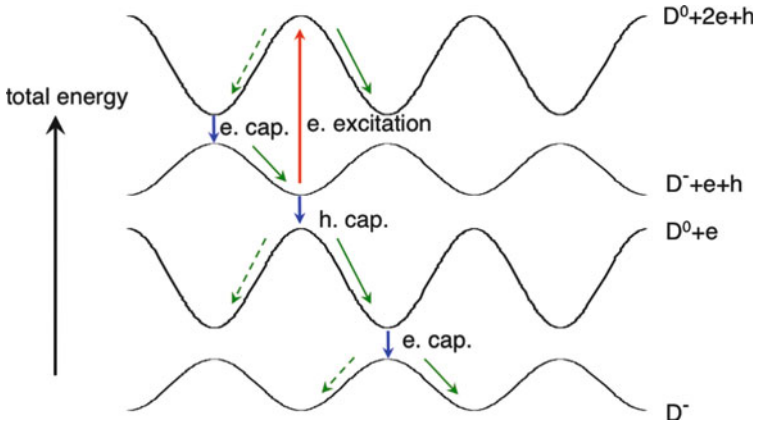



**Fig. 10.18** The Bourgoin-Corbett mechanism for impurity diffusion

### *10.4.2 Instability Mechanism for Defect Reaction*

Let us briefly discuss an impurity diffusion by the interstitial mechanism, where an impurity atom diffuses via interstitial positions in the lattice. An interstitial impurity may occupy a larger vacant site in the lattice such as **A** and **B** in Fig. 10.17. It is probable that these sites have high point symmetry, and then we can apply the results in Sect. 10.3. If a localized electronic state which shows the off-center J-T instability at a position **A** which continuously changes into a state which does not show the instability at another position **B** and vice versa, the diffusion is enhanced by the electronic excitation and deexcitation. The self-interstitial in Si is found to be this type [36]. This mechanism is similar to the Bourgoin-Corbett (B-C) mechanism [32], but the number of bound electrons (holes) here remains constant. The B-C mechanism has been proposed for an enhancement of an impurity diffusion, where the stable position of an impurity is unstable for another charged state and vice versa (Fig. 10.18).

### 10.4.3 Annihilation of an Impurity Complex

A hydrogen atom can passivate or activate donors, acceptors, and other impurities in semiconductors. In silicon, a hydrogen impurity can occupy several stable positions in the diamond lattice. If it occupies a bond center position between a substitutional carbon (isoelectronic) impurity and a silicon atom, this H-C complex acts as a shallow donor whose thermal depth is $E_{\mathrm{th}}{}^{\mathrm{e}}$ =0.15 eV. The positively charged H-C complex is stable above the room temperature, while the neutral state is unstable at the room temperature, and the activation energy for the annihilation is estimated as 0.5–0.7 eV [37]. Since the "united atom" of an H-C complex is a nitrogen atom in the molecular chemistry, the effective potential seen by a conduction electron away from the $(\mathrm{H\text{-}C})^{+}$ complex is similar to that of the $N^{+}$ impurity. Then, we can derive an analogy between the stability of an N impurity and that of an H-C complex in Si. For a shallow donor state or a positively charged state (an electron is away), the central position is stable for an N impurity, and the H-C complex is stable. When an electron comes close to them, on the other hand, it induces the off-center displacement of an N impurity and the off-center displacement of a hydrogen atom. That is, a hydrogen atom is pushed away from a carbon: the annihilation of an H-C complex. This would be a typical example of a defect reaction caused by the structural instability mechanism.

### 10.4.4 Phonon-Kick Mechanism

Let us next discuss the transient lattice vibration induced by carrier captures with CCD in Fig. 10.14. A nonradiative electron (hole) capture may occur when $Q_1(t)$ crosses at $Q_{\mathrm{e}}$ ($Q_{\mathrm{h}}$) in the c-c diagram with a capture rate $p_{\mathrm{e}}$ ($p_{\mathrm{h}}$) per crossing. The first capture process needs an activation from $Q = 0$ to $Q = Q_{\mathrm{e}}$. This first nonradiative capture probability $p_{\mathrm{e}}{}^{\mathrm{thermal}}$ has been investigated theoretically by Henry and Lang [27], Sumi [35], and Kayanuma [38, 39]. It is found that $p_i^{\mathrm{thermal}} = n_i\, v_i^{\mathrm{T}}\, \sigma_i$, where $n_i$ is the carrier density of $i = (\mathrm{e}, \mathrm{h})$, $v_i{}^{\mathrm{T}}$ is the thermal velocity, and $\sigma_i = \sigma_i^{\infty} \exp[-E_i^{\mathrm{act}}/k_{\mathrm{B}}T]$ is the capture cross section. The capture rate $p_i$, which will be used in this chapter, is given by $p_i = n_i\, v_i^{\mathrm{T}}\, \sigma_i^{\infty}$.

Consider a situation where $N$ electrons and $N$ holes are sequentially captured by a defect at $Q = Q_{\mathrm{e}}$ or $Q_{\mathrm{h}}$, each occurred at $t_{\mathrm{e}1} < t_{\mathrm{h}1} < t_{\mathrm{e}2} < t_{\mathrm{h}2} < \cdots < t_{\mathrm{e}N} < t_{\mathrm{h}N}$; the time evolution of $Q_1(t)$ for $t_{\mathrm{h}N} < t$ is shown to be given by [29, 30, 40]

$$Q_1(t) = Q_1^{\mathrm{before}}(t) + \bar{Q}_1 \left\{ 1 - \sum_j^N \theta\left(t - t_{\mathrm{e}j}\right) \exp\left[\frac{-\Delta\omega^2 \left(t - t_{\mathrm{e}j}\right)^2}{4}\right] \cos\omega_0\left(t - t_{\mathrm{e}j}\right) + \sum_j^N \theta\left(t - t_{\mathrm{h}j}\right) \exp\left[\frac{-\Delta\omega^2 \left(t - t_{\mathrm{h}j}\right)^2}{4}\right] \cos\omega_0\left(t - t_{\mathrm{h}j}\right) \right\}, \tag{10.34}$$

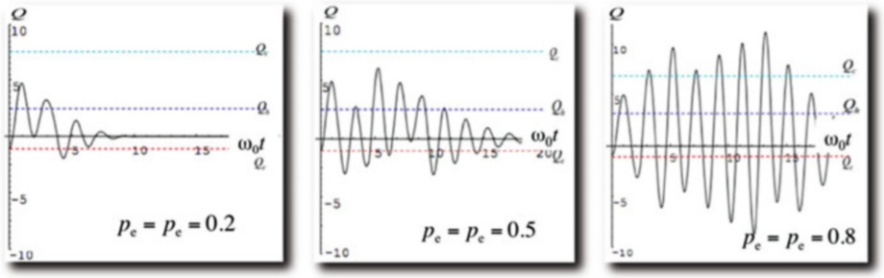



**Fig. 10.19** Examples of transient lattice vibration induced by carrier captures for $\Delta\omega/\omega_0 = 0.1$ [29]

where the lattice motion is treated classically and obeys the Newton equation in the adiabatic potential for each electronic state. Here, the interaction mode $Q_1$ is assumed to be composed of many normal modes whose phonon frequency distribution is given by a Gaussian with a central frequency $\omega_0$ and a width $\Delta\omega$. In Eq. 10.34, $\theta(t)$ is a step function and $Q_1^{\text{before}}(t)$ is the time dependence if all the carrier captures would not have occurred at all: $Q_1^{\text{before}}(t)$ stands for thermal fluctuation. When an electron (hole) capture occurs at $t_{ej}$ ($t_{hj}$), the amplitude of $Q_1(t)$ increases stepwise, which is followed by a damping oscillation around $Q_1 = \bar{Q}_1$ ($Q_1 = 0$) with a decay time $\tau \sim 2\pi/\Delta\omega$ due to dephasing. If $Q_1(t)$ can reach $Q_h$ ($Q_e$) during the damping oscillation, the next nonradiative hole capture is possible with a rate $p_h$ ( $p_e$) per crossing (Fig. 10.19b). If $Q_1(t)$ cannot reach $Q_h$ ($Q_e$), next nonradiative electron (hole) capture will not take place, and then $Q_1(t)$ will be damped out to $Q_1(t)=\bar{Q}_1$ ($Q_1(t) = 0$) (Fig. 10.19a). Thus, the capture times $t_{ej}$ and $t_{hj}$ in $Q_1(t)$ (Eq. 10.35) depend on the time evolution of $Q_1(t)$ itself and the capture probability process. If the amplitude of $Q_1(t)$ increases remarkably by successive carrier captures and $Q_1(t)$ can reach another critical point $Q_c$, then a defect reaction may occur (Fig. 10.19c).

We have performed simulations on the transient $Q_1(t)$ for various values of $p_e$, $p_h$, and $\Delta\omega/\omega_0$ and the relative positions of $Q_e$, $Q_h$, $\bar{Q}_1$, and $Q_c$ using the following assumptions [29]:

(a) The first electron capture takes place by a thermal activation to $Q_1 = Q_e$ with the activation energy $E_e^{\text{act}}$.
(b) If $Q_1(t)$ crosses at the point $Q_e$ ($Q_h$) of the adiabatic potentials, there is a rate $p_e$ ( $p_h$) per crossing to capture an electron (hole) nonradiatively. In the adiabatic limit, $p_i \sim n_i/n_i^T$, where $n_i^T$ is the effective carrier density ($i = e, h$) given by $2(m_i^* k_B T/2\pi\hbar^2)^{\frac{3}{2}}$.
(c) If $Q_1(t)$ reaches the critical point $Q_c$, the defect reaction takes place.

Figure 10.20 shows the obtained criteria whether a defect reaction occurs or not. The condition for the coherent captures turns out to be $(\omega_o/2\pi)p_i\tau \sim \omega_0/\Delta\omega > 1$. Then, if $N$ pairs of electron and hole are captured within a short period $\sim 2\pi/\Delta\omega$, the amplitude of the interaction mode $Q_1(t)$ increases remarkably. By a series of coherent carrier captures, more than the band gap energy $E_g = E_e^{\text{th}} + E_h^{\text{th}}$ can be transformed into the lattice vibration energy. The defect reaction rate is given by

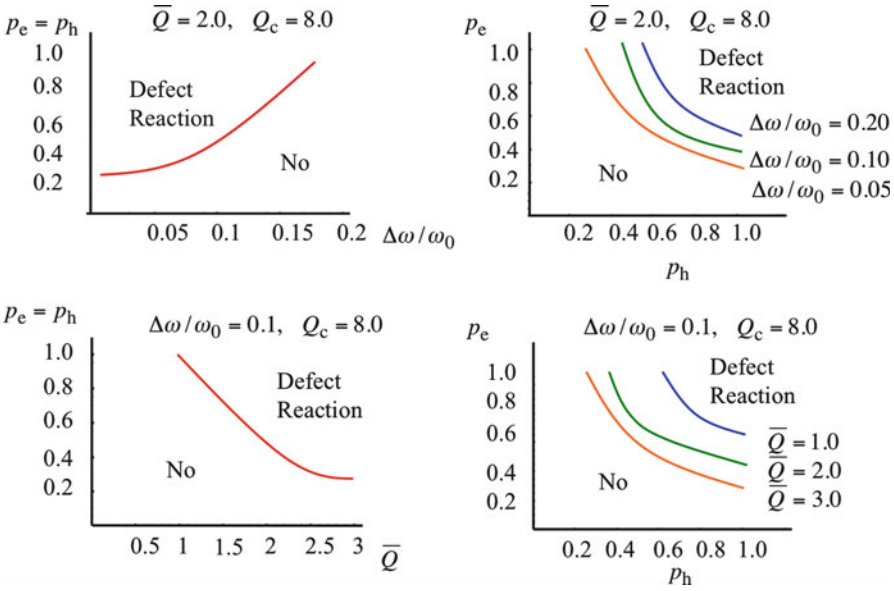



**Fig. 10.20** The obtained criteria whether a defect reaction occurs or not [29]

$(\omega_0/2\pi)\exp(-E_i^{\text{act}}/k_BT)$ because only the first capture ($i$ = e, h) is to be activated. Various situations are possible depending on the variety of charged states of a defect, the magnitude of $E_{\text{th}}^{\text{e}}$ and $E_{\text{th}}^{\text{h}}$, the relation between the relaxation mode $Q_1$ and the reaction coordinate $Q_R$, the magnitude of capture cross sections, and so on.

### *10.4.5 Dynamics of Defect Reactions*

If the relaxation mode $Q_1$ is a symmetric mode such as a breathing mode, $Q_1$ is then orthogonal to $Q_R$ that is a symmetry-breaking mode by definition. Then, the symmetrical relaxation process does not affect defect reactions. When $Q_1$ is not orthogonal to $Q_R$, Sumi [35] has shown theoretically that the transient lattice vibration by a capture reduces the activation energy $E_t$ for a defect reaction to $E_t - \left(E_{\text{act}}^i + E_{\text{th}}^i\right)$ or $\left(\sqrt{E_t} - |\cos\theta|\sqrt{E_{\text{act}}^i + E_{\text{th}}^i}\right)^2 \Big/ (1 - \cos^2\theta)$ where $\theta$ is the angle between $Q_1$ and $Q_R$.

When the off-center instability takes place by the charge change of a defect, it is important which corresponds to the off-center configuration, $Q_1 = 0$ or $\bar{Q}_1$? If the reaction coordinate $Q_R$ is almost parallel to $Q_1$, a defect reaction in p-type semiconductors occurs by the structural instability mechanism induced by a minority electron capture (Fig. 10.21). If $E_{\text{th}}^{\text{e}}$ is large and $Q_1 // Q_R$ but there is a potential barrier, the phonon-kick mechanism takes place with the aid of $E_{\text{th}}^{\text{e}}$ induced by an

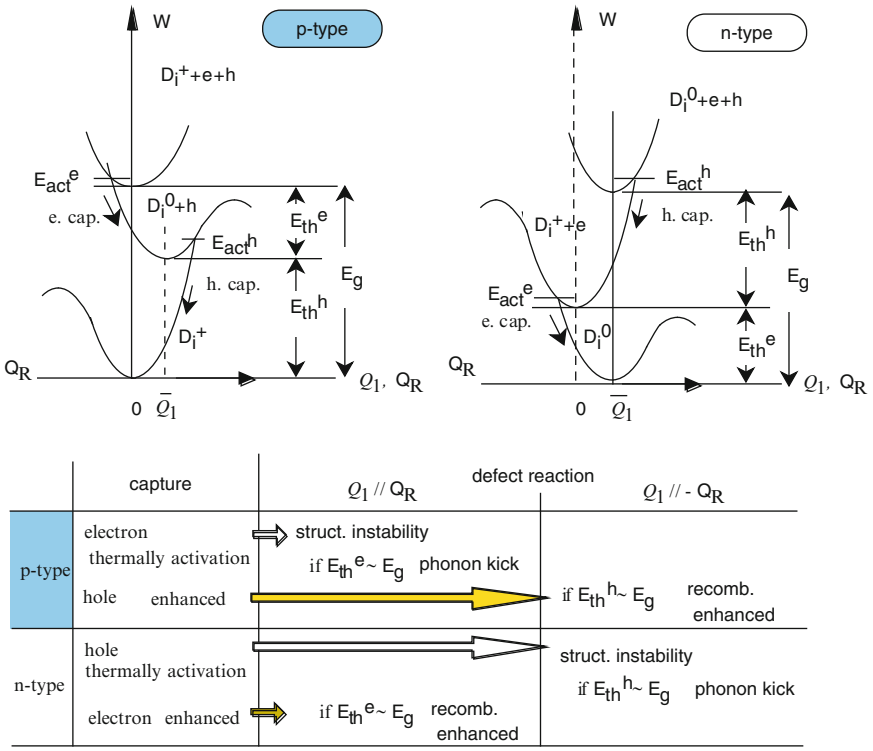



**Fig. 10.21** An example of the configuration coordinate diagrams for the same defect in p-type and n-type semiconductors. Thick adiabatic potentials are for the interaction mode $Q_1$ (relaxation mode) and thin adiabatic potentials for the reaction coordinate $Q_R$. The table summarizes the most probable path for DRIEE

electron capture. If $E_{th}^h$ is large and $Q_1 // -Q_R$, the phonon-kick mechanism takes place by a hole capture which has been enhanced by an electron capture. If a capture of a majority carrier occurs very quickly within the local transient vibration, the band gap energy $E_g = E_{th}^e + E_{th}^h$ can be used for an excitation of a defect reaction. This could be the most effective path to the defect reaction and nothing but the so-called recombination-enhanced mechanism [41,42]. In n-type semiconductors, defect reactions can be discussed in a similar way as above, starting from $Q_1 = \bar{Q}_1$ and $D_i^0 + e + h$. In contrast to p-type, the first capture of a minority hole by a neutral center $D_i^0$ is not accelerated by the Coulomb attraction, while the second electron capture by $D_i^+$ is accelerated. We have assumed that the defect has $D_i^0$ and $D_i^+$ states. If it has $D_i^-$ and $D_i^0$, the above discussion can be also valid with respective translations with $D_i^0 \rightarrow D_i^-$ and $D_i^+ \rightarrow D_i^0$. As we have shown, either in n-type or p-type, the capture of a majority carrier is always enhanced by a minority carrier capture and vice versa. But only a hole capture is accelerated by the Coulomb attraction of $D_i^-$.

## 10.5 Extrinsic Origin for Defect Reaction

As has been mentioned in the introduction, defect reactions in semiconductors are induced or enhanced also by extrinsic origins, irradiations from outside. Irradiation with photon, electron, ion, etc., usually excites electrons in target materials. When the excitation energy is larger than the fundamental energy gap of the material, a pair of an electron and a hole is created. The created carriers interact with the lattice and then sometimes cause defect reactions, as have been discussed in the previous sections.

### *10.5.1 Core Excitation*

When the excitation energy is so high to excite the core electrons of the constituent atoms of the material, a core hole is created. After a very short lifetime ($<10^{-15}$ s), the core hole is annihilated by the Auger process. There are two types of Auger recombination. One is the participant Auger where the excited electron from the core state excites a valence electron and again recombines with the core hole, leaving one valence hole. The other is the spectator Auger where another valence electron recombines with the core hole, leaving two valence holes and one conduction electron. The spectator Auger process is an effective trigger for defect reactions because the resulting two holes are very effective to interact with the lattice (bipolaron mechanism) in addition to the leaving conduction electron which helps to weaken the bond around it.

### *10.5.2 Knock-On Process*

Irradiation by a particle with nonzero mass causes a momentum transfer to constituent atoms. For electron irradiation whose kinetic energy is larger than ~10 keV, it enables to knock on an atom in semiconductors from the stable position to the interstitial position: a Frenkel pair defect (a vacancy and an interstitial) is created.

## 10.6 Summary

In this chapter, we have discussed mechanisms of defect reactions (defect creation, annihilation, multiplication, reconstruction, impurity diffusion, etc.) in semiconductors starting with the origin of the electron-lattice interactions. Strong localization of carriers accompanied with large lattice distortion is explained in terms of the extrinsic self-trapping. Symmetry-breaking structural instability is then

discussed in connection with the Jahn-Teller effect. It is found that the tetrahedral coordination is not so rigid as it is considered, especially for hole localization. Next, we have explained the true meaning of the so-called CCD, which is sometimes misunderstood in literature. The proper relation is explained among the lattice distortion, the level position of a defect in the band gap, and the thermal and optical depth. The multiphonon carrier capture and the following induced lattice relaxation processes are explained. Finally, we showed a simple simulation on the phonon-kick mechanism. Energy released by a carrier capture can enhance the next capture process, and then for high carrier density, a positive feedback may occur to create a rapid increase of lattice vibrations, which then causes defect reactions.

**Acknowledgments** The author would like to thank Professor K. Maeda of University of Tokyo and Professor O. Ueda of Kanazawa Institute of Technology for their valuable discussion.

## References

1. Y. Shinozuka, Relaxations of excited states and photo-induced structural phase transitions, in *Proceedings of 19th Taniguchi Symposium*, ed. by K. Nasu (Springer, Berlin, 1997), p. 229
2. J. Dabrowiski, M. Scheffler, Phys. Rev. Lett. **60**, 2183 (1988)
3. D.J. Chadi, K.J. Chang, Phys. Rev. Lett. **60**, 2187 (1988)
4. O. Ueda, *Reliability and Degradation of III-V Optical Devices* (Artech House Publishers, Boston-London, 1996)
5. O. Ueda, Jpn. J. Appl. Phys. **49**, 090001(2010)
6. K. Maeda, S. Takeuchi, in *Dislocations in Solids*, ed. by F.R.N. Nabarro, M.S. Duesbery, vol. 10 (North-Holland, Amsterdam, 1996), pp. 444–504
7. Y. Shinozuka, Jpn. J. Appl. Phys. **32**, 4560 (1993)
8. Y. Toyozawa, Chapter 4, in *Excitonic Processes in Condensed Matter*, ed. by M. Ueta, H. Kanzaki, K. Kobayashi, Y. Toyozawa, E. Hanamura (Springer, Berlin, 1986)
9. Y. Toyozawa, *Optical Processes in Solids* (Cambridge University Press, Cambridge, 2003)
10. Y. Toyozawa, Physica **116B**, 7 (1983)
11. Y. Toyozawa, Solid State Electron. **21**, 1313 (1978)
12. Y. Shinozuka, Y. Toyozawa, J. Phys. Soc. Jpn. **46**, 505 (1979)
13. K.L. Brower, Phys. Rev. **B26**, 6040 (1982)
14. K. Murakami, H. Kuribayashi, K. Masuda, Phys. Rev. **B38**, 1589 (1988)
15. D.J. Chadi, K.J. Chang, Phys. Rev. Lett. **61**, 873 (1988), Phys. Rev. B **39**, 10063 (1989)
16. H.A. Jahn, E. Teller, Proc. Roy. Soc. **A161**, 220 (1937)
17. Y. Onodera, *Group Theory and Application to Quantum Mechanics* (Springer, Berlin, 1995)
18. G.D. Watkins, Proc. Defect. Semicond. 228 (1972); Physica **117B & 118B**, 9 (1983)
19. M. Sprenger, S. Muller, E. Sieverts, C. Ammerlaan, Phys. Rev. **B35**, 1566 (1987)
20. Y. Shinozuka, Mater. Sci. Forum **83–87**, 527 (1992)
21. M. Saito, A. Oshiyama, O. Sugino, Phys. Rev. **B47**, 13205 (1993)
22. Y. Toyozawa, J. Phys. Soc. Jpn. **51**, 1861 (1981)
23. M. El-Maghraby, Y. Shinozuka, Proc Mater. Sci. Forum **258–263**, 647 (1997)
24. M. El-Maghraby, Y. Shinozuka, J. Phys. Soc. Jpn **67**, 3524 (1998)
25. I. Takahashi, Y. Shinozuka, Physica B **340–342**, 349 (2003)
26. A.M. Stoneham, *Theory of Defects in Solids* (Clarendon, Oxford, 1975)
27. C.H. Henry, D.V. Lang, Phys. Rev. **B15**, 989 (1977)
28. Y. Shinozuka, J. Phys. Soc. Jpn. **51**, 2852 (1982)

29. Y. Shinozuka, Mater. Res. Soc. Symp. Proc.. **1195**, B02–02 (2009)
30. Y. Shinozuka, in *Defects in Electronic Materials II*, ed. by J. Michel, T. Kennedy, K. Wada, K. Thonke. Proceedings of MRS 1996 Fall Meeting 442 (Materials Research Society, Pittsburg, 1996), p. 225
31. K. Maeda, M. Sato, A. Kubo, S. Takeuchi, J. Appl. Phys. **54**, 161 (1983)
32. J.C. Bourgoin, J.M. Corbett, Phys. Lett. **38A**, 135 (1972), and Radiat. Eff. **36**, 157 (1978)
33. M.K. Sheinkman, JETP Lett. **38**, 330 (1983), and M.K. Sheinkman, L.C. Kimerling, in *Defect Control in Semiconductors*, ed. by K. Sumino (North-Holland, Amsterdam, 1990), p. 97
34. J.D. Weeks, J.C. Tully, L.C. Kimerling, Phys. Rev. **B12**, 3286 (1975)
35. H. Sumi, Phys. Rev. **B29**, 4616 (1985), J. Phys. **C17**, 6071 (1984)
36. R. Car, P.J. Kelly, A. Oshiyama, S.T. Pantelides, Phys. Rev. Lett. **54**, 360 (1985)
37. Y. Kamiura, M. Tsutsune, M. Hayashi, Y. Yamashita, F. Hashimoto, Mater. Sci. Forum **196–201**, 903 (1995)
38. Y. Kayanuma, J. Phys. Soc. Jpn. **51**(11), 3526 (1982)
39. Y. Kayanuma, S. Fukuchi, J. Phys. Soc. Jpn. **53**(5), 1869 (1984)
40. Y. Shinozuka, T. Karatsu, Mater. Sci. Forum **258–263**, 659 (1998)
41. K. Maeda, Mater. Res. Soc. Symp. Proc. **1195**, B02–01 (2009)
42. Y. Shinozuka, Chapter 8, in *Defects in Optoelectronic Materials*, ed. by K. Wada, S.W. Pang (Gordon and Breach Science Publishers, London, 2001)

# Part II
# Materials Issues and Reliability of Electron Devices

# Chapter 11
# Reliability Studies in the Real World

**William J. Roesch**

**Abstract** This chapter is intended to give a nontraditional look at reliability of compound semiconductors. It is a review of the five main stages of reliability analysis. Although essential, the first three stages are separated as those traditional tests which are particularly applicable for process development and material selection phases. The remaining two reliability stages are the focus of real life. These later stages are the tools used for improving reliability of established processes and families of products where significant changes and switches to new materials are not feasible. While all five steps are necessary in reliability studies, the methods of establishing and improving reliability are different.

## 11.1 Vocabulary and Introduction

The definition of reliability is an interesting enigma. So often, reliability gets defined by collections of other words ending with "ability," such as dependability, availability, or maintainability. Rarely do definitions focus on the "rely" aspect of the reliability word. Rely sets a demanding expectation. For example, in human reliability, baseball legend Cal Ripken Jr. was nicknamed the "iron man." In considering the *rely* aspect for an iron man, most people almost immediately think of the measurement units. In Cal's case: How many years did he play in every game? (17 seasons) Or, how many consecutive games did he play? (2,632).

Common reliability definitions don't concern mortal humans but instead focus entirely on equipment, machinery, systems, or other apparatus. Since the first transistor was invented in the mid-twentieth century, reliability has been an expectation for electronics. Semiconductor discussions about reliability quickly jump to statistics and probability. Most often, the "rely" is traded for quality. In fact, the

W.J. Roesch (✉)
TriQuint Semiconductor, Inc., 2300 N.E. Brookwood Parkway, Hillsboro, OR 97124-5300, USA
e-mail: Bill.Roesch@tqs.com

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_11,
© Springer Science+Business Media New York 2013

**Table 11.1** Perspectives and vocabulary

| | Types of engineers | | |
|---|---|---|---|
| Customer | Process and product | Quality | Reliability |
| Price | Performance parametric | Yield loss | Fallout |
| Performance | | | |
| Delivery | | | |
| Absence of failures throughout the warranty period | Yield % of "good" things | # of "bad" things | Rate of "bad" things (FITs) |
| | | Parts per million defects | Defects per million |
| | | SPC ($6\sigma$) | Probability |
| | | | Confidence levels |
| | | | Acceleration factors |
| | | | Distributions |

simplest definition of reliability is quality-ability, meaning some measure of performance (quality) for a period of time. Time is a key essential in defining reliability, which in effect, does extend the quality requirement over a finite demand period. In addition to time, folks who seek to measure or predict reliability will utilize various types of stress to emulate worst case use or to achieve acceleration beyond a use expectation (Table 11.1).

In the end, reliability engineers study degradation and failures. Reliability folks typically don't care about rates of success; they are obsessed with failure rates. Unfortunately, the vocabulary between reliability engineers and their coworkers get unnecessarily complicated. Predictions of failures are made using confidence levels and the statistical measures of probability. Customers are taught to focus on abstract terms such as mean time between failures (MTBF) which is a horrible approximation of lifetimes for semiconductor reliability. But in the real world, the actions to eliminate failures are the key to improvement. The reliability engineering job is only just beginning when the equations are determined, the calculations are made, and the confidence is claimed. True reliability is not probability of success; it is the *absence of failure*. Reliability engineering in the real world is all about reducing customer fallout rates until failures are eliminated.

Because of their obsession with time, reliability engineers are always seeking ways to manipulate normal expected lifetimes. Both traditional and nontraditional techniques use methods of acceleration to achieve three goals:

1. To predict future failures
2. To improve reliability
3. To ensure reliability

Acceleration is achieved by environmental stressing and electrical stressing. But in order to make improvements, the reliability engineer needs help. Changes come

about through collaborative efforts between reliability engineers and designers, quality engineers, process engineers, and product engineers.

Unlike quality and reliability folks, the process and product engineers are eternally optimistic, so they typically focus on yield. They like to focus on measuring and improving the percent of good things – so yield is typically the measurement that gets the most attention. Quality engineers also utilize yield, but they measure distributions, statistics, and capabilities as well as shifting some attention from the percentage of "good" to the number of "bad," by looking at defects or defect populations. Reliability engineers take the quality engineer perspective a little bit further. While still focusing on broken populations, the reliability engineer adds the perspective of time (rates), and they will look for innovative, yet meaningful, ways to warp time with the acceleration of stress.

Reliability investigates both wearout and defects, looking for "special lots," mavericks, and outliers. Corrective actions are initiated by failures, regardless of type. In the absence of degradation and destruction, reliability engineers move on to different aspects of stress and degradation.

Is reliability the opposite of insanity? Before you answer, consider how many times failing reliability tests are repeated by reliability optimists. Even though good scientific practice demands multiple trials to verify and validate, human nature rarely evokes a desire to repeat a passing qualification.

*Reliability*: the extent to which the same experiment, test, or measuring procedure yields the same results on repeated trials.

*Insanity*: the expectation that the same experiment, test, or measuring procedure yields different results on repeated trials (i.e., doing the same thing and expecting different results).

*Optimism*: the expectation that the same (failing) experiment, test, or measuring procedure yields different (passing) results on repeated trials (of a different sample).

Although silicon integrated circuits have dominated the semiconductor industry and the reliability focus, compound semiconductors have been around since the invention of solid state technology. However, unlike silicon, compound semiconductor circuits experienced a longer development ramp and languished as a curiosity for many years. Ignited by LEDs, the optical devices pulled compounds into the forefront. As a result, practical integrated circuits from compound semiconductor materials didn't enter the "real world" until the 1980s. For example, GaAs devices generated their own niche at about the time the commercialization and "catch up" phase started in earnest in the 1980s. Although it has a certain allure, new technology unfortunately carries a suspicion of reliability risk.

Semiconductors are constantly evolving and changing. The biggest drive for change has been the endless pursuit to compress circuit features in order to reduce costs and increase functionality of devices, which is commonly referred to as Moore's Law. The desire to cram more components onto integrated circuits, with higher reliability and lower cost, constantly drives improvement in semiconductor processing and fabrication methods [1].

**Table 11.2** Historical eras of reliability

| Era/ stage | Definition | Silicon example | Compound semiconductor example |
|---|---|---|---|
| 1 | Materials | 1975 | 1980 |
| | | Introduction of new VLSI materials: Si, Al, $SiO_2$ | Introduction of GaAs FET materials: GaAs, Au, $Si_xN_x$ |
| 2 | Mechanisms | 1980 | 1985 |
| | (Major reliability problems) | Mobile ions, Electromigration. Stress migration, TDDB, corrosion, cracked die, Broken-bonds, ESD, soft errors | Sinking gates, ohmic contacts |
| 3 | Physics | 1985 | 1990 |
| | (Major reliability physics effort) | Models developed for EM, mobile ions, SM, TDDB, corrosion, temp-cycling, alpha particles, EOS/ESD | Thermal diffusion, JEP118, hydrogen |
| 4 | Engineering | 1990 | 1995 |
| | (Major reliability engineering effort) | Building-in reliability, with emphasis on WLR and DIR | WLR, passives, BIR |
| 5 | Defects | 1995 | 2000 |
| | (Major defect-reduction effort) | SPC and 6 sigma outliers | Capacitors, interconnects |

Reliability counts.
In almost every case integrated electronics has demonstrated high reliability.

Gordon Moore [1]

In studying the history of semiconductor reliability, there have been five activities in reliability that are copied over and over with each new development. Semiconductor integrated circuit reliability typically goes through these five stages of reliability. These stages, or eras, have historically proceeded chronologically, with each era taking up to 5 years to complete. A reliability era is intended to show the maturation of a technology. To further describe reliability eras, consider stages of reliability for silicon from 1975 through 2000 as shown in Table 11.2 [2].

Prior to 1975, reliability improvement on silicon devices had already proceeded through emerging generations leading up to VLSI. In the first era of silicon reliability improvement, from 1975 to 1980, the *introduction of "new" VLSI materials* was at the forefront. Everyone was learning about material properties of Si, Al, and $SiO_2$ and their various interactions. The study and selection of sets of materials for semiconductor construction is focused upon integration of circuits into a manufacturable configuration – by shrinking feature sizes. However, the material selection is also the formative stage of reliability. After the original generations of small-scale, medium-scale, and large-scale integration, the silicon material set was generally established for VLSI.

In the second reliability era, from 1980 to 1985, VLSI silicon technology was faced with *major reliability problems*. For silicon technologies, some of these major problems were mobile ions, aluminum electromigration, stress migration, time-dependent dielectric breakdown (TDDB), cracked die, broken bond wires, purple plague, and soft errors. Problems can be considered as individual failure mechanisms.

In the third era, from 1985 to 1990, *reliability physics* was the focus for the silicon folks. The reliability professionals had developed degradation models for all the previously identified mechanisms, and they also derived acceleration factors for the special environmental stresses of temperature cycling and corrosion. These were the showcase years for electromigration theories, gate oxide research, and ESD studies on silicon-based devices.

In the fourth era, from 1990 to 1995, the silicon folks adopted a new phrase, *building-in reliability*. This was a period of reliability engineering, with a lot of emphasis on wafer-level reliability. The focus was on control of design, processes, and materials as inputs of reliability versus evaluating the end product. Built-in reliability was a phrase used in the silicon industry to emphasize process control, in-line screening, and wafer-level reliability (WLR) testing.

In the fifth era, from 1995 to 2000, the silicon folks were completing their 20-year cycle by merging the metrics of reliability and quality with focus on a *major defect reduction effort*. This was characterized by lots of statistical process control, various six sigma programs, and a relentless search for "outliers" of measured distributions.

As the previously described eras were coming to a close, the reemergence of compound semiconductors was just beginning – generally with discrete transistors and then with the start of integration of monolithic microwave integrated circuits (MMICs).

This was also the time that compound semiconductor circuits were introduced. For the most part, the compounds followed the silicon pattern, starting most simply with MEtal Semiconductor Field Effect Transistors (MESFETs). Compound semiconductor materials were scarce, and expensive, so sample sizes were small, and the development folks immediately took up the reliability physics methodologies established by the preceding five eras of the silicon folks and focused on measuring distributions instead of large sample sizes.

Even though the reliability era template was well established and proven, the patience to spend 5 years on each stage of reliability was gone. So the first generation of reliability was compressed. There were some interesting parallels between the technologies, but since the timeframe was changed, we will consider each stage as the important aspect. While the techniques of the silicon cycles have been utilized, it is important to note that skipping a stage is not recommended. For example, some elements of the first three stages are natural; questions still remain in the basic material properties of the compound semiconductors and the noble metallizations used. Moore's Law has not driven developments for compound semiconductor interconnects or dielectrics, and those have been basically stagnant even though epitaxial material formulations have changed. Instead of

**Table 11.3** Stages of semiconductor reliability

| (Eras) Reliability stages | Short description | Volume mindset | Bathtub curve emphasis | Focus | Qualification styles |
|---|---|---|---|---|---|
| 1 | Materials | ↑ | ↑ | | Process integration compatibility |
| 2 | Mechanisms (Major reliability problems) | Low | Back (wearout) | Eliminating failures | One dimensional stress to failure |
| 3 | Physics (Major reliability physics effort) | ↑ ..... | ↓ ..... | ↓ ↓ | Success testing |
| 4 | Engineering (Major reliability engineering effort) | ↓ High | ↓ Front (early/ infant) | Predicting unreliability ↓ ↓ | Capability edges and simultaneous stressing |
| 5 | Defects (Major defect-reduction effort) | ↓ | ↓ | Eliminating failures | Continuous improvement |

shrinking features appreciably, the compound semiconductor industry has tended to squeeze performance out of epitaxial enhancements and blending of additional "new" materials in the substrates and epitaxy. The stages of "building-in reliability" and "major defect reduction effort" still have not been fully embraced by many of the compound semiconductor fabricators.

> Silicon is likely to remain the basic material, although others will be of use in specific applications. For example, gallium arsenide will be important in integrated microwave functions.
>
> Gordon Moore [1]

As a consequence of applying the historical eras of reliability toward mandatory stages of reliability development, we can produce a reliability prescription for each new technology. These stages are shown in Table 11.3.

For the most part, the first three stages have become second nature for compound semiconductor folks. Stage 1, the search of appropriate materials, is contained in tight proximity with stage 2, understanding of failure mechanisms, and with stage 3, reliability physics. In fact, stages 2 and 3 are easiest to accomplish by testing individual circuit elements and not products or complex integrated circuits. In other words, individual discrete transistors are common development samples for investigations into failure mechanisms and finding acceleration factors. The first three stages are synergistic with each other and with process development as shown in Fig. 11.1. By focusing on transistors or other individual circuit elements, the number of potential failure mechanisms can be segregated for easier analysis. Actual circuits, on the other hand, can jumble a multitude of various mechanisms together, complicating the analysis immensely.

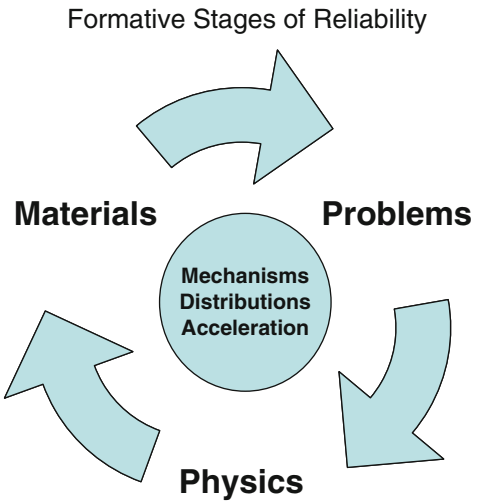



**Fig. 11.1** Diagram showing the first three stages of reliability

## 11.2 Stage 1 Materials

> The difficulty in defining quality is to translate future needs of the user into measurable characteristics, so that a product can be designed and turned out to give satisfaction at a price that the user will pay. This is not easy, and as soon as one feels fairly successful in the endeavor, he finds that the needs of the consumer have changed, competitors have moved in, *there are new materials to work with*, some better than the old ones, some worse; some cheaper than the old ones, some dearer.
>
> W. Edwards Deming [3]

From the third era of reliability, everyone has learned the methodology of determining acceleration factors and lifetime predictions. For compound semiconductors, the application of new materials is an attractive alternative to Moore's Law since feature sizes remain relatively constant as we learned in Table 11.4. Figure 11.2 indicates how semiconductor materials are used to achieve key performance attributes.

However, during the development of semiconductor processes, there are many more materials to consider than just the semiconductor. For example, contacts, interconnects, and dielectrics are a few of the integrated circuit elements than need material engineering – regardless if the process is shrinking feature sizes or changing semiconductor mixes.

Changes in materials are a common practice during semiconductor process development. Integration of various circuit elements requires careful compatibility verification and the application of all types of environmental stresses and acceleration methods. In general, materials are constrained by some level of capability. This constraint is easily envisioned, such as in Fig. 11.3, when considering mechanical material strengths.

**Table 11.4** Moore's Law doesn't exactly apply to compound semiconductors

| Technology | Improve performance by | Improve reliability by | Yield focus |
|---|---|---|---|
| Silicon | Shrink Physical size | New materials | Defects |
| Compound semiconductors | New materials | Continuous improvement | Parametric |

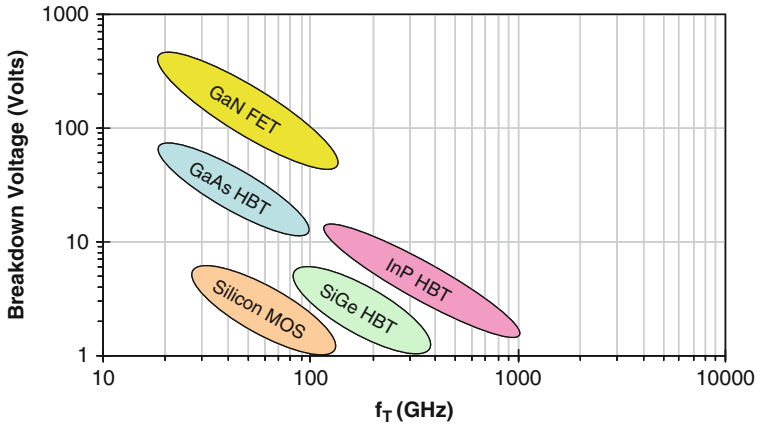



**Fig. 11.2** Performance attributes of operational frequency and breakdown voltage for some semiconductor materials

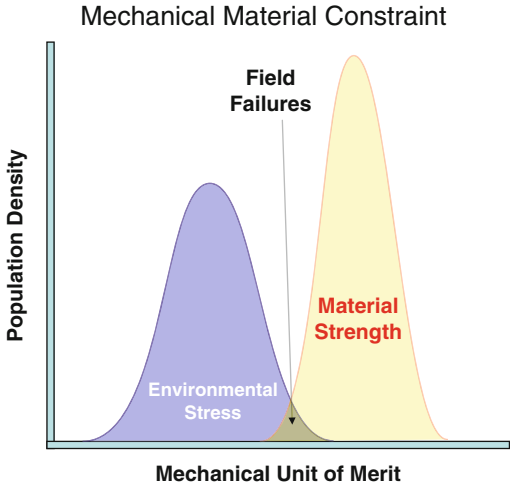



**Fig. 11.3** Overlap of distributions of environmental stress and material strength leads to failures in the field

The intent of material selection is to find materials that have no stress-strength overlap with all potential stress environments, including all types of electrical stress. So while it is easier to consider stress and strength for mechanical properties,

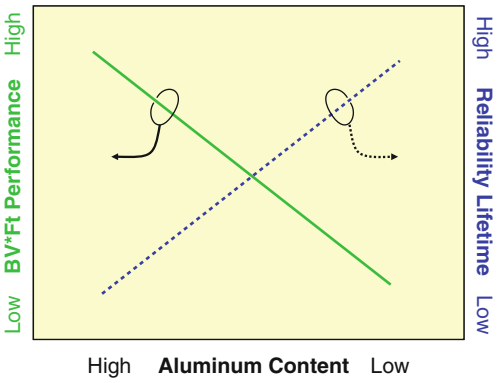



**Fig. 11.4** Performance versus reliability relationships for percent aluminum in GaN buffer layers

the reliability engineer must also consider all electrical properties and parameters as well.

Besides the stress-strength comparison of a single parameter of merit, the selection of material systems can sometimes result in trade-offs of desirable parameters that aren't directly comparable. When one of the trade-offs is reliability, then the estimation of absolute lifetime is immediately important. This need for a "number" drives investigation forward through the failure mechanism and physics stages so that acceleration factors can be applied. For example, consider the aluminum content in GaN buffer layers. Unfortunately, the reliability and performance are inversely proportional such that high aluminum content makes very high performance devices with very low reliability. Low aluminum epitaxy makes higher reliability devices with lower performance. As a result, folks immediately want to know the trade-offs in absolute units so they can select the material based upon the requirements of performance and lifetime of the application. In this situation, relative relationships, such as "GHz per year" are not useful regardless of the accuracy – see Fig. 11.4.

In reality, this trade-off problem is typically solved by finding a new material. So, when the trade-offs become important considerations for management, then absolute measurements are essential in terms of both performance and reliable lifetimes. This desire to know the numbers is a catalyst to move beyond material studies and into the next two stages of reliability.

## 11.3 Stage 2 Failure Mechanisms

From the third era of reliability, everyone has learned the methodology of determining reliability. The three keys to measuring and predicting absolute reliability are:

1. Knowledge of the root cause failure mechanisms
2. Measurement of degradation distributions
3. Characterization of acceleration factors

**Table 11.5** Relationships of materials-mechanisms-physics

| Mechanism | Material | Stress | Physics/model |
|---|---|---|---|
| Diffusion, physiochemical | Metals, semiconductor | Temperature | Arrhenius activation energy (exponential) |
| Corrosion, delamination | Metals, dielectrics | Humidity/ moisture | Peck (power) |
| Thermal excursion, creep, crack, TCE | Metal, semiconductor, dielectric | Cycle, shock, reflow | Coffin-Manson (power) |
| Mechanical | Metal, semiconductor, dielectric | Drop, shock, vibration, | |
| Electromigration | Metal | Current density | Black (power) |
| Dielectric breakdown | Dielectric, oxide, nitride | Voltage field | Crook |

Obviously, understanding mechanisms is the first key of the reliability physics measurement triad. The compound semiconductor folks used all of the previous knowledge available from the prior silicon reliability eras, but the failure mechanisms are not necessarily the same. This is not unexpected, since we have already acknowledged the use of different materials in stage 1. There are a multitude of failure mechanisms for compound semiconductors, and recognizing that there are more than just one mechanism is an important point (Table 11.5).

Here are a few of the significant compound semiconductor mechanisms in a relative chronological order of discovery:

### *11.3.1 Sinking Gates*

This diffusion mechanism is thermally driven. It is one of the oldest reported for FET-style compound semiconductors. This material diffusion mechanism has been substantiated with physical evidence for both MESFETs [4] and pHEMTs [5]. Because the sinking mechanism is so easy to accelerate with temperature, sinking gates are often identified as *the* cause of FET wearout. Even though some refractory gate metals have been aged and reported, titanium is still a common material for FET gates, and Ti diffusion with GaAs has a very high activation energy (the diffusion is exponentially accelerated with increasing temperatures). Even though the activation energy has been empirically measured at 2.56 eV [6], that high number is not unexpected based upon diffusion of other transition metals in GaAs; Mn = 2.49 eV, Au = 2.64 eV, and Ag = 2.27 eV. So if we were to make an estimate of the activation energy for Ti diffusion in GaAs from the ballistic model, it wouldn't be a stretch to 2.5 6 eV since the enthalpy of formation of a gallium site has been estimated by Van Vechten to be 2.31 eV [7]. This mechanism is quite unique to the Schottky contact FETs, and nothing like the sinking gate mechanism is found and addressed in the MOS-type world. Sinking gates and other

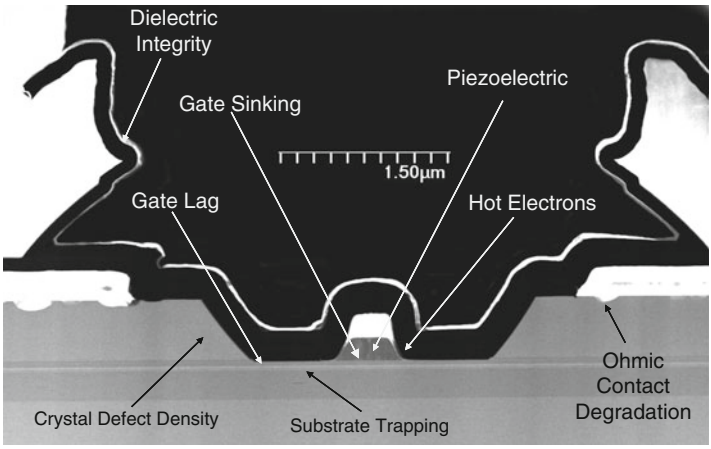



**Fig. 11.5** FET-style cross section with some of the physical locations for various failure mechanisms indicated

FET mechanisms are located in Fig. 11.5. The diffusion of Ti into compound semiconductors is so highly accelerated by temperature; it is arguably the mechanism which has driven the compound semiconductor folks' fascination with lifetesting at multiple high temperatures to predict wearout.

### *11.3.2 Gate Lag*

A common affliction for FETs. This phenomenon is thought to be a result of traps on the GaAs channel surface beside the gate. But is gate lag a failure mechanism or merely an ugly property of trap-filled compound semiconductor substrates? These trap studies have focused on traps in various locations which result in various effects such as drain lag, side-gating, back-gating, and substrate-gating. Investigations into the aging of traps have indicated no particular degradation over time [8].

### *11.3.3 Hot Electrons*

Borrowed from problems suffered by silicon gate dielectrics, this is a charging effect, sometimes called "power slump." This is indeed an interesting mechanism, since it is normally exacerbated by *low* temperatures – at least for oxide dielectrics. The degradation is caused by impact ionization at the gate edge, which injects hot carriers into the interface between the semiconductor and insulator. These carriers effectively expand the surface depletion layer and reduce Idss and gm without much of a change to the threshold voltage [9].

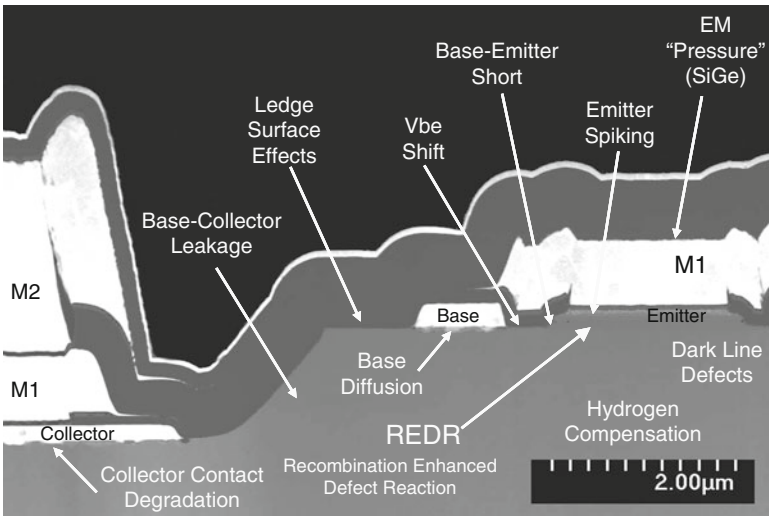



**Fig. 11.6** HBT-style cross section with some of the physical locations for various failure mechanisms indicated

### *11.3.4 Mechanical Stress*

Compound semiconductors are piezoelectric [10]. Local mechanical stresses will cause changes in fields and characteristics of active devices. Yet when hydrogen contamination was found to cause threshold shift degradation in FET devices (in 1989), this piezoelectric property was passed over for a more *sexy* theory "...conversion of gaseous hydrogen into atomic hydrogen in the Pt layer of the gate metallization, and the subsequent doping compensation and carrier removal from the active channel as the atomic hydrogen diffuses into the channel" [11]. It has taken more than 10 years to bust the original beautiful theory with the simple ugly fact that reduction of the gate metals (particularly titanium) in hydrogen produces a compressive stress that causes a piezoelectric polarization charge and results in the threshold shift [12]. This is another mechanism foreign to the silicon guys.

### *11.3.5 Electromigration*

This is an example of a failure mechanism born from silicon devices. Aluminum metallization is particularly susceptible to electromigration because of its relatively high resistivity. Aluminum also has a grain structure that makes it vulnerable to failure. Most compound semiconductors use gold instead of aluminum. The electromigration mechanism can be induced in gold under very high current densities, but

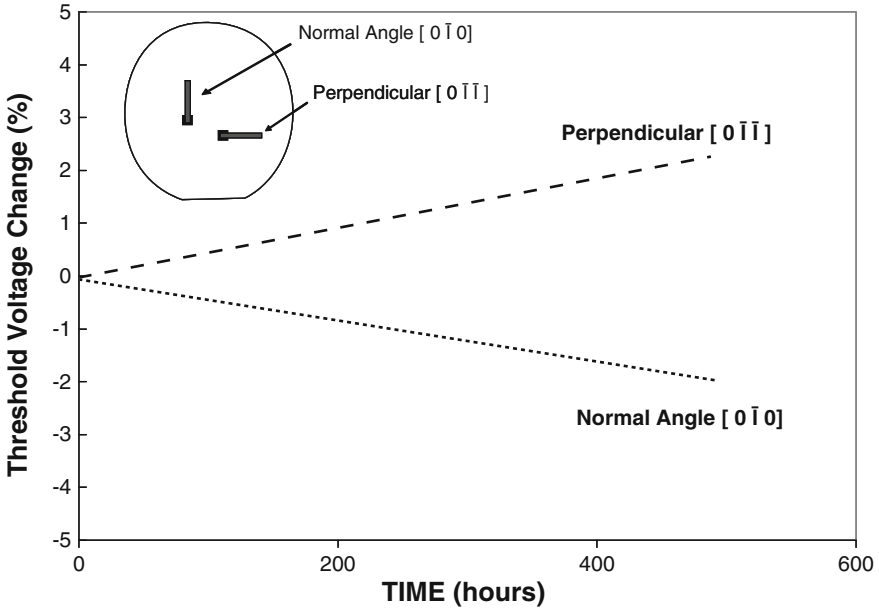



**Fig. 11.7** Threshold shift dependence on gate orientation for degradation on GaAs pHEMTs aged unbiased at 85°C and 85% R.H

electromigration has been largely disregarded for compound semiconductor interconnects at current densities below 500,000 A/cm$^2$.

On the other hand, current density does seem to have an accelerating effect on HBTs [13]. But there are two diverse mechanisms being theorized as to the cause. From the GaAs blend folks, the latest account for increasing base current is the "tunneling-recombination conduction mechanism" [14]. From the SiGe folks, "change in ß is attributed to electromigration-induced *pressure* on the emitter contact" [15]. There continues to be significant experimentation with thermal acceleration aging, and both theories speak in terms of activation energy instead of current density. HBT studies and results usually define current density conditions, and the current density is known to be the primary stress to accelerate degradation in HBT devices. In general, there are several mechanisms affecting HBTs, as shown in Fig. 11.6.

### *11.3.6 Corrosion*

Moisture ingression has only recently been considered as an issue for compound semiconductors [16]. Although gold and nitride are thought to be relatively inert, many of the compound semiconductor contact materials are susceptible to moisture

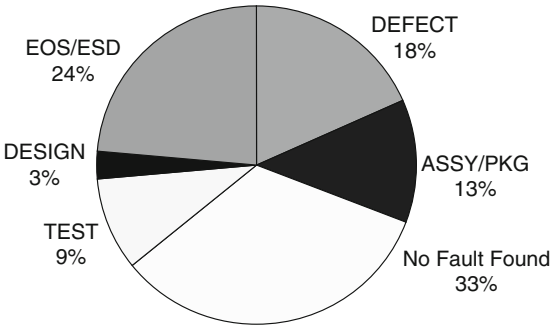



**Fig. 11.8** Results of root cause failure analysis of field returns 1999–2005

as are some of the semiconductor materials. In the past couple of years, humidity testing has come to the forefront. Once again, the mechanisms were found to be radically different from those characterized and modeled originally with silicon material families [17]. Nevertheless, significant lifetime improvements based upon various surface treatments are still being made for compound semiconductor FETs [18]. It is interesting that even within this short review of selected failure mechanisms, a 2004 look at moisture takes us back to the mechanical stress piezoelectric mechanism – see Fig. 11.7.

How does one define *importance*? What should be the priority for reliability testing? How do reliability engineers know "what matters"? For any situation, a good answer is to ask your customer. For a reliability issue, just look at what causes your customers to return devices. Figure 11.8 shows the causes of customer returns for a 5-year time span. Note that none of the problems that our customers report match with the first six mechanisms previously mentioned in the historical recap! This shouldn't be a surprise because most lifetime projections on those initial mechanisms don't predict any wearout for hundreds (even thousands) of years.

What can be more important than a customer failure? If the investigative stresses don't simulate what a customer does to the device, and the failures aren't the same type that our customers report, then how can "standard" methodologies predict reliability in actual use? Following are some of the actual causes of failures from Figs. 11.8 and 11.11.

### *11.3.7 ESD*

From Fig. 11.8, electrostatic discharge is one of the leading causes of problems reported by customers. This has been true as long as we have measured the causes. However, the compound semiconductor folks have tried to explain away this weakness as a technology difference or an education problem. ESD is definitely an important subject, and it is being discussed more openly in recent years [19, 20]. ESD is an issue that spans both silicon and compound semiconductors as the bulk size (volume) of transistors decrease. As an example, see Fig. 11.9 and consider

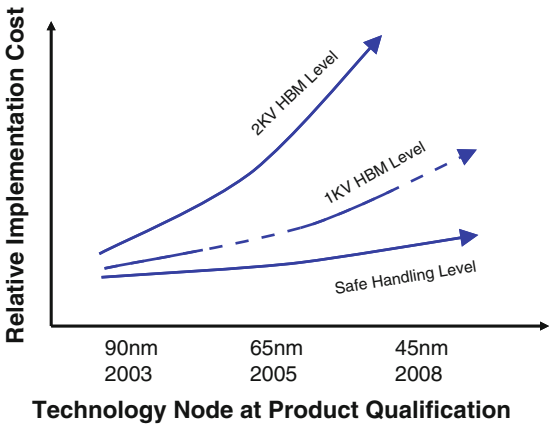



**Fig. 11.9** Cost argument to reduce ESD expectations. The projected cost of ESD requirements as a function of the calendar year and the silicon technology node for three ESD targets. Costs include chip area to implement ESD protection, performance degradation, and time to market

how the ESD robustness requirements have decreased in recent years under pressure of cost – once again driven by the decreasing physical sizes experienced by silicon devices under the influence of Moore's Law [1].

### *11.3.8 Defects*

Another leading cause of customer problems is defects. However, this is not as clear an issue as ESD. Defects seem to rise up in the Pareto chart as all the other causes get solved and processes mature. Assembly and packaging or testing issues will sometimes outweigh defect issues in the development phases of semiconductor technologies. Remember the fifth era of silicon reliability improvement? It was 5 years of "major defect reduction effort." So, acknowledgment of defects as an issue is a positive sign that they will be addressed appropriately.

The most commonly reported compound semiconductor defects are found in capacitors. Several compound semiconductor manufacturers have reported studies on capacitors [21–24]. The mechanisms, distributions, and acceleration factors are understood and similar to what has been found on silicon devices – particularly for the focused improvements on ever thinner gate dielectrics.

Other defects involving interconnect integrity and metal-to-metal leakages are newer issues for compound semiconductors. The distributions are familiar, but the mechanisms and acceleration factors appear to be specific to the technologies involved. Understanding these differences is likely to be one of the next challenges for reliability [25].

Reliability investigations and product qualifications are typically built on a set of "standard" aging methods which are designed to extract degradation that can be

statistically analyzed in order to predict lifetimes. The other perspective is to investigate what actually happens to devices during use by the customer, by performing detailed root cause failure analysis on devices returned by customers. This information is provided as an account of experiences with supplier-customer relationships. It is intended to describe expectations (from the supplier viewpoint) and to guide aging and qualification methodologies toward meeting those expectations. Comparisons to accelerated aging results are critical.

As we have noted, product reliability investigations typically include accelerated aging studies on population samples, evaluation circuits, or characterization vehicles with arbitrary failure criterion [26]. The best source of data is often overlooked – field returns. The customer uses 100% of the population under actual operating conditions while continuously measuring critical performance attributes. What better source of data exists than what the customer experiences? This customer fallout mechanisms are identified as "natural" because they are not tainted by acceleration, not altered by data manipulation, and not conditioned, treated, disguised, enhanced or artificial compared to any other source of reliability information.

Customer returns are of high interest because they contribute to the cost of doing business and because returns have a direct influence on customer opinions. The interactions and discussions of failure analysis results offer a prime measurement of customer satisfaction. Reliability engineers should seek out field failures to obtain a glimpse of reality. We can compare field return data to accelerated stress results for the same time period. This will help us add more failure mechanisms to our list.

Devices are returned from customer to supplier for various reasons. Sometimes the parts are excess inventory, demand decreased, or the need has changed. Other times, the packaging may not be useable, the quantity is wrong, or the label is incorrect. Of most interest are the devices sent back from the field which are the result of quality anomalies. These customer-returned devices can be divided into two main categories: product issues or administrative issues.

List 1: Examples of administrative issues

A. Wrong label on the part, package, or shipping container
B. Shipment errors involving containers, dry pack, or products
C. Paperwork issues involving the invoice, billing, part count, delivery time, packing list, order process, etc.
D. Stock rotation. Refresh of inventory at distributors. Date code expiration

The focus of this discussion is on product issues. Those issues would include anything that affects the form, fit, function, quality, or reliability of the actual product. This distinction is not intended to diminish the importance of administrative aspects of customer satisfaction, but rather to focus on technical issues of device performance and manufacturing factors of interest in semiconductor reliability. Our focus is to look at traditional reliability failure mechanisms.

Another concern about returns is in deciding what proportion of failures are actually measured and communicated to suppliers. An area of constant debate is the level of representation of customer returns. What portions of natural failures are eventually returned by customers? There are many factors which affect this answer. From a casual survey, it seems that returns are rarely a high percentage of actual failures. In other words, customers rarely return low-level or "routine" fallout. Generally, customers will have some level of fallout during the manufacture of a product. Fallout is defined as failure to meet expectations. But every customer seems to deal with the fallout differently.

List 2. A few of the possible factors influencing return rates are

1. Quality expectation by the customer
2. Cost to troubleshoot versus shotgun replacement
3. Cost to analyze product failures to root cause
4. Cost to repair products versus cost to replace
5. Cost of the components that fail relative to the repair
6. Cost to verify failures from fallout
7. Total volume of failures versus volume of fallout
8. Yield expectation of the end product
9. Trigger level for return to supplier action
10. Trigger level for epidemic failure action
11. Overall relationship with suppliers
12. Track record of specific supplier on corrective action
13. Rate of fallout compared to highest failure rate items
14. Cost to return fallout to suppliers
15. Fallout cost versus other supplier management costs

For example, if the cost of the product is low relative of the cost of repair, it is an easy financial decision to simply throw away fallout. If the product is modular or has multiple circuit boards, then these subcomponents might be cheap enough to throw away instead of repairing. Another example is a factory that accepts a consistently low level of fallout as a scrap cost without any additional analysis. In this case, it would only be an "epidemic" level of fallout which would trigger the offending products to be analyzed further. Of course, the epidemic failures would get attention and be the focus of corrective action activities – leaving the lower levels of fallout without enough resources to return to the supplier.

If component replacement is economical, the substitution still may not be an accurate identifier of actual failures. Some troubleshooting methods (nicknamed "shotgun") involve a form of guilt by association, where repair technicians get in the habit of removing certain parts for certain failure modes without actually confirming which individual part is causing the fallout. This practice could destroy evidence of bad solder joints, electrical shorts between connections, or simply an effect of "stacked" component marginalities among several related components. Knowing these particular shortcomings, many customers choose to scrap the fallout rather than return parts. In other words, the cost of determining the root cause of all fallout is too high. Without the accurate data of the actual root cause, then statistics are vague or nonexistent.

**Table 11.6** Example fallout levels to trigger returns

| Customer category | Product type | Volume | Claimed trigger | Actual trigger |
|---|---|---|---|---|
| Automotive | PA | Low | > Zero | Every failure |
| Standard product | Various | Large | 100 DPM | 100–2,000 DPM |
| Cell phone | PA | High | 100 DPM | 200–2,000 DPM |
| Cell phone | RF | High | 100 DPM | 20,000 DPM |
| Long haul | Digital | Medium | 100 DPM | 20,000 DPM |
| PC/LAN | LNA | High | 200 DPM | 20,000 DPM |
| Cell phone | PA | High | 40 DPM | 20,000 DPM |

**Table 11.7** Definition of volume levels for semiconductors

| General description | Volume | Units shipped/year |
|---|---|---|
| Laboratory/space | Research | Less than 1,000 |
| Prototype | Small | 1,000 |
| High reliability/military | Low | 10,000 |
| Automotive | Medium | 100,000 |
| Industrial | Large | 1 Million |
| Commercial/consumer | High | 10 Million |
| Mass production | Very high | 100 Million |
| "Killer application" | Ultra high | 1 Billion |
| Nirvana | Ludicrous | 10 Billion |

In effect, each one of the factors shown above is an excuse or a reason for not returning natural failures for analysis. However, if a customer has high expectations for quality or if that customer is particularly interested in continuous improvement, they may return higher percentages of fallout or analyze all fallout and return every single failure. In fact, the most concerned customers may even perform some root cause failure analyses internally, without returning all devices to the supplier. Table 11.6 shows examples of return representations by our most discerning customers. This table was based upon experiences with more than 100 different customers [27].

By examining customer categories, business areas, and sizes, there is no particular aspect that defines their propensity to return devices. So far, there are no categories, statistics, or logic which can describe or predict which customers will return which levels of fallout. Clearly, the expectation for semiconductors is heading toward zero fallout or "perfect quality" [28]. However, there is one essential ingredient that is required to increase the likelihood of customer fallout and information – volume. Many smaller customers cannot return field fallout because they simply do not have experience with enough devices. This is an extension of a supplier's assumption that accelerated reliability tests on 77 samples will represent a customer's experience. In reality, the detection level is purely a function of sample volume.

For the purpose of discussing volume levels universally for semiconductors, Table 11.7 is presented. By definition, sample sizes exceeding 1,000 devices are needed to begin to detect fallout, and samples of 10,000 or more are needed to break

into the realm of measuring at the 100 DPM, or lower, failure rates. At the time of this study, 100 DPM is a common level of concern, but this is expected to change as customer expectations are continually improving.

What is a description of a "natural" failure mechanism? There are many possible failure mechanisms for any given device. Standard reliability investigations tend to focus on wearout mechanisms which can be accelerated by temperature, bias conditions, humidity, thermal transitions, or other accelerated aging techniques. In contrast, the natural failure mechanisms are rarely (if ever?) caused by wearout mechanisms.

Experience shows that the categories of natural failures follow a progression associated with the development of the technology. For example, early in development, most fallout is associated with application-type mechanisms. These include electrical, thermal, and mechanical overstress. This is the time when sub-suppliers of raw materials are also blamed for many of the fallout and failures. As the products begin to find homes in the market and volumes increase, the mechanical application issues become dominated by packaging and assembly issues. Eventually, testing, test coverage, or test correlation issues are encountered. Typically, the last type of failure mechanisms is related to defects. Even though the defects are likely to be present during the whole development process, they won't become apparent until the previously mentioned mechanism types are discovered and corrected down to low enough levels so that defects are noticed. It should be noted again that volume is the primary driver of detection levels among all the failure mechanism types [29].

### *11.3.9 Electrical Overstress Mechanisms*

Generally, electrical overstress can be categorized into two types: electrical overstress (EOS) and electrostatic discharge (ESD). Basically, ESD is defined as a short duration phenomena (less than 1 uS), and EOS covers everything else of a longer duration: from a part plugged in backward, to noise spikes on a power supply, to severe mismatch of output terminations. Often, true application errors include timing sequences which exceed specified limits or even external electrical stimulus which is not specified.

During analysis of a field return, the determination of EOS or ESD can be very subjective. Often, this determination can be a relative measure of the extent of the damage. For example, pure ESD failures are often invisible upon optical inspections up to 5,000×. But these mechanisms can overlap when an ESD event results in a succession of cascading reactions that end up severe catastrophic damage. There are other mechanisms, such as defects, that cause EOS-type signatures. The root cause determination is difficult because the structures involved are obliterated. The analyst is left to make judgments of the debris based upon experience or from similar results that are obtained during experiments to duplicate the failure events.

### 11.3.10 Thermal Overstress Mechanisms

Thermal overstress is nominally manifested in damage to packaging or assembly components. In general, semiconductors can survive overstress conditions exceeding 350°C with no observable degradation. However, plastic packages will suffer from "popcorn" damage above 100°C, and they will begin to "melt" above their glass transition temperatures (~160°C). Both of these thermal overstress conditions are intentionally exceeded during the normal solder reflow process. Device self-heating can also be a cause for thermal overstress if special heat sinking is expected but neglected or if normal thermal grounding is omitted by die attach or package soldering anomalies.

### 11.3.11 Mechanical Overstress Mechanisms

Mechanical failure mechanisms can also occur on the packaging materials or on the die itself. Compound semiconductor devices are notoriously sensitive because the crystal is more brittle than silicon. Since compound semiconductors are more thermally resistive than silicon, they are often thinned considerably, which increases the susceptibility to mechanical damage. Damage can easily be induced by die eject tools, pick and place collets, or by flexing of finished packages. Often, mechanical removal (for repair) or rough handling of packages can result in mechanical damage that is subsequent to the actual failure mechanism.

Some compound semiconductor processes still employ airbridge interconnects and/or top-surface interconnects which are especially susceptible to mechanical damage from pick and place collets or wire bond damage. There have also been reports of damage to gold interconnects by filler materials used in injection molding assembly operations for plastic encapsulated devices or overmolded laminate modules.

### 11.3.12 Design Mechanisms

Although design mechanisms are rare, they can be discovered among field returns. Typical errors include layout, spacing, contact, or interconnect stacking violations; inadequate margins for voltage fields or current density; and omissions in simulation of use conditions of proper terminations, bias sequences, or deployment of known ESD countermeasures.

### 11.3.13 Test Mechanisms

The most commonly caused test failures are combinations of the physical and electrical test operation with the application issues already described. For example, mechanical issues involving the formation of contact between the component and the electrical test socket. These socketing operations could induce ESD, EOS, or mechanical damage to the leads, package, or die. These application-type errors can also happen between test and shipping operations such as in tape and reel or other kitting or packing processes. Additionally, test mechanisms can evolve from errors of omission, such as test coverage, fault coverage, or determination of device stability. One of the more embarrassing (but still common) test escape problems is the mix-up of passing and failing samples in the posttest populations.

### 11.3.14 No Fault Found (NFF)

If nothing anomalous can be found electrically or mechanically with a return, it should be sent back to the customer to confirm it is a failing device. Without the confirming action, there is a possibility that the return is a correlation mismatch between the customer's expectation and the supplier's ability to measure.

As customers and suppliers work to develop return programs for low-level fallout, engineering materials, or samples, they should both recognize that the probability of NFF devices will increase. The communication of NFF data and the verification action on samples is needed to improve and refine the troubleshooting and fallout detection methods at the customer factory and field repair centers.

### 11.3.15 Assembly and Packaging Mechanisms

Table 11.8 shows the common failure mechanisms that are encountered through various types of assembly and packaging situations. During the development process, these mechanisms can occur each time a new subcontractor is utilized. An extensive set of experiments is needed to optimize all of the process parameters and equipment that is utilized in an assembly process.

### 11.3.16 Other Mechanisms

The other types of fallout include untestable devices, samples that are destroyed during analysis and undetermined causes.

**Table 11.8** Assembly and packaging failure mechanisms

| # | Assembly | Failure mechanisms |
|---|---|---|
| 1 | Wafer carrier, wafer on tape, waffle pack, gel pack, tray | Shipment damage, loose chips, cover slip issues, tweezer damage, particles, surface damage, die eject force, die pickup/placement force |
| 2 | Cavity package: ceramic, glass, frit, epoxy, alumina, aluminum, metal, welded, or hybrids | Die eject force, die pickup/placement force, hermeticity, hydrogen contamination, RGA, die attach (solder or epoxy), wire bond parameters |
| 3 | Plastic package: plastic encapsulated device, injection molded with leadframe | Die eject force, die pickup/placement force, moisture ingress, wire sweep, Tg stress, wire bond adhesion, leadframe TCE mismatch, leadframe plating, wire bond force/energy |
| 4 | Laminate-based modules, SIP | Die eject force, die pickup/placement force, moisture sensitivity level (MSL), mold fill, popcorning, die crack, conductive epoxy bleed out |
| 5 | Flip chip | Die eject force, die pickup/placement force, underfill, pillar/bump planarity, thermal management, flux |
| 6 | Chip scale, wafer scale, SOC | Die eject force, die pickup/placement force, die hermeticity, TCE mismatch, mechanical integrity, electrical connectivity |

### *11.3.17 Defect Mechanisms*

The most common defect types involve capacitors with underlying anomalies which perturb the dielectric such as particles, stains, or voids. These are followed by similar anomalies which affect metal – usually with shorts or leakages caused by filaments, debris, or damage – either interlayer or intralayer.

Let's compare accelerated and natural failure mechanisms. Although the types of mechanisms observed during accelerated testing are similar to those seen in the field fallout, the proportions are somewhat different. Most all fallout experienced by customers has occurred in factories, immediately after assembly into systems or products. The "aging" at this point is primarily a simple solder reflow and application of nominal bias. The comparison of rates of return and rates of fallout in aggregated accelerated aging makes the assessment of return representation possible. With data from reliability monitoring and fallout data from customers who are most diligent with returns, we can see that fallout is occurring that is not returned. For example, the internal fallout rates compare favorably with our most diligent customers (see Table 11.6). While we expect that customer applications and detection methods may vary – it would not be enough to explain why the return rates vary from zero to 1,000 DPM. Some high-volume customers admit they cannot return every single device. Therefore, from lack of reported returns, it is assumed that most customers employ the factors shown in List 2 to "filter" the total fallout before returning samples to their suppliers.

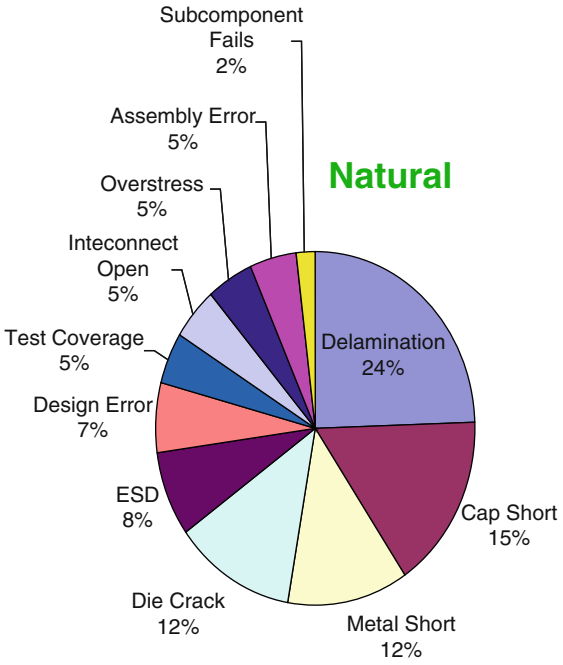



**Fig. 11.10** Causes for recent customer returns, the natural failure mechanisms

For a direct comparison of internal and external failure causes, an example group of 100 recent failure analyses were compared with all of the accelerated test fallout of that same period. Of course, there is some delay between the manufacture of production and the in-service timing of customer implementation. Our experience shows that on average, this delay can typically be 6 months. The causes of returns are shown in Fig. 11.10, and the causes of internal accelerated testing are shown in Fig. 11.11 [30].

Upon comparison of Figs. 11.10 and 11.11, the failure causes were found to be remarkably similar in type and proportion. The exclusive natural mechanisms for the returns were die cracks, ESD, test coverage, overstress, and some subcomponent fallout. The accelerated mechanisms had corrosion and bin mix which were not duplicated in the returns. In subsequent failure analyses, these mechanisms which were momentarily exclusive could be found from both sources, albeit in somewhat different proportions. For example, natural corrosion failures are extremely rare.

Had there been more significant differences in the failure mechanisms, this would have been an indication to change accelerated test conditions to improve the overlap and proportion of mechanisms. One of the reasons of conducting accelerated tests is to simulate customer experiences. For example, if accelerated testing could produce the same mechanisms seen by customers, then the internal test fallout could compensate for the low-level fallout that is sometimes unreported or not returned from customers.

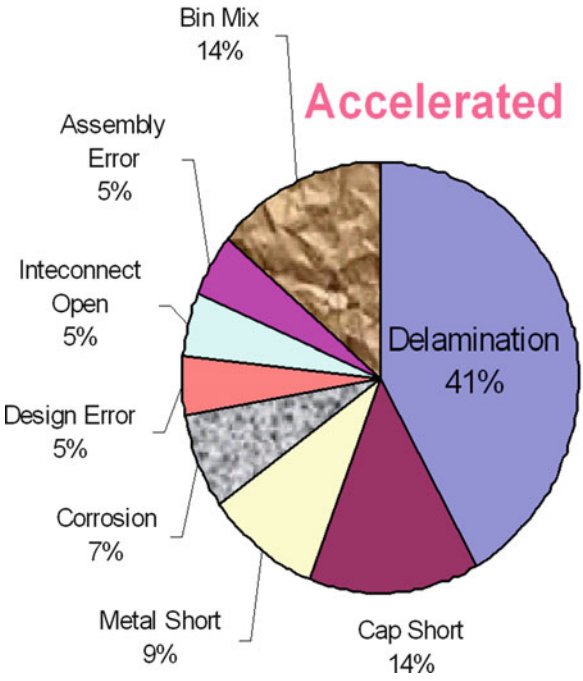



**Fig. 11.11** Causes for recent accelerated lifetests, the accelerated failure mechanisms

It is a remarkable result to find the matching of customer's feedback from so many devices delivered in production with natural fallout results to accelerated results on more than one million samples. It is only through root cause failure analyses that comparisons such as this can be made.

Comparisons of customer returns and accelerated test fallout are influenced by dynamics of the customer returns process. Relationships between customers and suppliers are critical to effective analysis and continuous improvement. These relationships and the ensuing discussions have evolved a completely different perspective about customer returns. In small volume start-ups, customers were "blamed" for causing nonrepresentative failures and being abusive to the samples [31]. From the higher volume data, the customer return is now considered the most valuable source of relevant failures to focus improvement efforts and develop optimum accelerated test conditions. The following points are concluded:

- Root cause analysis and corrective and preventive action on natural failure mechanisms are highly desired.
- Customers and suppliers must communicate and work together to obtain the information so as to reduce fallout and occurrence of quality and reliability excursions.
- The types and mixture of natural failures have changed over time and are expected to change as technology evolves.

- There are more similarities between recent accelerated and natural mechanisms than expected.
- The exclusive natural mechanisms should be addressed to make future acceleration studies more effective and reliability investigations more representative of actual customer experiences.

## 11.4 Stage 3 Reliability Physics

In describing the first two stages, we have referred to the three key aspects of measuring and predicting reliability using the physics approach which includes a good knowledge of the root cause failure mechanisms, measurement of degradation distributions, and characterization of acceleration factors.

So far, we have shown that the first three stages are intertwined by the reliability physics approach such that each material can have multiple mechanisms that are characterized by:

- The application of stresses
- The measurement of resulting degradation
- The determination of how various stresses accelerate the mechanisms

Degradation distributions and statistics are important tools for studying reliability physics. To understand reliability physics, folks need to know about distributions. Significant degradation is needed during aging, and the ability to separate fallout by each mechanism is often the issue. This can be especially difficult in a complex IC where multiple mechanisms are all degrading at various rates (and sometimes in opposite directions). Reliability engineers tend to immerse these data analysis problems with complicated statistical formulas and probability estimates. While this approach may seem to bring out unbiased and objective interpretations of the results, it often is where the non-reliability person gets lost. Unfortunately, the reliability engineer often loses sight of what might be most important for customers.

Even though reliability engineers and statisticians will enjoy analyses of distributions using normal, lognormal, Weibull, Poisson, maximum likelihood, or some other techniques, they all tend to show nearly identical (albeit astronomical) results after the acceleration factors are considered. In other words, small differences between distributions aren't significant when each method projects median lifetimes that are 180× longer than the age of the universe!

The important point to keep in mind is that the graphical representation (a line predicting reliability) is only a surrogate for the distribution data that generates the line. See Fig. 11.12 for a different viewpoint of the reliability prediction line. As the purity of the distribution is contaminated by overlapping failure mechanisms, then the accuracy of the prediction suffers.

The important thing to note is that degradation is necessary to apply the reliability physics methodology. Lack of degradation means no data.

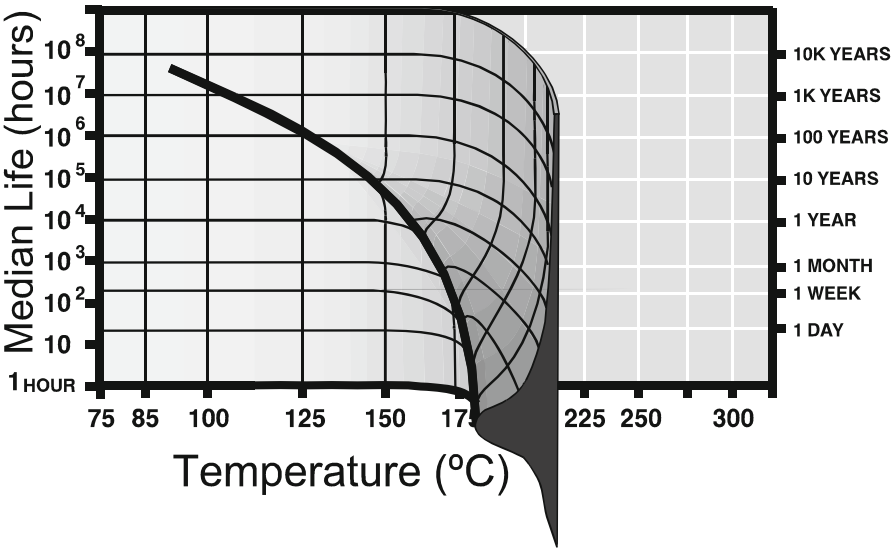



**Fig. 11.12** Representation of distribution data that underlies simple lines showing median life dependence on temperature

Once the distribution is known, how do we make use of it? One of the most controversial issues in reliability is interpretation of results. For example, reliability engineers speak in terms of "median" lifetimes – when *half* of the population will fail. The issue is that customers want to know when the *first* part will degrade. Typically, this question begins another mathematical diversion into confidence levels and probabilities – more gobbledygook for those not familiar with reliability. A not-so-obvious answer could be to reduce the dispersion of the distribution, so that the time to the first failure and time to 50% failure are closer together. Another idea would be to dump references to median time, and report data for the first failure, such as a $t_{ppm}$ (time to first part per million fail) instead of $t_{50}$ (time to 50% failure). Once the distribution shifts from the "middle" to the "front edge," any separations between quality and reliability start to come together – as was signaled in historical silicon reliability generations; this is an indicator of movement from the third to the fourth era.

Regardless of how the data distribution is incorporated in the reliability results, it must be first determined and scrutinized. Applying statistical analyses blindly without understanding the distribution is a very common mistake. Often, experimental results include multiple distributions because multiple failure mechanisms are causing degradation. Each mechanism and distribution must be separated for a proper analysis. Likewise, analysts should not necessarily assume the distribution is always "normal" or Gaussian.

If we are to make predictions about reliability, we must be able to accelerate the failure mechanisms without generating new issues or masking mechanisms that may exist under reduced stress. Thermal acceleration is easy, but that doesn't

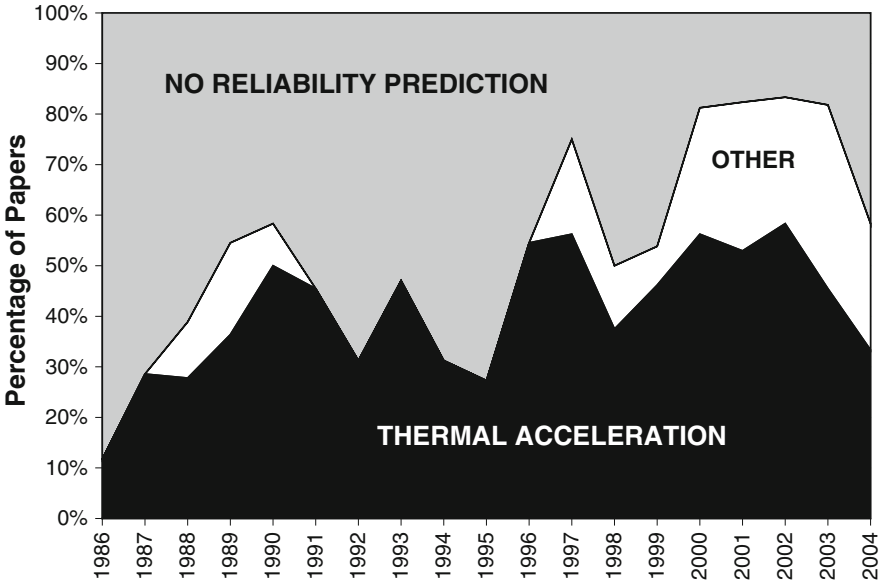



**Fig. 11.13** Type of acceleration reported at the Reliability of Compound Semiconductor Workshop by year

always make it the right thing to do. Understanding of mechanisms and degradation distributions can help us to look where acceleration might be less understood. To help, we need to talk more with our customers. Their use of the devices can give us clues as to what types of acceleration are more applicable. For example, thermal excursions, voltage, current density, and humidity might be preferable to high temperature acceleration. In other words, almost every part gets attached by a solder reflow (an extreme thermal excursion). Hopefully, they will all turn on (experience voltage fields and current flow). And eventually, they are likely to be exposed to an everyday environment (humidity).

We can investigate the applicability of various stresses by talking with customers or by looking into stresses which are most effective at causing degradation. Remember, reliability folks need some kind of degradation to measure and predict metrics, so effectiveness means: How well do aging tests cause failures that match our customer's stress? Since we've already admitted that changes in reliability are meager, looking into stress effectivity is useful to guide our improvement efforts.

With few exceptions, the reliability investigations on compound semiconductor circuits over the past two decades have evolved to rely on thermally accelerated wearout failure mechanisms – see Fig. 11.13. Regardless of the measured lifetimes, there have been very few wearout failures reported during use of the circuits. Typically, the application issues of ESD, assembly issues of interfacing the devices to systems, and manufacturing and logistics issues of getting the packaging and shipping systems synchronized will capture most of the attention. Buried under

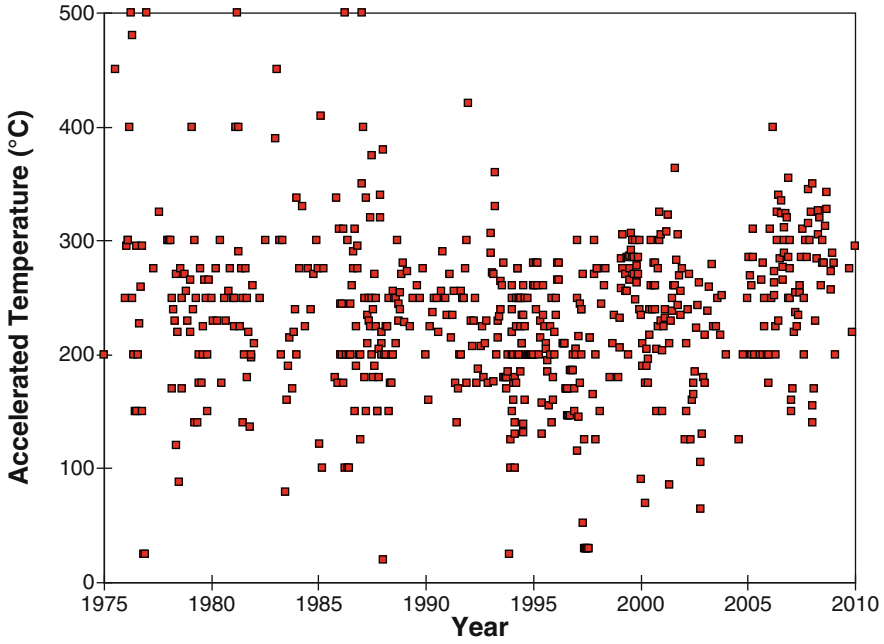



**Fig. 11.14** Reported accelerated test temperatures; 92% participation

these initial start-up problems, customers do report measurable defect rates and early life failures that often match up with yield fallout failure mechanisms. For the highest volume (>100 M/year) applications (such as cell phones), the early life mechanisms are significant. These volume applications have brought attention to the fifth era for compound semiconductors.

Figure 11.13 shows the historical fascination with thermal acceleration by compound semiconductor aging. Perhaps the reason for the popularity of thermal acceleration testing can be illustrated by the following compendium of data from the Reliability of Compound Semiconductor Workshop and other reliability publications [32]. Specific data on the actual accelerated test temperatures utilized is shown in Fig. 11.14. The data for Fig. 11.14 and the following seven summary graphs were collected over a 35-year history from 1975 to 2010 [33]. In all, 1,300 reliability publications were researched regarding applicable information. In order to be included in Figs. 11.14, 11.15, 11.16, 11.17, 11.18, 11.19, and 11.20, publications had to include at least one of the following parameters:

- Temperature of accelerated test
- Duration of accelerated test
- Sample size
- Estimate of use temperature
- Activation energy of primary degradation mechanism
- Projected lifetime at use conditions

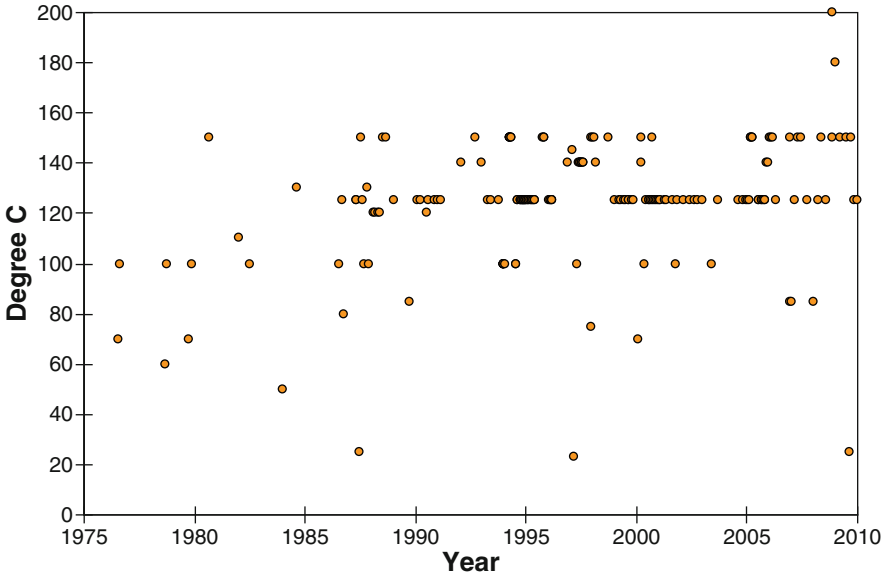



**Fig. 11.15** Reported temperature of operation (use temperature); 28% participation

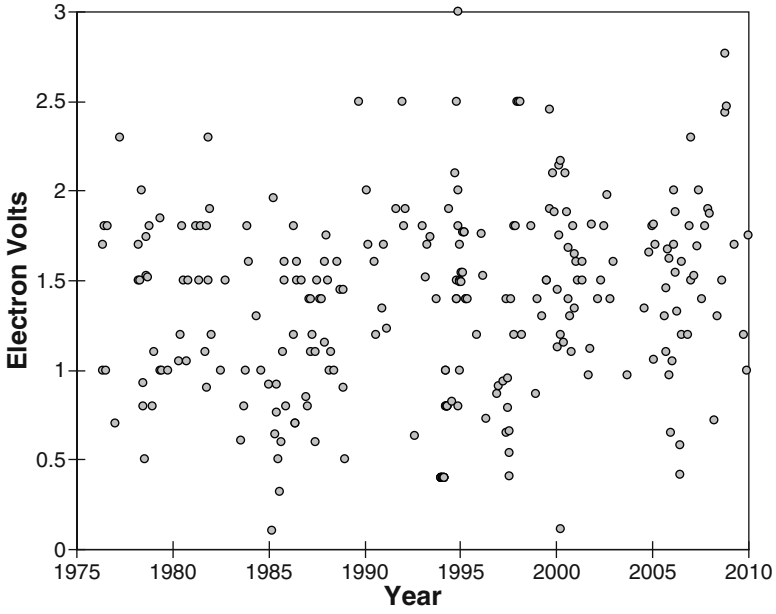



**Fig. 11.16** Reported activation energies; 39% participation

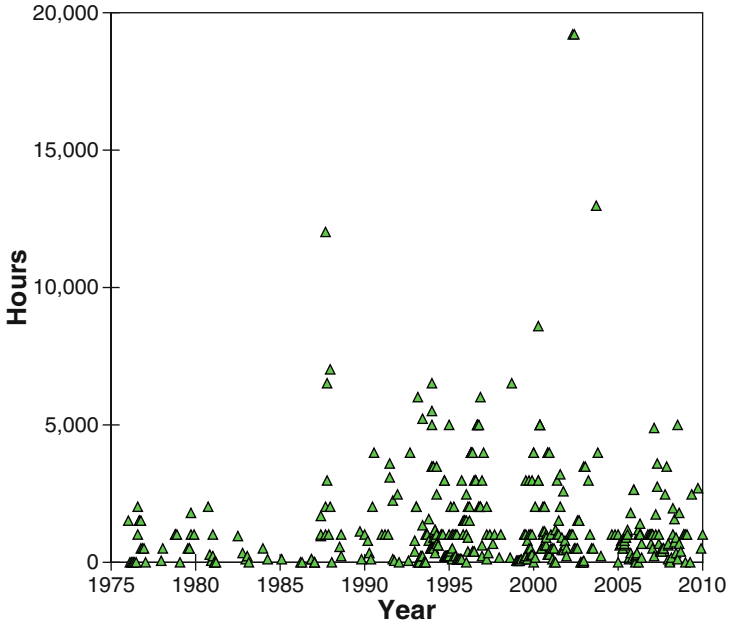



**Fig. 11.17** Reported test durations of reliability aging; participation = 60.1%

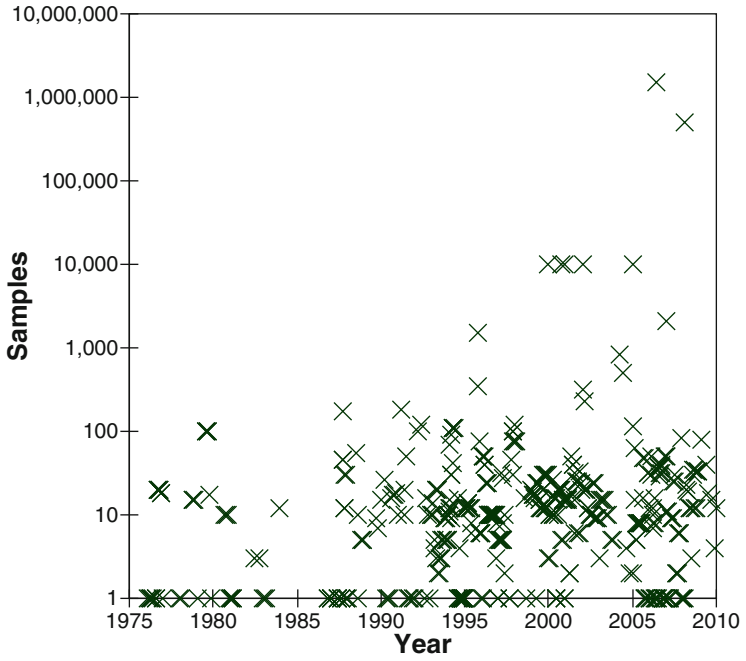



**Fig. 11.18** Reported sample size for reliability testing; 67% participation

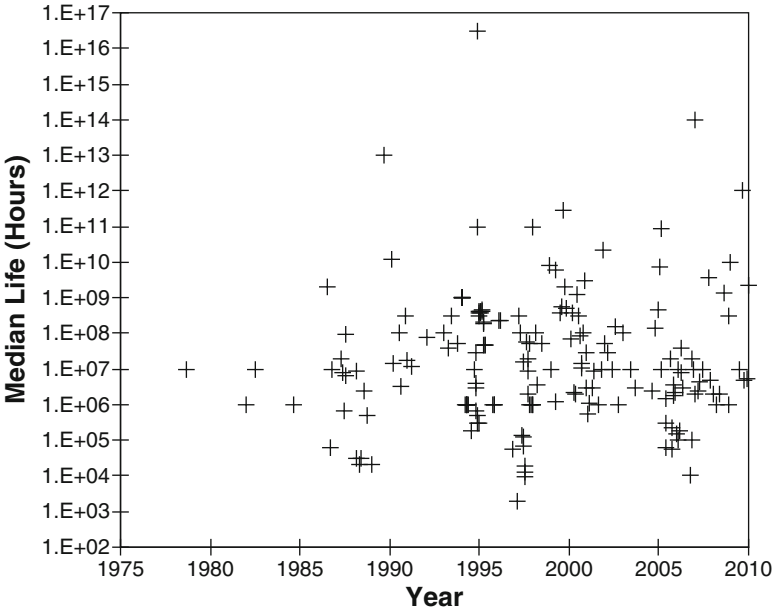



**Fig. 11.19** Lifetime predictions reported; 26.6% participation

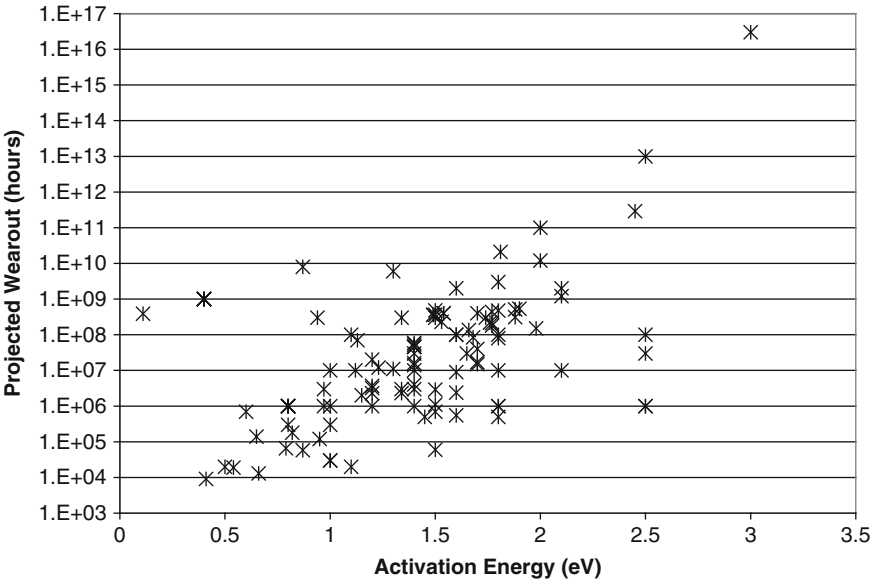



**Fig. 11.20** Participation is 19.2% in both projected life and acceleration factor

Of all the 1,300 compound reliability papers examined, only 673 included at least one of the six applicable data parameters listed above. The primary source of the reliability results data is the ROCS Workshop, but most of the compound semiconductor reliability reports from the International Reliability Physics Symposium, Compound Semiconductor Manufacturing Technology (ManTech) Conference, and the Compound Semiconductor IC Symposium are also included. These data represent much more than 372 million device hours of testing. The most common reported parameter was the temperature of the accelerated test. As shown in Fig. 11.14, 92% of the papers included the stress temperature.

Although testing has ranged from 0 to 500°C, most of the evaluations in recent years have been held between 200 and 300°C. While most accelerated stress publications will report the test temperature, few of the studies will report the expected operating temperature. Figure 11.15 shows the operating temperatures that have been reported. It is obvious that of those reporting use temperatures, the consensus is becoming 125°C. The median of this data is 124°C and the average is 123°C.

Figure 11.16 shows the reported activation energies over the review period. We would not expect this graph to have common results since each compound semiconductor fab has unique recipes and materials. The various thermal acceleration factors demonstrate the diversity of failure mechanisms.

Figure 11.17 shows the third most commonly reported parameter, duration. In all, just over 60% of the publications mentioned the length of the applied stress. As a result, the average length of test was 1,388 h. There were a few particularly long tests, which have pulled up the average, but it is nice to see that most tests are running at least 1,000 h.

Figure 11.18 shows the sample size of parts aged. The sample size was reported in 67% of all papers researched, making it the second most highly reported parameter. The average here is a little embarrassing at ten samples. In spite of a couple of million piece evaluations and several publications utilizing 10,000 devices, the sample of devices is fairly small. These samples will need to be increased to provide statistical significance to measurements nearing zero defects.

Figure 11.19 shows the predicted lifetimes for a few of the referenced publications. It is a bit surprising that only 26.6% of all the publications included a prediction. Overall, the median predicted lifetime is just at 10 million h, or approximately 1,000 years.

Figure 11.20 shows the relationship between reported activation energies and predicted lifetimes. This is the most interesting graph to ponder. Should there be any relationship expected? If yes, should it look like this one?

From the reported history, there is an obvious preference for thermally accelerated lifetesting. The median accelerated test temperature of compiled results is 235°C, and the range is decreasing with time. The average of the expected use temperature is 125°C, and the data indicates that there is now a "standard" of 125°C. The history of all results indicates a threshold for projected lifetimes to exceed 1 million h (114 years). This "requirement" has become an expectation through competition and example. Even though the desire to reach 1 million h

lifetimes is not driven by the historical data, the data does favor activation energies above 1.5 eV to ensure that goal. However, it seems remarkable that less than one-fifth of all reported results actually include both an activation energy and a projected lifetime!

The average reported lifetime is 2.1e14 hours, or more than one hundred million times 1 million h. This is a ludicrous number, and perhaps that is the reason that sample sizes and tests durations have remained woefully low. The sample size goes unreported in most recent testing, and the median is only ten parts per test. The average test duration is 1,388 h, which seems insignificant compared to the predicted lifetimes.

While examining all the reported results, one would be hard-pressed to find indications of overall improvements in methods or results over the past 35 years. There are a couple of de facto standards that are apparent, such as the test temperature and, particularly, the use temperature.

Accelerated tests are used to predict reliability when such predictions are available from no other means. Accelerated tests can sometimes help us find failure modes or mechanisms that ought to be explored to see if they might occur in ordinary use. But beware of the use of inappropriate accelerated tests that do not match the results that our customers experience in the field.

Why have most aging experiments on GaAs devices involved accelerated temperature? Temperature was originally studied by the Swedish physical chemist Svante Arrhenius back in 1887 because it has a universal acceleration effect on chemical reaction rates of approximately 2× per 10°C [34]. Of course, this "universal" effect is apparent in laboratory experiments that exhibit similar activation energies. Because of Arrhenius, by convention, out of habit, or from standardization, the accelerated temperature test is most commonly applied today – see Fig. 11.14.

The use of reliability predictions based on temperature could be sufficient if the failure mechanisms of the accelerated tests matched the failure mechanisms that are experienced in normal use. For the most part, the "wearout" mechanisms involve diffusion-type degradation, which is typically physically manifested by phenomenon such as "sinking gates" [35, 36] or perhaps metallic interconnect electromigration [37, 38]. However, measurement of these mechanisms in this manner is actually self-fulfilling as they result from a syndrome described by the popular fable of "looking where the light is." So goes the story:

> A drunk lost his car keys at night and was searching frantically for them near a lamppost. A passerby attempted to assist him. After a long search around the lamppost to no avail, the passerby asked "are you sure you lost your keys here?" "Nope," the drunk replied. "Then why are we looking here?" asked the helper. The response was, "I can see my way around under this lamppost because the light is better here."

Fortunately, high temperature lifetesting is not the only source of reliability data. A ready source of device failures comes from field returns. Experience has shown that customer returns have some dependency on volume. If devices fail during engineering, prototyping, or initial development phases, they are rarely returned to

the manufacturer. The tendency to capture and return fallout increases with volume. For example, a single failure out of a 100, 1000, or even 10,000 devices doesn't generate nearly as much customer interest as a 100 failures per day on a volume production line – even though the failure rates could be identical or even less in volume production than in the lower volume scenario.

So what does the accelerated lifetest predict as an earlier failure mechanism? This is the problem – accelerated lifetesting on small samples generally does not indicate any other mechanisms exist, except wearout or an occasional ESD event. If normal convention is applied and pure thermal acceleration is utilized, a failure mechanism is identified and predicted not to occur until our next millennium celebration! What is the alternative?

If there weren't ever any failures caused by mechanisms other than wearout, reliability engineers would end up like Maytag repairmen. However, accelerated tests other than high temperature lifetests *do* experience real fallout and, occasionally, failures. The gap must be filled between testing for wearout and the reality of earlier extrinsic failures. One method is to utilize different types of acceleration, for example, thermal excursions (cycling), voltage, and moisture. Real world experience shows the "effectiveness" of various accelerated stress tests in terms of their ability to cause fallout. This fallout can include ESD, mechanical damage, and test errors, but "real" non-wearout failures are included as well. Using a suite of qualification tests, such as preconditioning, temperature cycling, solder reflow simulations, thermal shock, and both biased and unbiased humidity testing, will generally be effective in evoking various failure mechanisms over time – especially when applied to very large sample sizes.

It's important to note that the effectiveness of each stress varies, particularly when causes of failure are corrected upon discovery. So, why don't these real field failure mechanisms show up in accelerated lifetests? The answer regarding any fallout is elusive until the volume picks up. It is only through the sheer volume of sample size do the non-wearout failure mechanisms show up in thermally accelerated lifetesting. If only a few hundred parts are tested every month, it may take a year to run across a failure caused by a mechanism other than thermal wearout. With only one "unexplained" failure per year, the device is more likely to be thrown out, than get analyzed down to the root cause. Additionally, some of these "real" mechanisms are related to manufacturing anomalies, and they don't wait for wearout. Following are a few examples of discovery regarding defects.

Although thermal acceleration is most commonly applied to transistor wearout, each discovered failure mechanism presents new opportunities for learning. One such new mechanism involved the integrity of multilayered interconnects or vias. This mechanism was discovered by a customer at the crossing of another one of those volume thresholds – ten million units in the warranty period. In this case, we found that field failures could be duplicated first by utilizing larger sample sizes during reliability stressing and then by applying a different form of acceleration. Results of inducing fallout by utilizing various types of thermal shock and temperature cycling methods are shown in Fig. 11.21. Note that this special population of devices had remarkably high fallout during any of the applied conditions of

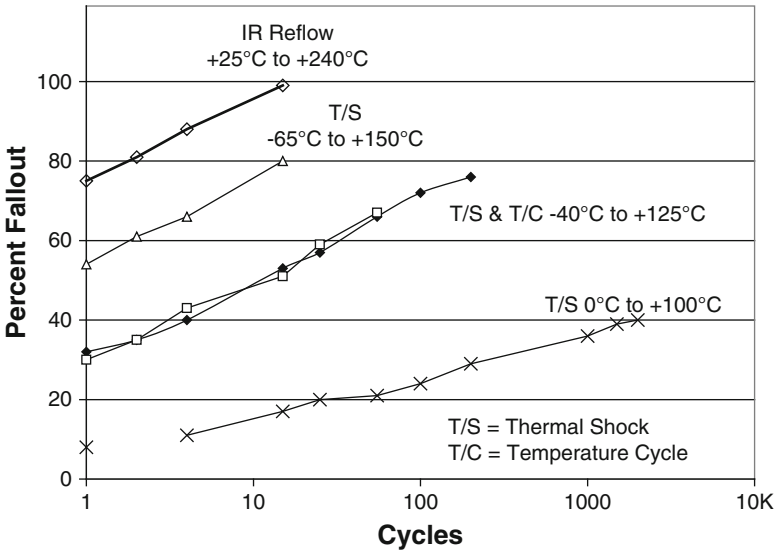



**Fig. 11.21** Thermal excursion accelerated test results for a maverick lot

thermal excursion testing, but fallout during high temperature lifetesting was nonexistent [39].

The thermal excursions are clearly an effective acceleration factor, such that for this very special sample of devices; one infrared reflow is roughly equivalent to 8 thermal shocks from −65 to +150°C or 144 thermal cycles from −40 to +125°C or 5 million cycles from 0 to 100°C! Armed with a solder reflow simulator, the root cause of the interconnect failure mechanism was found quickly, and containment was possible by lot screening with the optimum thermal excursion stress test. In this case, both field returns and reliability test failures provided an effective combination to focus attention for corrective action.

There have been several similar improvement projects, where none of the root cause failure mechanisms have been identified by conventional high temperature lifetesting. So far, the mechanisms have been discovered by either high volume use in the field or by using special accelerated tests on enhanced structures in the factory.

The importance of correlating mechanisms and their failure physics cannot be overstated. In most all historical data that is presented here, work to understand distributions and acceleration factors assumes that the majority of the failure population will fail for the one particular mechanism under study. As we move beyond the first three stages of reliability, the reliability methods can begin to comprehend more than a single failure mechanism which afflicts an entire population and is most successfully accelerated by temperature. In other words, we must move beyond measuring thermally accelerated lifetesting of transistors. We must move beyond simply determining reliability – to improving reliability. There is no

disputing that the inherent reliability of each technology is critical; however, the technology must be integrated into circuits, modules, and systems that can be useful for any particular application. If the first three stages of reliability are performed correctly, then the reliability of the application is likely to be much different than the fundamental limits of the transistor technology. Some would say that this is where the real world reliability begins.

With the methodology of the first three stages of reliability described, and the history of results revealed, we are now ready to move on the final two stages. Just as the first three stages have synergy and are intertwined, so are the remaining two stages. In some respects, built-in reliability is an outcome of the material-mechanism-physics steps, and the defects stage is merely a subset of the building-in process. In the real world, all five steps begin at the beginning, and they can each feed one another independently. Who would not want to be eliminating defects right from the beginning? Who would not want to build in reliability during research and development? Who would not want to understand the material-mechanism-physics aspect of how a defect is formed, what causes it to degrade performance, and when it should be expected to manifest?

## 11.5 Stage 4: Building-In Reliability

In the previous review of available published work on compound semiconductor reliability, there seemed to be no clear indications of reliability improvements over the past 35 years. Is reliability actually being "built-in," and if so, how is it done?

One of the interesting things about compound semiconductors is that there are few examples of legacy products which exist for more than a few years – especially in commercial and consumer applications. However, there are some processes that survive considerably longer than silicon processes that are constantly shrinking and evolving. For example, consider the traditional "high reliability" applications in defense and aerospace markets. These niches are especially averse to any changes, and as a result, reliability is predictably stagnant. Figure 11.22 demonstrates process reliability over a 14-year period.

The data in Fig. 11.22 is accumulated by testing individual transistors pulled from several wafer lots each week from 1997 through 2011. These transistors are tested at extremely high temperatures to evoke significant wearout-type degradation in a short period of time. Any samples degrading more than 10% in the first 15 h are unacceptable. Most samples exceed the criteria by 2× life, and some last more than 10× longer. While the distribution has become obviously tighter, there is no significant change in the average lifetime at 320°C for the entire 14-year history.

Figure 11.22 is a good example of the distribution of measured lifetimes in a production process. Figure 11.22 is the antithesis of building-in reliability. First of all, notice there is a distribution of lifetimes that exceeds an order of magnitude. This should not be unexpected, since characterization of lifetimes is part of the development and qualification activities that occur in the first three stages

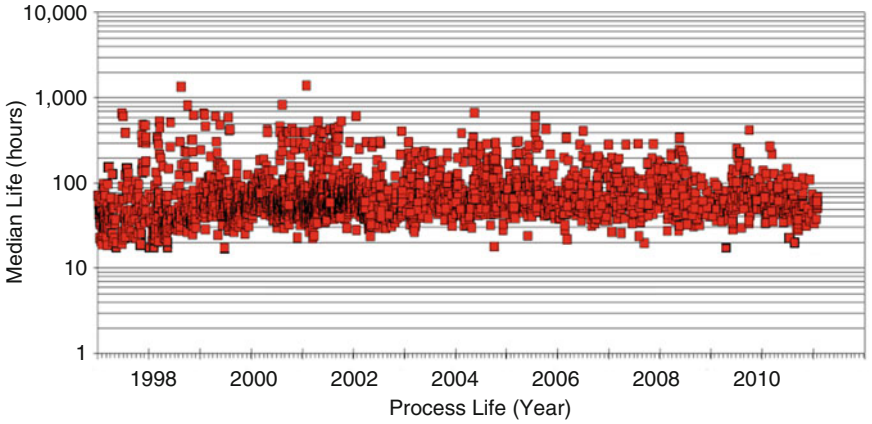



**Fig. 11.22** Median life versus lot sample from pHEMT nominal DC lifetest at 320°C. Failure criterion is 10% current degradation

of reliability. Overall, it is the general summation of "process maturity" that accounts for gradual improvement in the spread of the distribution. Typically, the maturity is achieved by establishing statistical process control and applying minor adjustments and alterations over an extended time period. Note that major changes of continuous improvement efforts are typically not allowed once a process is developed, characterized, and fixed (or "frozen") – particularly for high reliability applications.

For the most part, all of the "built-in" aspects of semiconductor processing occur during the engineering, development, and pilot production phases. Subsequent adjustments involve tuning that does not affect the form, fit, or function aspects which are under strict change control. As a result, most adjustments involve maintenance and control functions without effecting material or physical properties of the products – thus, there cannot be any change in reliability.

In reality, continuous improvement involves change. Without change, you are likely to do what you've always done and get what you always got. High-volume consumer markets have expectations that building-in reliability is continuous and does not end when the product is introduced.

If we compare technologies over time, we can see the effect of learning cycles in the reliability lifetime improvement. The generation-to-generation and era-to-era improvements in reliability and quality have been demonstrated over and over in the history of reported results. By adding the technology information to Fig. 11.19, we can indicate how each new compound semiconductor material might improve in Fig. 11.23. As technology improvements occurred for MESFETs, then HBTs, and then GaN, each one of these cycles began with relatively low reliability, followed by improvements in the reported lifetimes.

For specific cycles of cellular phone power amplifiers, the same improvement cycles were experienced, but not necessarily for the same type of reliability. Up to this point, the reliability focus has been on lifetimes. For commercial high-volume

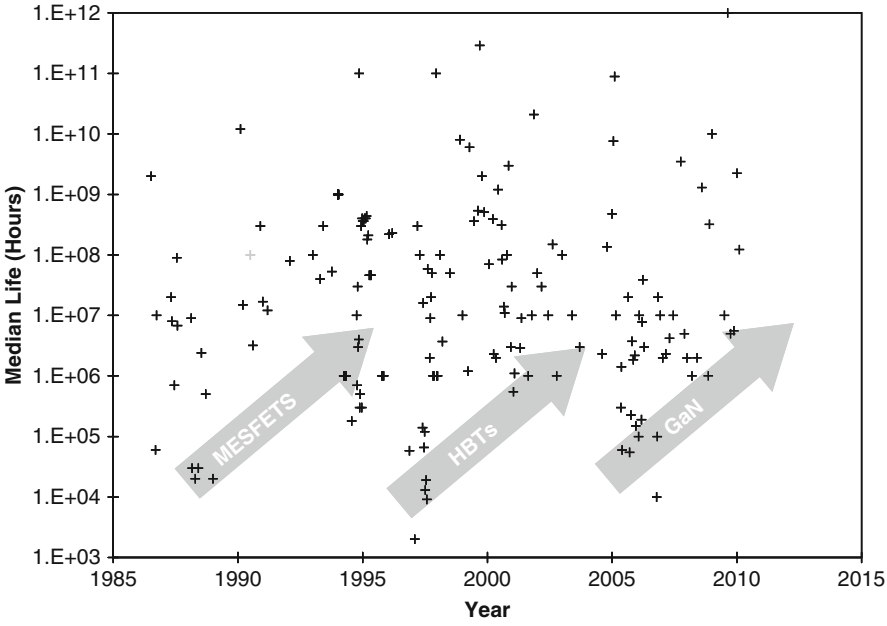



**Fig. 11.23** Summary of published reliability results for the past 30 years. New technologies will typical report lower expected lifetimes but improve with maturity

applications, there is a shift in reliability concern away from wearout median lifetimes and toward early life fallout and infant failure rates.

As an example, consider fallout in a cell phone factory. Expectations have changed two orders of magnitude over the past decade. Defects per million expectations in the factory dropped from around 1,000 in the late 1990s to less than 50 by 2010. Figure 11.24 shows that subsequent generations learn from previous generations, and the cycle time for improvement can be accelerated. However, when the technology changes significantly (or the materials change), the cycle time can be extended and the defectivity is typically set backward to effectively repeat previous cycles.

These cycles of learning are similar regardless of the technology. Figure 11.25 shows that even major mainstream silicon suppliers experience quality improvement resetting with each new development along the path of miniaturization. Each technology node can suffer an order of magnitude more fallout at the launch. Even though Figs. 11.24 and 11.25 have different units (fallout vs quality), the similarity in cycles is undeniable.

The previous three graphs all show that maturity gained within each generation has an impact on quality and reliability levels. However, the absolute level of reliability must meet the requirements for early life and wearout during use for each application.

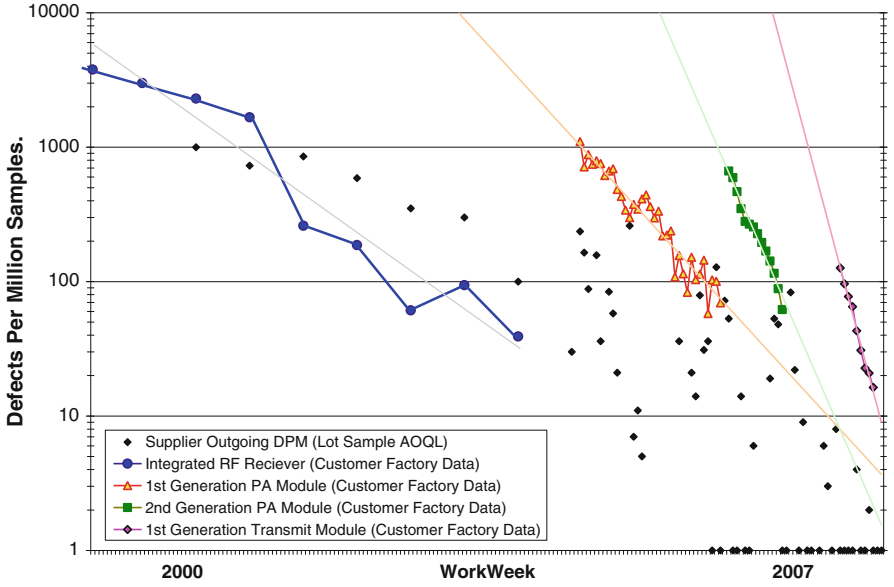



**Fig. 11.24** Example of reduction in fallout during product development

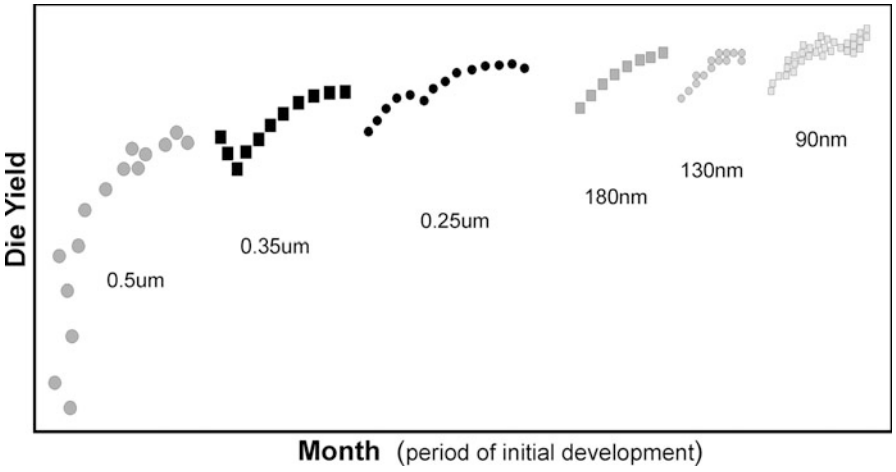



**Fig. 11.25** Example of quality improvement during progress of shrinking node sizes for silicon

One of the first considerations of stage 4, building-in reliability, is to determine the cause of unreliability (or unquality). One categorization of causes is to consider common cause versus special cause.

Common causes are all of the natural variation that exists in device fabrication, assembly, and testing. Special causes are unique events, incidents, or accidents which produce some result which is neither predictable nor assignable to the

**Table 11.9** Common and special causes of variation

| Common causes | Special causes |
|---|---|
| Normal fab tool degradation between maintenance events | Skipped maintenance |
| Normal tool variation | Machine malfunction |
| Raw material variation | Unapproved source of raw material |
| Variability in settings | Operator error |
| Machine-to-machine variation | Skipped step or doubled step |
| Normal process vacuum variation | Leak to atmosphere |
| Poor design | Power surge (ESD or EOS) |
| Within-family | Out-of-family |
| Present in all batches | Rare occurrence |
| Population straddles spec limit | Nonstandard rework to center population |

general population of devices. Some comparisons of common causes and special causes are shown in Table 11.9.

The reason for determining the type of cause is that the cure (or the mitigation) is quite different for each type. Reducing common causes is accomplished by reducing variation. Improvements in one area may affect other areas, and there is an overall synergy. Semiconductor processing will naturally mature by repetition. Process improvement experiments can pinpoint sources of variation and provide methods for reducing variation. Statistical process control provides ongoing measurement of process capability and measurement of inherent variability.

On the other hand, special causes may not have synergy with anything, either with common causes or with other special causes. Special causes can occur on mature processes which already exhibit very low common cause variation. While many special causes can be prevented by rigorous methods and disciplined processes, the best mitigation is to permanently eliminate the root cause and verify the elimination by improving the detection methods and by preventing recurrence.

One of the obvious and largest groupings of special causes is defects – but that is the subject of stage 5.

For modern semiconductor processes, detection is achieved by the use of escape points. Escape points are defined as steps in a manufacturing/fabrication process where fallout can be caught. Escape points are often measurements of physical attributes or electrical properties. The idea here is to consider not only the root cause of a problem but, equally importantly, what went wrong with the control system in allowing this problem to escape downstream. Reliability engineers must work with their process engineering teammates to identify and verify all of the applicable escape points. In semiconductor processing, the point is more precisely defined as the earliest control point in the fab process following the appearance of the root cause that should have detected the problem but failed to do so. The potential value in considering escape points is the redundancy of not only solving the root cause of problems but also in improving the detection of future problems. A robust problem-solving process would include selection of corrective action,

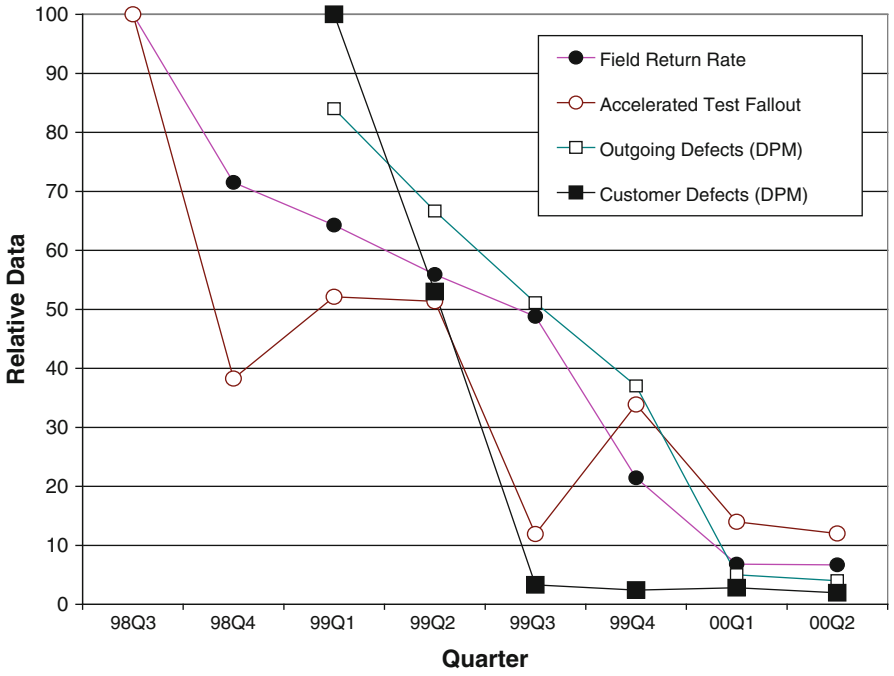



**Fig. 11.26** Volume, quality, and reliability relationships for 2.5 years

verification, implementation, and validation of both the method to eliminate the root cause but also the method to improve the effectiveness of the escape points.

Another important correlation that is important in reducing customer reliability failures is the relationship between outliers and mavericks. In many cases, outliers and maverick are even considered as the same thing. However, a maverick population is one that causes fallout in the final application and otherwise goes undetected during fabrication, assembly, and manufacturing. On the other hand, outliers are a population that can be within specification but are unusual compared to the prevailing population. The definitions of outliers and mavericks depend critically on what is being measured and where it is measured. The correlation is important since most of the process control monitor and component measurements are considerably different than the system measurements in the final application.

The implication of improving the detection of new failure mechanisms by enhancing escape points such that weak devices can be caught is that the mechanisms are converted from early life fallout to simple yield loss. Besides the effect of "building-in" reliability, finding the associative ties between quality and reliability is one of the improvement methodologies for reducing defects. Examples of general quality and reliability associations are shown for the past 2 years in Fig. 11.26 along with yield and volume. The (relatively) mild increase in yield and the approximate 3× increase in volume are not shown, but the quality and reliability measures are normalized to their initial measurements in the 2-year

demonstration period. It seems more than a coincidence that both internal and external measures of quality track so closely with more than 10× improvements during this time frame. It is also noteworthy that external return rates and internal test fallout have very similar trends to each other, as well as to the quality levels.

Figure 11.26 shows the clear trends of decreasing defects, returns, and fallout over a period of increasing production. It is arguable whether the volume and yield increases *caused* the quality/reliability improvements, but there is no doubt that the methodologies of improvement are related and, in fact, synergistic. At least for our company, nothing short of shipping a million parts per week could demonstrate what type of failures are important, and then subsequently identify better tests for accelerated reliability evaluations and predictions [40].

One of the most important aspects of building-in reliability is the reliance of the relationship, or correlation, of quality and reliability. Every mechanism and its associated degradation has some kind of onset. Recall that our definition of reliability is "quality-ability," meaning some measure of performance (quality) for a period of time. Time is a key essential in defining reliability, which in effect, does extend the quality requirement over a finite demand period. For some mechanisms, an initial quality measurement can foretell the ensuing degradation and even predict the ensuing reliability – for example, certain types of leakage within a transistor. On the other hand, some mechanisms give no electrical indication of ensuing reliability – such as in time-dependent dielectric breakdown in capacitors.

The reason that quality-reliability relationships are important is because the methods of building-in quality will usually result in built-in reliability. This quality-reliability relationship is often essential as continuous improvement progresses during the development of a particular semiconductor technology. As each dominant failure mechanism is mitigated, the next mechanism will be more difficult to detect. If volume is low and/or customer feedback is minimal, then an alternative source of data can be the quality measurements and an alternative source of failures can be the yield fallout. Just like customer use, quality and performance measurements are typically made for 100% of a population. In contrast, reliability measurements are typically collected on a sample of the population. The sample size dictates the resolution of reliability measures, so a 100% sample offers the best possible resolution. We can show relationships of quality to reliability by simply measuring fallout, as shown in Fig. 11.27.

If all of the fallout is caused by defects and the lifetime is shorter than expected, then this type of graph may start to look like a bathtub curve.

## 11.6 Stage 5: Defects

One of the most simplistic ways to model reliability is to consider a basic ink pen. The pen has a visual wearout meter when the outside case is transparent, because you can see how close you are getting to wearout by watching the ink reservoir inside the pen. Like most other consumable items, you have a good feeling about

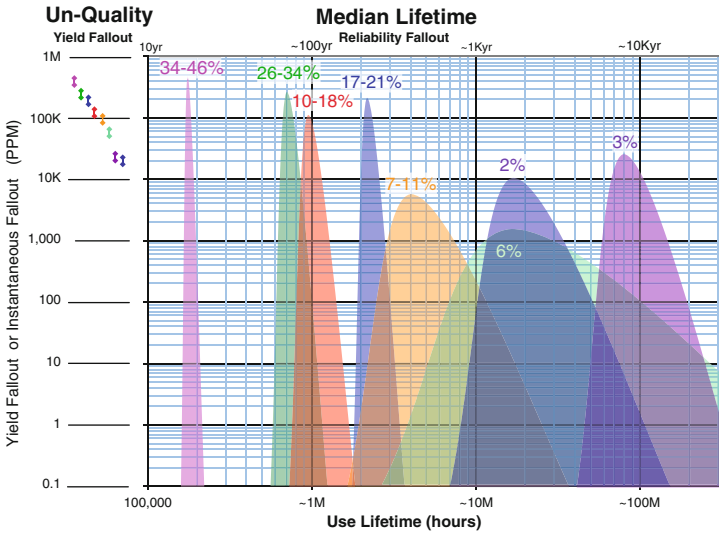



**Fig. 11.27** Quality and reliability relationships for yields ranging from 56 to 98%. Lowest fallout exhibits longest lifetimes, and highest fallout exhibits the shortest lifetimes

the reliability of a pen once you get the first sentence out of it. The first drop of ink is basically the early failure rate interval for our ink pen example. So unless you apply the pen incorrectly, such as using the pen as a can opener or a garden trowel, you can expect it work until you see the ink run out. Running out of ink is analogous to measuring end of life as we described in the reliability physics of stage 3. But how often do you actually run the pen completely out of ink? The reliability expectations for a pen can be summarized by something called a bathtub curve.

The original bathtub curve was modeled after human fatality rates [41]. As seen in Fig. 11.28, childbirth is a particularly life-threatening event. However, once born, infants experience a sharply decreasing fatality rate that levels off around adolescence. Teenage activities (particularly driving automobiles) signal the maturity phase of life where human fatality rates begin to increase.

As we previously discussed, a brand new pen might not always work properly, but once the ability to function is proven, there is some confidence the pen will perform reliably. The reliability of semiconductors has been historically modeled with a traditional "bathtub curve," as in Fig. 11.29 [42]. This curve maps the generic failure rate versus time. Because the failure rates are highest during the early life and at wearout, the curve suggests three regions of failure rate:

The first part of the bathtub curve is called "early." This has often been referred to as the "infant" region. This terminology obviously follows the origin of the human bathtub curve. The early life part of the curve is now part of our expectation of turning on an electrical device for the very first time. Once the new gizmo or gadget turns on successfully the first time, most consumers feel a sigh of relief and

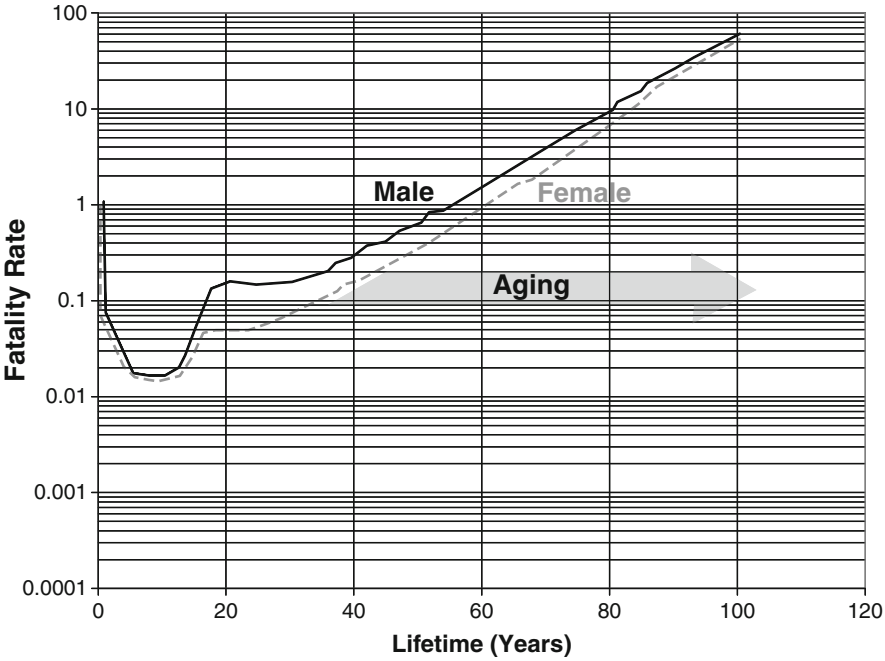



**Fig. 11.28** The original bathtub curve, showing human mortality rates (in percent) for males and females in the United States circa 1990

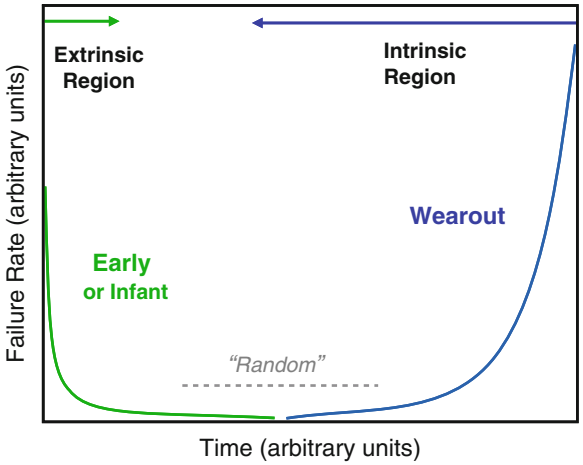



**Fig. 11.29** A generic bathtub curve for semiconductors. Failure rates are typically plotted on a log scale, and lifetimes are plotted linearly up to a 100 years or whatever is the life expectancy

move on to an expectation of reliable performance. The early failure rate is assumed to be caused by defects, and fallout rates are constantly decreasing throughout the lifetime of the device. For this discussion, we'll name this the "extrinsic" region. It is the region of specific interest for reliability stages 4 and 5, and it is also the region of most interest for customers since the early life is analogous to the warranty period for most applications of semiconductors.

The next region, in the middle, is called "random." It is the region attributed to a constant failure rate. No specific semiconductor failure mechanisms have been associated with this region, although "defects" have been blamed here as the underlying cause for "random events." Application anomalies, such as overstress or electrostatic damage, are likewise attributed to this region. Random failure mechanisms are the favorites of folks who must calculate fallout for complex electronic systems since constant failure rates are easy to add and accumulate estimates. Although some mechanical wearout mechanisms exhibit this type of "random" signature for machinery, we find no evidence that this constant failure rate region exists for normal operation of semiconductors, so we will not include a "flat bottom" on our bathtub and ignore the "random" region from here on.

The last region is called "wearout." This has been the focus of most semiconductor reliability studies, particularly during reliability stage 3. Failures are produced under extreme acceleration conditions – primarily driven by temperature. Instantaneous failure rates increase constantly throughout the useful life of the distribution. The rates may actually drop off after the population is depleted, but the entire population is expected to fail. In most experiments, wearout failure distributions are apparent under accelerated conditions. This becomes an important region to consider for new technologies utilizing new materials, but in general, wearout is not seen within the expected useful life of a semiconductor device or even within most human lifetimes. In a normal population of samples in a production environment, the majority of devices are expected to exhibit very long lifetimes and eventually degrade by wearout, so we can also call this the intrinsic region.

Each of the bathtub regions contains a summation of multiple failure mechanisms and their associated failure distributions that we described in reliability stage 2. For example, the wearout region would typically include diffusion, electromigration, thermal excursion fatigue, corrosion, or dielectric breakdown failure mechanisms. The extrinsic region might contain all the same mechanisms but under the influence of various defects. Detection of enough defects could produce sufficient failures to characterize a distribution – but only with sample sizes of many thousands. An empirical bathtub model for reliability would include only the extrinsic and wearout distributions, and it could be more accurately predicted by the empirical details added in Fig. 11.30.

Modern IC design practices utilize design layout rules which delay the occurrence of wearout – at least until some time that is much longer than the product's planned life (e.g., 100, 1,000, or even 10,000 years). The development of the appropriate layout rules, material qualities, or application stresses involves the extrapolated results of accelerated lifetesting (e.g., high temperature operating life – HTOL) to restrict the maximum operating temperature, environment, and

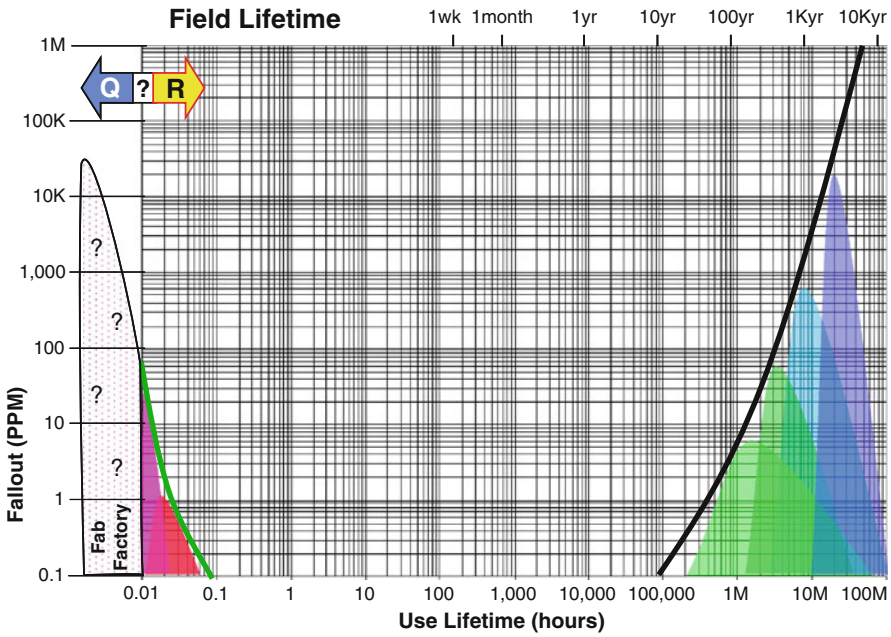



**Fig. 11.30** An empirical bathtub curve for semiconductors. Fallout is typically plotted on a log scale, and lifetimes are plotted logarithmically since actual wearout tends to exceed 100 years. Importance of early fallout depends on when deployment is started during the early life

electrical stress by applying operating restrictions. The thermal restrictions of plastic packages reduce the cumulative damage on nominal structures during unaccelerated use to a negligible level. Therefore, the actual failures of any well-designed semiconductor product would be determined solely by introduced defects. Thus, a simplified version of the empirical bathtub model would focus on the extrinsic region, or early life region, as identified in Fig. 11.29 and enhanced by Fig. 11.30.

Quality and reliability relationships can be investigated further by focusing on the early failure region of the bathtub curve. The definition of quality or reliability defects becomes rhetorical depending on the individual perspective. For example, does reliability life begin with the OEM or with the final consumer? This question is highlighted in Fig. 11.30 with the question marks in the area of the bathtub curve preceding the lifetime scale. In consideration of the actual data provided by customers, the expectation of increasing quality and reliability can be generalized for the early or infant region by significantly decreasing fallout. This expectation has been identified by a ratio named the K factor. A logical relationship between quality and reliability has been pondered for many years. Investigations into relationships between yield and reliability began with the first defect

characterizations by Stapper two decades ago [43, 44]. Later, a ratio of quality defects and reliability defects was proposed by Shirley [45] in the following form:

$$\boldsymbol{K} = \frac{\text{Reliability Defect Density}}{\text{Quality Defect Density}} \tag{11.1}$$

Shirley postulated that this ratio ($\boldsymbol{K}$) is a constant, and that each failure mechanism has its own ratio. He also suggested the ratio was at least 1/100. Estimates for this ratio have expanded to a region between 1/100 and 1/1,000 [46]. Simply put, for every 1,000 quality defects (yield at age zero), you should expect between 1 and 10 reliability defects. Using the $\boldsymbol{K}$ ratio, many studies have investigated and optimized burn-in.

The K factor is the ratio of defective parts that fallout at each escape point during fabrication, assembly, product test, OEM assembly, system test, during end user turn-on, early life, warranty period, useful life, and fallout. Historically, semiconductor fabricators have taken advantage of the decreasing early failure rate by a process called "burn-in." The intent of burn-in is to accelerate early failures using bias and temperature in order to evoke the defective fallout and move to a lower failure rate region of the bathtub curve. From another perspective, simply consider what part of the bathtub curve is likely to produce fallout that the customer would detect? The empirical curve shown in Fig. 11.30 has an early life peak fallout at about 3% during the fabrication process. This could be more simply described as 97% yield. Fallout drops by nearly a factor of 100 during the detection, screening, and protection phases of construction to produce a part with early life fallout less than 100 ppm. In this empirical example, both quality and reliability fallouts are measured on the same early life distribution – the delivery stages from supplier, to OEM, to end customer are all that separate the various definitions. As a result, the origin of the quality/reliability relationship becomes obvious. For silicon devices, the K factor of 100–1,000 is normal, but each situation and mechanism can be affected by the percentage of early failures present in any population and the associated detection methods that can be applied. As a result, K factors ranging from 0.5 to 1,000 are typical.

In contrast, consider that for more than two decades, the focus of GaAs and compound semiconductor reliability testing has been on high temperature lifetesting – which is geared to measure the wearout mechanisms which appear at the end of life – such as beyond 10-year lifetimes at the right side of Fig. 11.30. Several failure mechanisms are highly accelerated by temperature, so this methodology has produced data that is easy to analyze and straightforward to predict applicable lifetimes – albeit very long lifetimes. To the contrary, if any early failure mechanisms exist, those mechanisms tend to be quite different than what eventually falls at the end of a highly accelerated lifetest.

Other studies have shown similar relationships and correlations between yield and reliability results. For example, van der Pol et al. [47] showed correlations between intrinsic failure rates and early failure rates over the past 30 years. Compound semiconductors have similar results when comparing yield and reliability [40].

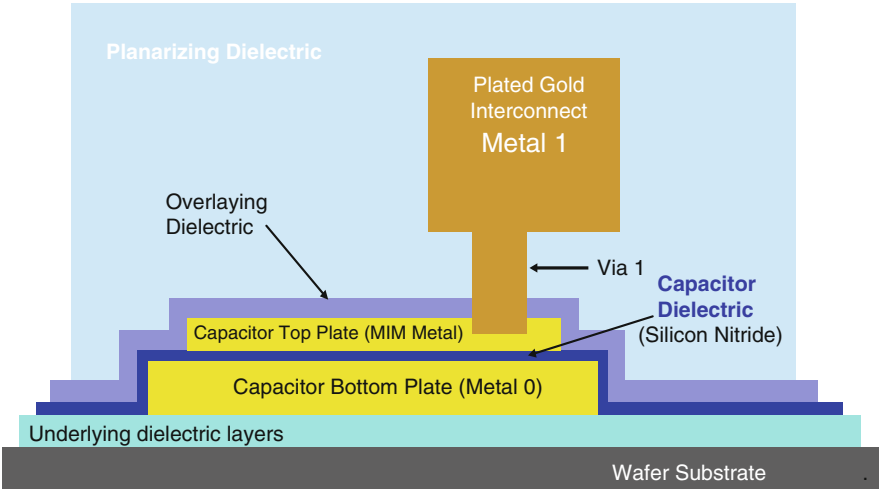



**Fig. 11.31** Cartoon drawing of MIM capacitor construction

By using the bathtub curve, we can explain the important shift of focus from wearout to defects. Defects are those rare artifacts that are present in every process but seldom discussed in reliability physics investigations. The important extrinsic reliability failures occur as defects that degrade to a complete failure under nominal aging and use conditions. The most prevalent defect types which fall into this category are:

I. Capacitor dielectric defects
II. Metal intralevel connection defects
III. Adjacent contact short defects
IV. Metal interconnect short defects
V. Contact or interconnect step coverage defects

Ten years prior to discovery of defects in compound semiconductors, similar defect types were reported for silicon. Shirley reported that metal defects and particles were the top cause of reliability defects, as determined by reliability testing and by a yield analysis [45].

Several examples of defect improvements are described by the following stories and seven figures. The improvements are referenced in terms of continuous improvement methodology and progress in overall reduction of defectives. The units of measure will be percent fallout, defects per million (DPMs), or parts per million (PPM) fallout. Most are measured by a statistically significant sample or measured in use by the customers. In some cases, the units are normalized to show relative improvements.

As one of the circuit elements most affected by defects, let's look first at capacitors. The construction of MIM capacitors is shown in Fig. 11.31. Several types of capacitors are available in various processes, where the construction and

thickness might be somewhat different. The common configurations are constructed from evaporated or sputtered metal plates and plasma-enhanced chemical vapor deposition (PECVD) silicon nitride dielectric.

During the formative stages of reliability, many thousands of capacitors were tested historically as individual components and within circuits without experiencing failure. Millions of capacitors were shipped into the field for more than 10 years in relatively low-volume applications. The first indication of capacitor failure mechanisms came from the field after 10 years of industrial and commercial volume shipment levels. Based upon inputs from customers, the failure rate was estimated to be approximately 200 defectives per million devices. At that rate, we would have expected a 50% chance of seeing three capacitor failures in the 15,000 devices that were lifetested between 1985 and 1995 – but we didn't experience any capacitor failures up to that time. Evidently, probability wasn't helping us with lower volumes. Based upon the initial field returns, an improvement project was undertaken. Instead of measuring capacitor reliability test failures (which had not yet been detected), particle density was measured as a new escape point during fabrication and utilized to track improvement. Over the course of a 2-year improvement project, the particle density had been reduced by almost two orders of magnitude – still without generating a single accelerated reliability test failure! It's important to reiterate that the improvement in Fig. 11.32 was initiated by the volume resolution achieved by crossing the million-unit warranty period "volume threshold." The improvement itself was a result of techniques in improving capacitor manufacturability and yield – this is the typical methodology of building-in reliability. Subsequently, the incidence of field returns for the particle mechanism was reduced to less than 1/20th of the original fallout rate.

Subsequently, the capacitor dielectric was improved, and the thickness was reduced by 1/4 so that voltage-accelerated tests became more and more effective and the challenges of thinner dielectrics came to the forefront. Additionally, special larger-area capacitors were added to investigate reliability and yield. These changes led to discoveries that were never apparent on actual products – even from larger samples used in the field. In effect, the combination of unique structures, effective stress factors, and yield improvements have carried capacitor reliability performance well beyond the next volume observation threshold.

All capacitors, even perfect ones, have a finite lifetime. The dielectric (silicon nitride) lifetime for MIM capacitors can be characterized as the overall "charge to breakdown." Although intrinsic defects within the dielectric can influence a capacitor's lifetime, the extrinsic defects, which result in changes to the dielectric thickness, are the primary cause for early life failures. In fabrication, the extrinsic defects are typically caused by stray particles from all process steps that precede the formation of the capacitor dielectric.

The best method of characterizing capacitors is to ramp voltage on individual capacitors for an entire population across an entire wafer. Unfortunately, a voltage ramp is destructive. Nevertheless, the results of voltage ramping is a powerful investigative method for quality and reliability analyses – see Fig. 11.33.

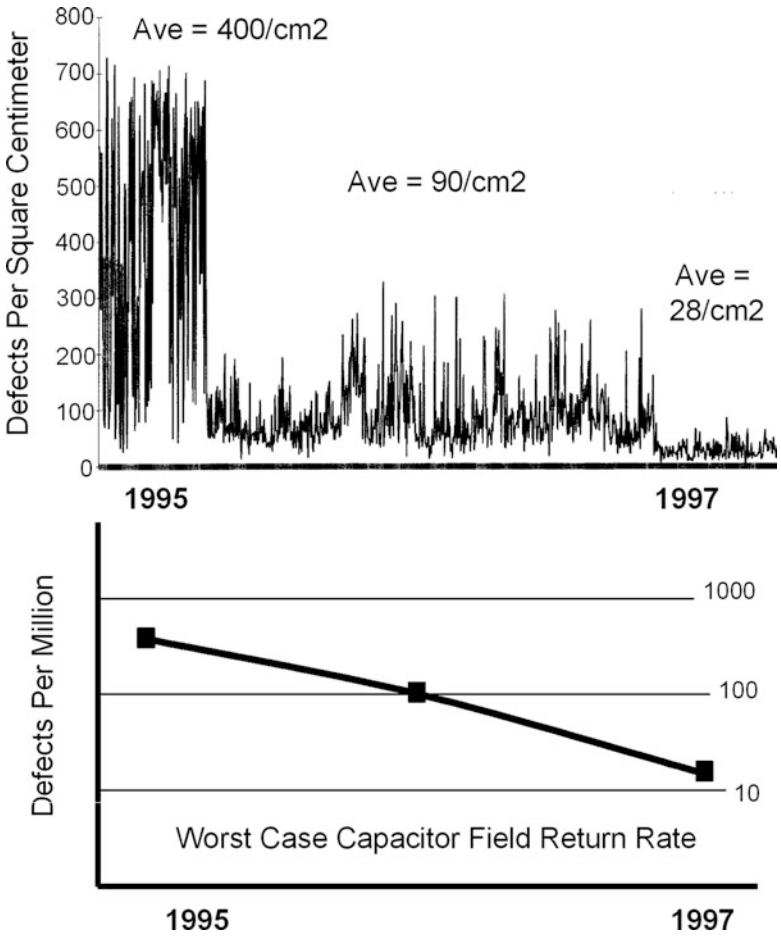



**Fig. 11.32** Number of defects and resulting failure rates of capacitors

The ramp results shown in Fig. 11.33 are for a population of 110 very large (>1 sq mm) area capacitors that are nominally 500 Å thick. In this population of large capacitors, approximately 15% of the capacitors exhibited defects (shown in Fig. 11.33) causing lower than expected breakdown. The remaining 85% of the population had breakdown voltages between 50 and 56 V – which is an indication of the nominal (or intrinsic) variation in capacitor thickness across the wafer.

An important consideration for conducting ramp to failure testing is the size of the test capacitors. Larger-area capacitors are desirable to study extrinsic defects because larger areas are more likely to include these defects. On the contrary, small capacitors are more useful to look at intrinsic properties of capacitors.

Wafer-level testing techniques offer a rapid and easy means for evaluating the overall quality and reliability aspects of capacitor populations. Comparisons from one capacitor type to another are easy, once a baseline expectation is established.

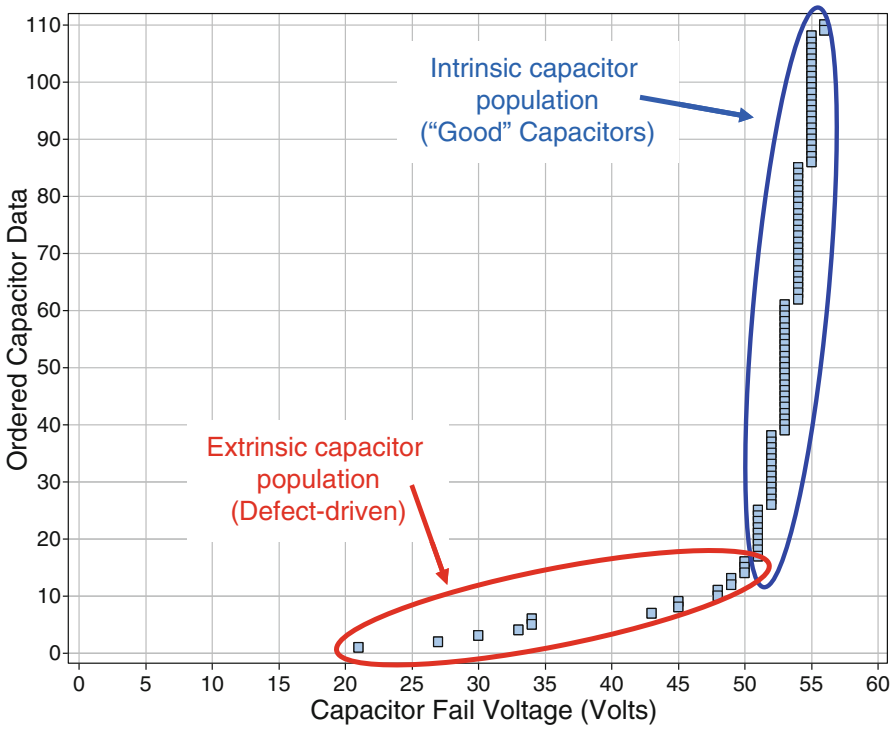



**Fig. 11.33** Example results of ramping capacitors to failure

For example, various iterations of a process in development can be evaluated and compared. This quick feedback loop can be a successful component of process improvement.

The reduction of capacitor defects requires continuous improvement across various engineering disciplines. For example, the reduction of susceptibility to defects by design engineering, the elimination of defect types by process engineering, and the screening out of defects by product engineering. If the improvement efforts are applied together, then a synergy of effects and interactions of variables can multiply the results.

For example, the design aspects of capacitor layout and capacitor connection can have an influence on capacitor defects. For a connection example, consider that the top plate and bottom plates of all MIM capacitors must be connected. In some cases, the size and manner of the connections can produce different quality levels or introduce significant defects. By using the connection information, designers can increase yield and reduce the capacitor susceptibility to defects by adjusting the area and type of connections to the metal that forms the top plate and bottom plate of capacitors. The highest yielding connection would be sought, and for reliability, a redundancy connection might also be recommended – so, two optimized connections are suggested.

In addition to connections to the capacitors, designers should be advised if capacitor defects are found in specific locations other than a uniform random distributions. For example, a typical layout interaction with capacitor defects is the overlap of bottom and top plates. A lift-off metallization cross section is traditionally trapezoidal, not rectangular. Most lift-off metal edges include winglike features at the end of a gradual tapered edge. As a result, the proximity of the bottom plate of a capacitor must be carefully extended beyond the upper layer so that edge features do not interfere with the formation of a smooth, uniform, capacitor dielectric. The minimum requirement for the separation of the edges must be determined and maintained.

Designers can significantly enhance capacitor screening capability by providing utility connections on capacitors that are buried within circuits or are otherwise inaccessible from external connections to a die or on a finished device. Consider that typical circuit layouts have some capacitors that can be individually accessed and stressed with voltage and some capacitors are buried or placed in parallel with sensitive elements, so they cannot be accessed. Those capacitors within sensitive circuits and those capacitors lacking accessibility points might be reachable with the addition of utility probe pads that are contacted only during screening. Obviously, designers can help by adding the ability to screen all capacitors.

The most obvious improvement for process engineers to improve capacitor reliability can be by defect reduction. Every process step that precedes the formation of a MIM capacitor is an opportunity to form an extrinsic defect that could reduce a capacitor's lifetime. Metal patterning processes for transistor contacts, resistors, and metal interconnect features are the primary sources of defects found under capacitors. Additionally, photolithography processes, etching steps, and cleaning operations are opportunities to leave residues and particles that can reduce capacitor thickness. Refinements within each of these processes and adjustments to equipment and maintenance procedures can affect the generation and removal of defects. Consequently, interactions of processes, procedures, and maintenance functions can produce uniquely clean or exceptionally dirty results. So, the investigative and experimental process of defect reduction must include consideration of many fabrication aspects.

Process, test, and product engineers have at least two opportunities to improve screening methods to increase capacitor reliability: visual inspection and electrical stress.

MIM capacitor improvements are achievable by a successful combination of effort in three broad areas: (A) design and layout, (B) semiconductor fabrication processing, and (C) screening. Individual choices in each area can significantly reduce defects, but all efforts have a cost-benefit trade-off that makes an exclusive approach impractical or economically prohibitive. Fortunately, efforts in all areas simultaneously can be synergistic and effective.

The next improvement is another example of how defects can affect other elements besides capacitors. This is also a classic example of how removal of defects helps in more ways than expected. For a particular power amplifier design, compliance to requirements on some of the noise parameters was causing excessive fallout.

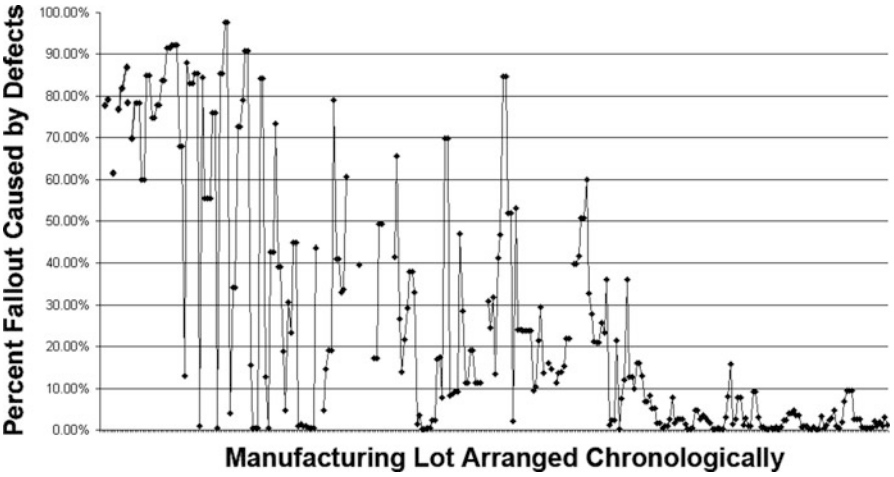



**Fig. 11.34** Reduction in percent fallout per lot for photolithography improvement for a lift-off anomaly

A dramatic improvement was obtained by eliminating the top cause of physical defects. At the time of this particular project, the defect was a type of lift-off anomaly left behind after the formation of the gate metallization. An adjustment in photolithography resulted in the improvement shown in Fig. 11.34. This is a typical result for physical defect reductions. However, the quality improved by *more* than the fallout amount caused by the particular defect that was removed.

Following capacitors and other types of physical defects, another leading cause of failures in sample testing, field returns, and reliability aging is electrostatic discharge (ESD). ESD can be mitigated in several ways. First, the sources of ESD can be reduced by well-known countermeasures such as use of wrist straps, heel straps, conductive smocks, ESD-safe workstations, shielded bags, air ionizers, humidifiers, and others. Alternately, the device design can be modified to reduce susceptibility. The following two graphs show examples of each type of improvement. Figure 11.35 shows how the relative number of DPMs is observed through duplicate assembly lines where some of the countermeasures have been inadvertently disengaged at a high-volume manufacturer. Figure 11.36 shows how a design modification to improve ESD robustness can reduce the DPM levels at a factory.

The improvement shown in Fig. 11.37 is the result of a specific team effort to install ESD countermeasures on a specific part handler. ESD was known as the leading cause of DPMs during final test. The team performed an audit of the test area and found the handler was a source of ESD fields. In an effort to improve the handler, metal parts were fabricated to replace some of the nonconductive parts that came in contact with the sensitive devices during testing. The results of this simple audit and upgrade are shown in Fig. 11.37.

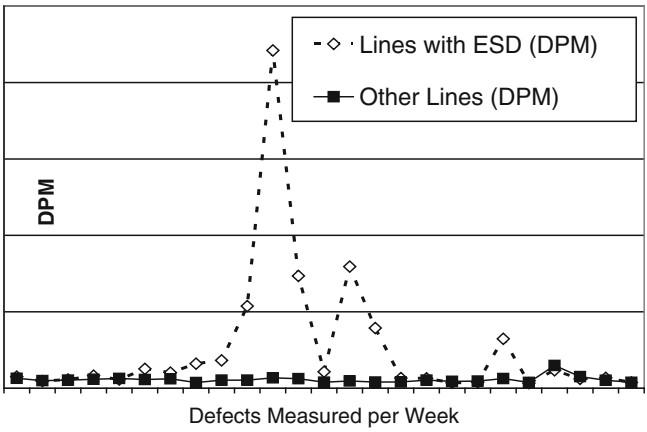



**Fig. 11.35** Effect of ESD countermeasures at a factory

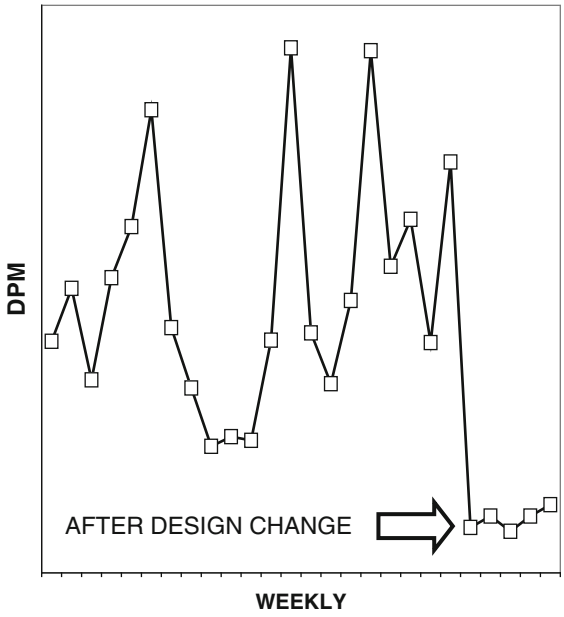



**Fig. 11.36** Effect of a design change on DPMs caused by ESD (measured at an OEM factory)

These examples represent projects resulting in the reduction in defects [48]. Quality and reliability are directly impacted by defects and certain yield loss, and fallout mechanisms are correlated to defects detected during manufacturing. In reality, these examples were just a small representation of more than 120 improvement teams operating over a 24-month time frame. Although many of the

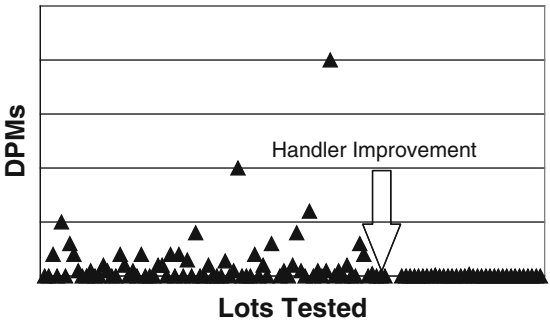



**Fig. 11.37** Lot-by-lot DPMs measured before and after countermeasure improvement to a handler

teams were focused on general improvements other than the reduction of DPMs, the work of many more teams than the few described here resulted in overall factory DPM improvement of three orders of magnitude over a 2-year time span.

It isn't easy to learn about defects from reading examples in a book. Written accounts don't often provide the insight of an actual learning experience of living through one of those scenarios mentioned in the cited examples. The experiences described previously show that high temperature lifetests usually aren't the most effective tests to determine actual reliability – particularly for early life and for typical customer applications and especially in the building-in reliability and defect investigation stages of reliability study. Two alternatives are offered: (1) work closely with customers who are using the highest volume of your parts or (2) evaluate statistically significant sample sizes periodically through tests involving different forms of acceleration – not just temperature. Note that there is not a particular stress that will work best all the time. The variability of effectiveness is under constant change, especially while continuous improvement reshapes the ranking of important mechanisms. For example, high temperature operating lifetests (HTOL) may not have been very effective in measuring reliability over periods of several years at a time, but HTOL may become the most effective stress for the next design, package, application, or generation of devices, so we still run HTOL stresses. However, the difference now is that we don't care much about finding the onset of wearout. It is far more important to find the reliability limiting mechanisms that occur during early life.

High temperature lifetesting is obviously *not* the only reliability test. Other accelerated tests are likely to be more effective in generating failures prior to wearout – especially failures that are likely to occur in typical use conditions. Ever-increasing thermal acceleration and outrageous extrapolations of lifetimes rarely correspond to what concerns customers under typical use conditions. The work of reliability engineering and defect reduction stages rely on volume and quality aspects of semiconductor manufacturing to achieve improved reliability results. Communication with customers and utilization of larger sample sizes are offered as methods for finding "real" failure mechanisms so that they can be mitigated.

## 11.7 Summary

So now that some of the history has been reviewed, what should be done next? For 35 years, reliability studies have touted the technology differences as a defense for the compound semiconductor niche. But why not use the reliability differences as compelling advantages in reliability? Because of the unique mechanisms and absence of the silicon problems brought on by relentless adherence to Moore's Law, compound semiconductors have an opportunity to use diversity of materials to optimize the bathtub curve.

In spite of our silicon and compound semiconductors mismatched sizes, it is not so wild a notion that the compound semiconductor technologies could actually prevail in terms of a reliability comparison. Even with repetitive reliability eras and generations under their belt, the silicon folks have a number of new issues of their own yet to solve as each generation shrinks. The continuous pursuit of miniaturization has contributed to a recurring cycle of reliability challenges for the silicon technologies. Materials for the latest silicon nodes are new once more, and the engineers have compressed the stages of reliability and are developing technology nodes in parallel.

There are several indications that reliability is under direct attack by Moore's Law. Consider the recent paradigm shift in reliability expectations:

> . . .aging has traditionally been the bailiwick of engineers who guarantee the transistor will operate for 10 years or so . . . But as transistors are scaled down further and operated with thinner voltage margins, it's becoming harder to make those guarantees. . . transistor aging is emerging as a circuit designer's problem.
>
> Samuel K. Moore [49]

While data summarized here shows initial improvements during compound semiconductor development for each new material, there are reports that silicon reliability is heading the other direction. Some have speculated that the peak of mainstream silicon reliability occurred around 1995, and that failure rates have been increasing at about 2 FIT per year since then [50]. Figures 11.38 and 11.39 suggest that the increasing failure rates are attributable directly to shrinking technology modes for silicon [51]. If true, this means the right side of the bathtub is shrinking with the silicon dimensions.

If reliability is to be measured by the silicon yardstick, then there are two areas of focus with several important tasks ahead:

Formalize the compound semiconductor knowledge from the initial three reliability eras. Identify, define, and characterize the failure mechanisms for all materials in use – not just the predominant mechanism that degrade under high temperature acceleration. It's also important to match up the mechanisms with appropriate degradation distributions and acceleration factors. For example, this implies that high temperature stress shouldn't be overutilized when current density matters! The rigors of reliability physics are necessary for all mechanisms, not just when high temperature acceleration is employed.

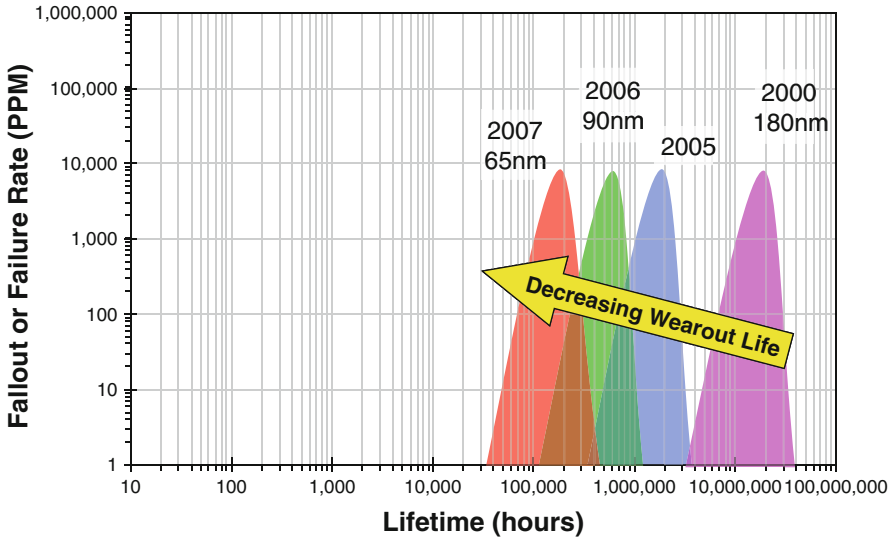



**Fig. 11.38** Perceived reductions in wearout lifetimes for silicon's shrinking features. Trading reliability for feature size in silicon-based integrated circuits

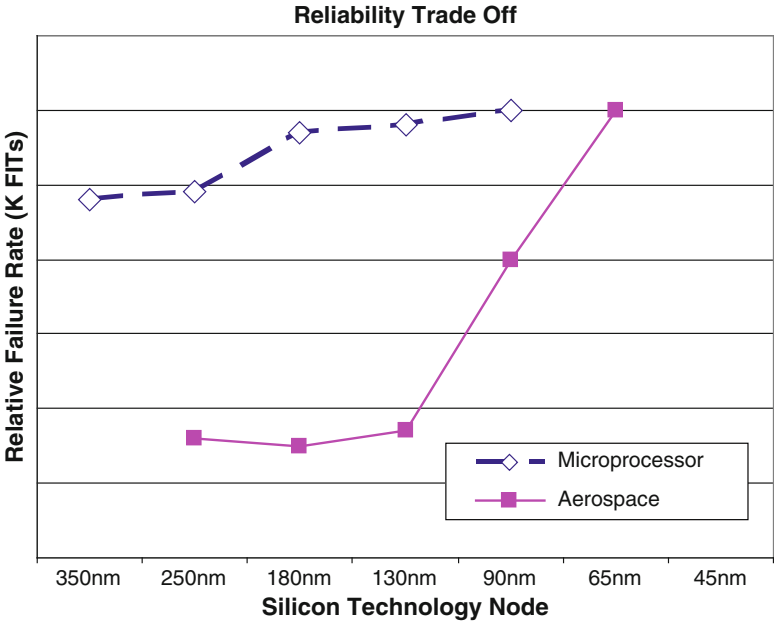



**Fig. 11.39** Trading reliability for feature size in silicon-based integrated circuits

Embrace the tactics of reliability stages 4 and 5 quickly and openly. Stage 4 means more effort on building-in reliability. Compound semiconductor folks need to showcase progress on process control and the relentless search for mavericks, rouge lots, and outliers. There must be continued development of fast, on wafer, test structures to speed reliability advancements and new material implementations. It will be important to apply knowledge of telltale failure signatures and customer returns with the use of in-line screening. Biasing methods of overvoltage, overcurrent, and overpower (RF) are expected, while quiescent and leakage current monitoring show promise. The defect reduction efforts of era 5 offer new territory to apply innovative tactics such as defect amplification, use of new environments, and/or pioneering accelerations to advance reliability knowledge ahead of the material and miniaturization technology drivers.

Customers are familiar with the silicon generation of reliability progress accomplished by the past 40 years of continuous improvement. Customers also have experience with problems of new technology and a perception that compound semiconductor reliability lagged the main stream. As feature sizes continue to decrease, resulting reliability impacts will loom larger for condensed silicon technologies, but the new GaN materials present similar material challenges for the compound semiconductor technology.

Reliability engineers need to be innovative. Armed with history and a goal of continuous reliability improvement, there are a few obvious approaches for compound devices which are different from mainstream silicon. As the overall reliability of a particular material improves with maturity and volume learnings, measuring degradation becomes elusive. If the compound semiconductor folks intend to build experience that is familiar to the customers, they should proceed with tactics of building-in reliability and efforts to reduce defects. An illustration might be a more visible use of yield-to-reliability correlations as an example to predict failure rates. If yield fallout also disappears, we can use new tricks such as physical amplification of defects in order to extend our predictive capability [52]. These tricks use the measurement of quality as a function of voltage and amplified (reduced) spacing on test structures. Other recent compound semiconductor breakthroughs with innovative stress such as bubble tests [53] and power cycling [54] are the kind of creative tactics that can take advantage of unique mechanisms and keep compound semiconductor technology on continually improving path.

This chapter has gone beyond the traditional look at reliability semiconductors by applying high temperatures. During a review of the five main stages of reliability, we have shown the first three stages are separated as those traditional tests which are particularly applicable for process development and material selection phases. While all five steps are necessary in reliability studies, the methods of establishing and improving reliability are different. The reliability stages to build-in reliability and eliminate defects are the focus of real-life measurements and efforts to ensure customers enjoy reliability by experiencing the absence of failures.

## References


1. G. Moore, Cramming more components onto integrated circuits, Electron. Mag., **38**(8), 19 (Apr 1965)
2. H. Stork, Reliability challenges for sub-100nm system on chip technologies, keynote address, in *International Reliability Physics Symposium*, Phoenix, Arizona *(IRPS)*, 2004
3. E. Deming, in *Out of the Crisis*, Massachusetts Institute of Technology, Center for Advanced Engineering Study, 1986
4. W. Roesch, T. Rubalcava, C. Hanson, Lifetesting GaAs MMICs under RF stimulus. IEEE Trans. Microw. Theory. Tech. **40**(12), 2452–2460 (1992)
5. Y.C. Chou et al., On the investigation of gate metal interdiffusion in GaAs HEMTs, in *GaAs IC Symposium*, San Diego, California pp. 63–66, 2003
6. W. Roesch, D. Stunkard, Proving GaAs reliability with IC element testing, in *US Conference on GaAs Manufacturing Technology*, Nashville, Tennessee (Mantech) pp. 91–95, 1988
7. H.C. Casey, G.L. Pearson, Diffusion in semiconductors, in *Point Defects in Solids*, vol. Z (Plenum, London, 1975)
8. H.Cholan, D. Stunkard, T. Rubalcava, Long-term effects of sidegating on GaAs MESFETs, in *IRPS*, San Diego, California pp. 49–53, 1992
9. A. Watanabe et al., Rapid degradation of WSi self-aligned gate GaAs MESFET by hot carrier Effect, in *IRPS*, San Diego, California pp. 127–130, 1992
10. P. Asbeck, C. Lee, M. Chang, Piezoelectric effects on GaAs FETs and their role in orientation-dependent device characteristics. IEEE Trans. Electron Devices **ED-31**, 1377–1380 (1984)
11. W. Camp et al., Hydrogen effects on reliability of GaAs MMICs, in *IEEE GaAs IC Symposium*, San Diego, California pp. 203–206, 1989
12. S. Mertens, J. del Alamo, A model for hydrogen-induced piezoelectric effect in InP HEMTs and GaAs PHEMTs. IEEE Trans. Electron Devices **49**(11), 1849–1855 ( 2002)
13. B. Surridge et al., Accelerated reliability testing of InGaP/GaAs HBTs, in *GaAs Mantech Conference*, San Diego, California pp. 37–41, 2002
14. T. Henderson, A physics-based model for excess leakage current in GaAs/AlGaAs HBTs after bias stress, in *GaAs REL Workshop*, Orlando, Florida pp. 2–10, 1996
15. J. Dunn et al., Trends in silicon germanium BiCMOS integration and reliability, in *IRPS*, San Jose, California pp. 237–242, 2000
16. W. Roesch, R. Winters, T. Rubalcava, B. Ingle, Humidity resistance of GaAs ICs, in *IEEE GaAs IC Symposium*, Philadelphia, PA pp. 251–254, 1994
17. P. Ersland, H. Jen, X. Yang, Lifetime acceleration model for HAST tests on a pHEMT process, in *GaAs REL Workshop*, San Diego pp. 3–6, 2003
18. T. Hisaka, et al., Degradation mechanisms of GaAs pHEMTs in high humidity conditions, in *Reliability of Compound Semiconductors (ROCS) Workshop*, Monterey, California pp. 81–88, 2004
19. A. Poe, S. Brockett, T. Rubalcava, Failure analysis of CDM ESD damage in a GaAs RFIC, in *International Symposium for Testing and Failure Analysis*, Bellevue, WA, Nov 2000
20. C.S. Whitman, T. Gilbert, A. Rahn, J.A. Antonell, Determining factors affecting ESD failure voltage using DOE, in *ROCS Workshop*, Palm Springs, California 2005
21. J. Scarpulla, E. Ahlers, D. Eng, D. Leung, S. Olsen, C. Wu, Dielectric breakdown, defects, and reliability in SiN MIMCAPs, in *GaAs REL Workshop*, Atlanta, Georgia pp. 92–105, 1998
22. H. Cramer, J. Oliver, G. Dix, MMIC capacitor dielectric reliability, in *GaAs REL Workshop*, Atlanta, Georgia pp. 46–51, 1998
23. B. Yeats, Assessing the reliability of silicon nitride capacitors in a GaAs IC process. IEEE Trans. Electron Devices **45**(4), 939–946 (1998)
24. M. Brophy, A. Torrejon, S. Peterson, K. Avala, L. Liu, MIM's the word – capacitors for fun and profit, in *GaAs Mantech*, San Diego, Ca pp. 57–59, 2003
25. W. Roesch, D. Hamada, Studying yield and reliability relationships for metal defects, in *ROCS Workshop*, Monterey, California pp. 121–133, 2004

26. W. Roesch, Reliability of compound semiconductor historical review, in *Reliability of Compound Semiconductors Workshop*, Palm Springs, CA, 30 Oct 2005, pp. 1–16
27. W. Roesch, S. Brockett, Natural failure mechanisms: analysis of actual field returns, in *Reliability of Compound Semiconductors Workshop*, San Antonio, TX, 12 Nov 2006, pp. 55–71
28. N-T. Shen, Perfect quality for free?!?, in *Compound Semiconductor Mantech Conference*, Vancouver, BC, Canada, 26 Apr 2006, pp. 75–78
29. W. Roesch, Volume & quality impacts on reliability: a new game for GaAs, in *GaAs REL Workshop*, Seattle, WA, 5 Nov 2000, pp. 137–145
30. W. Roesch, S. Brockett, Field returns, a source of natural failure mechanisms. Microelectron. Reliab. **47**, 1156–1165 (2007)
31. W. Roesch, A.L. Rubalcava, R.A. Winters. GaAs IC reliability returns: a story of abuse, in *GaAs REL Workshop*, Miami Beach, FL, 4 Oct 1992, pp. 30–34
32. W. Roesch, Outstanding issues in compound semiconductor reliability, in *International Conference on Compound Semiconductor Manufacturing Technology*, New Orleans, LA, 13 Apr 2005, pp. 49–52
33. W. Roesch, The ROCS workshop and 25 years of compound semiconductor reliability. Microelectron. Reliab. **51**(2), 188–194 (2011)
34. W. Masterton, E.J. Slowinski, *Chemical Principles Textbook* (W.B. Sanders Co, Philadelphia, 1977), pp. 388–389
35. W. Roesch, T. Rubalcava, C. Hanson, Lifetesting GaAs MMICs under RF stimulus. IEEE Trans Microw Theory Techn **40**(12) Albuquerque, NM MTT(1992) pp. 2452 – 2460
36. C. Canali, F. Castaldo, D. Fantini, L. Ogliari, L. Umena, E. Zanoni, Gate metallization "sinking" into the active channel in Ti/W/Au metallized power MESFETs. IEEE Electron Device Lett. **EDL-7**(3) (1986)
37. W. Roesch, D. Stunkard, Proving GaAs reliability with IC element testing, in *GaAs MANTECH Conference*, Nashville, TN, Nov 1988
38. J. Black, Electromigration – a brief survey and some recent results. IEEE Trans. Electron Devices **ED-16** (1969) pp. 338-347
39. W. Roesch, Thermal excursion accelerating factors, in *1999 GaAs REL Workshop*, Monterey, CA, 17 Oct 1999
40. W. Roesch, Volume impacts on GaAs reliability improvements. Microelectron. Reliab. **41**(8), 1123–1127 (2001)
41. Source = Age-specific death rates from the Human Mortality Database (HMD) http://www.mortality.org
42. W. Roesch, Reliability basics, in *International Conference on Compound Semiconductor Manufacturing Technology Workshop*, Atlanta, GA, 1999, pp. 1–60
43. C.H. Stapper, Modeling of integrated circuit defect sensitivities. IBM J. Res. Dev. **27**, 549–557 (1983)
44. C.H. Stapper, F.M. Armstrong, K. Saji, Integrated circuit yield statistics. Proc. IEEE **71**(4), 453 (1983)
45. C.G. Shirley, A defect model of reliability, IRPS Tutorial #3, Las Vegas, Nevada 1995
46. K.R. Forbes, P. Schani, Characterization of the time-dependent reliability as a function of yield for a 130 nm SRAM device and application to optimize production burn-in, in *IRPS*, Phoenix, Arizona 2004, pp. 165–170
47. J.A. van der Pol, E.R. Ooms, T. van 't Hof, F.G. Kuper, Impact of screening latent defects at electrical test on the yield-reliability relation and application to burn-in elimination, in *IRPS*, Reno, Nevada 1998, pp. 370–377
48. W. Roesch, Getting to zero – methods of reducing defects, in *GaAs IC Reliability Workshop*, Baltimore MD 2001, pp. 37–46
49. Samuel K. Moore, An odometer for silicon chips, IEEE Spectrum, June 2009
50. C. Hillman, Does silicon wearout: an OEM's perspective, in *IEEE SCV reliability*, Cupertino, California, 28 Oct 2009

51. E. Wyrwas, L. Condra, A. Hava, Accurate quantitative physics-of-failure approach to integrated circuit reliability, in *IPC APEX Expo*, Las Vegas, NV, 13 Apr 2011
52. W. Roesch, D. Hamada, Measuring liftoff quality and reliability with special test structures, in *International Conference on Compound Semiconductor Manufacturing Technology*, Wheeling, IL, 14–17 Apr 2008, pp. 199–202
53. W. Roesch, S. Peterson, A. Poe, S. Brockett, S. Mahon, J. Bruckner, Assessing circuit hermeticity by electrolysis, in *GaAs Mantech Conference*, Washington DC, 1 May 2000, pp. 121–124
54. W. Roesch, S. Jittinorasett, Investigating thermal excursion failure mechanisms for flip chip, in *GaAs IC Reliability Workshop*, San Diego, CA, 9 Nov 2003, pp. 107–116

# Chapter 12
# Strain Effects in AlGaN/GaN HEMTs

**Min Chu, Andrew D. Koehler, Amit Gupta, Srivatsan Parthasarathy, Mehmet Onur Baykan, Scott E. Thompson, and Toshikazu Nishida**

**Abstract** Since stress is a major factor in the operation, performance, and reliability in AlGaN/GaN HEMT devices, a thorough understanding of the impact of stress on performance and reliability can lead to improvements in device design. Mechanical wafer bending is a cost-effective method to investigate the effects of stress on semiconductor devices which has been extensively used to isolate and study the effect of stress in strain-engineered Si MOSFETs. In this chapter, a systematic study of the effects of externally applied mechanical stress on the AlGaN/GaN HEMT channel resistance and gate current is presented to provide insights into the physical mechanisms responsible for stress-related performance and reliability issues.

## 12.1 Introduction[1]

AlGaN/GaN high-electron-mobility transistors (HEMTs) stand out with superb advantages for high-power, high-temperature, and high-frequency applications. Unique advantages associated with GaN make AlGaN/GaN HEMTs desirable for high-speed, high-performance applications. Table 12.1 summarizes key material parameters for high-power, high-performance devices, displaying benefits of GaN compared to other relevant semiconductors. Although the effective mass ($m^*$) of GaN is larger than GaAs and InP, resulting in lower bulk effective low-field

[1] A. Koehler, *Impact of Mechanical Stress on AlGaN/GaN HEMT Performance: Channel Resistance and Gate Current*. University of Florida, 2011. Ann Arbor: UMI, 2011. AAT 3191701. Print. M. Chu,*Characterization and Modeling of Strained Si FET and GaN HEMT Devices*. University of Florida, 2011. Ann Arbor: UMI, 2011. AAT 3191701. Print.

M. Chu • A.D. Koehler • A. Gupta • S. Parthasarathy • M.O. Baykan • S.E. Thompson • T. Nishida (✉)
Interdisciplinary Microsystems Group, Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611-6200, USA
e-mail: nishida@ufl.edu

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_12,
© Springer Science+Business Media New York 2013

**Table 12.1** Key material parameters for high-power high-performance transistors

| Parameter | Units | GaN | Si | GaAs | InP | 4 H-SiC |
|---|---|---|---|---|---|---|
| $m^*$ | $m_0$kg | 0.22 [75] | 1.56 [75] | 0.06 [75] | 0.07 [75] | 0.58 [75] |
| $\mu_e$ | $cm^2/V{\cdot}s$ | 1,245 [75] | 1,750 [75] | 9,340 [75] | 6,460 [75] | 1,000 [75] |
| DOS | $10^{18}$ $cm^{-3}$ | 2.3 [76] | 32 [76] | 0.47 [76] | 0.57 [76] | 24.9 [76] |
| $v_{sat}$ | $10^7$ cm/s | 1.4 [77] | 1 [78] | 0.72 [78] | 0.67 [78] | 0.33 [76] |
| $\varepsilon_r$ | | 9.4 [1] | 11.9 [1] | 12.5 [1] | 12.9 [1] | 10.0 [1] |
| EG | eV | 3.4 [1] | 1.12 [75] | 1.43 [75] | 1.35 [75] | 5.4 [75] |
| $E_{br}$ | $10^5$ V/cm | 20 [1] | 3 [1] | 4 [1] | 4.5 [1] | 35 [1] |
| $\kappa$ | W/°K·cm | 1.12 [75] | 1.56 [75] | 0.45 [75] | 0.68 [75] | 3.7 [75] |

electron mobility ($\mu_e$), the higher conduction-band density of states (DOS) and large saturation velocity ($v_{sat}$) still permit large current densities. The lower dielectric constant ($\varepsilon_r$) reduces capacitive loading and allows for large area devices, increasing RF current and power [1]. The large bandgap ($E_G$) improves radiation resistance, results in high intrinsic temperature, and provides a very large breakdown field ($E_{br}$) necessary for handling high RF power. Heat dissipation benefits from the high thermal conductivity ($\kappa$), allowing for more efficient dissipation of heat away from the device.

Mechanical strain resulting from lattice mismatch between the AlGaN and GaN layers induces piezoelectric polarization. This polarization increases the two-dimensional electron gas (2DEG) sheet carrier density ($n_s$). Benefitting from mechanical strain, AlGaN/GaN HEMTs are capable of achieving $n_s$ greater than $10^{13}$ $cm^{-2}$, without intentional doping. This is significantly higher than other III-V systems due to strong piezoelectric polarization in the wurtzite GaN and AlN. A cross-section schematic of an AlGaN/GaN HEMT is shown in Fig. 12.1.

Although excellent AlGaN/GaN HEMT performance has been demonstrated, electrical stability and reliability issues of these devices remain obstacles to further development. Trapping centers at the AlGaN surface, AlGaN/GaN interface, and/or GaN bulk are considered the main origin of GaN reliability issues. Degradation effects associated with traps are characterized by low frequency noise [2–4], transconductance frequency dispersion [2, 5], current collapse [6, 7], gate-lag and drain-lag transients [8–10], threshold voltage shift [11], increased gate current [12], and light sensitivity [6, 8, 11]. Charging and discharging of traps can limit device performance. Generation of new traps can permanently degrade the device, even to breakdown. Understanding the degradation mechanisms is necessary for improving the performance and reliability of AlGaN/GaN HEMTs.

One of the most widely accepted theories explaining AlGaN/GaN HEMT degradation is based on the inverse-piezoelectric effect [13, 14]. During operation, the large vertical field under the gate creates strain in the AlGaN barrier, adding to the existing strain from lattice mismatch. Once the strain surpasses the material's critical limit (at the critical voltage), relaxation will occur through crystallographic defect formation. These generated defects act as trapping centers for electrons, degrading performance and reliability. Large amounts of mechanical strain can certainly cause cracks and defects to form; however, even nondestructive amounts

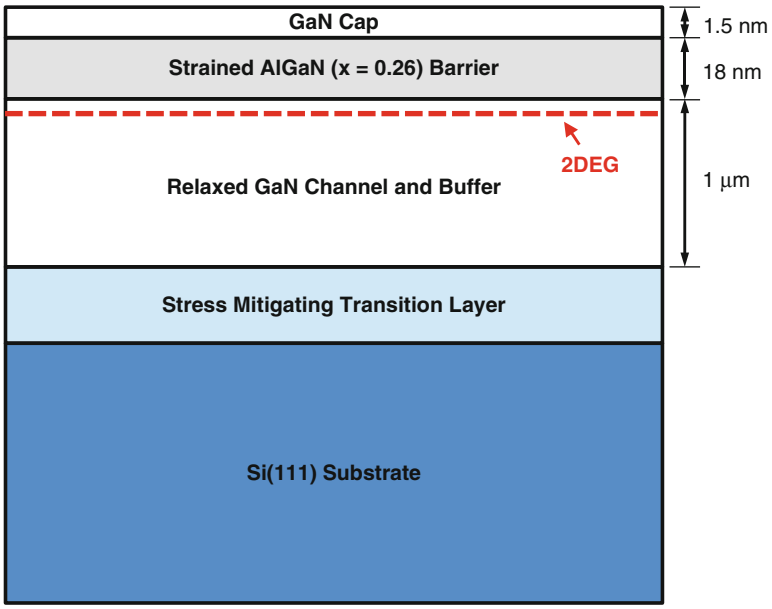



**Fig. 12.1** Cross-section schematic of AlGaN/GaN HEMT on Si(111) substrate

of strain also can impact performance and reliability of the device. Strain reduces crystal symmetry, reorienting the energy band structure with results in lifting of band degeneracies, shifting band energies, warping bands [15], and even altering trap energy levels [16]. This can affect carrier mobility by changing conductivity effective mass, density of states, and scattering, as well as impacting reliability, by increasing gate current, and by increasing hot-carrier effects [17].

Since stress is a major factor in the operation, performance, and reliability in AlGaN/GaN HEMT devices, a thorough understanding of the impact of stress on performance and reliability can lead to improvements in device design. The effects of strain in Si MOSFETs are well understood and have been used to improve the devices. Mechanical wafer bending is a cost-effective method to study the effects of stress on semiconductor devices which has been extensively used to isolate and study the effect of stress in Si MOSFETs. In this chapter, a systematic study of the effects of externally applied mechanical stress on the AlGaN/GaN HEMT channel resistance and gate current is presented to provide insights into the physical mechanisms responsible for stress-related performance and reliability issues.

## 12.2 Strain and Stress

To obtain insight into the underlying physics of strained HEMT devices, it is necessary to first introduce stress and strain.

### 12.2.1 Strain Definition

Strain in the context of a crystalline solid may be defined as the percentage change of the material lattice constant relative to its equilibrium value. Strain can result from lattice-mismatched film growth, phonon-induced lattice vibrations, and applied external mechanical stress. In silicon, beneficial strain reduces the crystal symmetry and thus lifts the energy band degeneracy and causes band warping in such a way to reduce the average conductivity effective mass. Any strain can be decomposed into a hydrostatic strain and two types of shear strain [18]. The hydrostatic strain does not break crystal symmetry and, therefore, only shifts energy levels without lifting band degeneracy. Thus, it is not important for carrier mobility enhancement. It is the shear component of strain that causes subband splitting and affects semiconductor transport properties. In the silicon semiconductor industry, strain is introduced into the device channel preferably by applying uniaxial stress. The uniaxial stress is longitudinal when parallel to the current channel and transverse when perpendicular to the channel.

### 12.2.2 Built-in Stress in HEMTs

Stress is an integral part of AlGaN/GaN HEMT devices. Large mechanical stress profiles are created during fabrication and are generated during device operation. These stresses can impact device performance and reliability.

#### 12.2.2.1 Lattice Mismatch Stress

In order for AlGaN/GaN HEMTs to be commercially competitive, low-cost, large-scale production must be achieved. Si (111) substrates offer pontential advantages of low cost, large size, and high quality over sapphire and SiC substrates.

However, large differences in lattice constants (~17%) and thermal expansion coefficients (TEC) (~56%) between GaN and Si (111) produce large strains, resulting in the formation of crystallographic defects [19]. High quality GaN layers on Si, free of cracks and dislocations, have been fabricated through implementation of stress mitigation using transition layers [19]. It is hypothesized that the lattice mismatch stress is primarily absorbed by the Al/Si interface, while the (Al, Ga) N transition layer absorbs the TEC mismatch stress, which occurs during processing [19].

Another type of stress induced during fabrication is biaxial tensile stress in the AlGaN barrier layer. The thin AlGaN barrier is pseudomorphically grown on the relaxed GaN channel/buffer. Lattice mismatch between AlGaN and GaN induces a biaxial tensile stress in the AlGaN barrier. For an Al concentration of 26%, the AlGaN barrier has ~3 GPa of biaxial tensile stress induced. This stress

is advantageous since wurtzite GaN and AlN grown in the (0001) orientation are both strongly piezoelectric [20]. The piezoelectric effect results in a polarization fixed charge at the AlGaN/GaN interface, inducing a mobile sheet charge layer which is termed a two-dimensional electron gas (2DEG). In addition, $SiO_2$ or $Si_3N_4$ passivation layers possess residual stress which also has been shown to induce stress, adding to the lattice mismatch stress and increasing the 2DEG density [21].

#### 12.2.2.2 Inverse-Piezoelectric Stress

During operation, the vertical electric field under the gate contact induces additional stress in the AlGaN barrier through the inverse-piezoelectric effect. This vertical field is the largest at the gate edges, where significant amounts of stress (500 MPa) can be generated in the AlGaN barrier during normal operation ($V_{GS} = 30$ V). This stress in addition to the lattice mismatch stress may vary the HEMT performance and reliability and therefore has attracted researchers' attention in the past decade. It has been proposed that stress generated from the inverse-piezoelectric effect can initiate defect formation, leading to irreversible degradation [13].

### *12.2.3 Wafer Bending Stress*

Mechanical wafer bending is a simple and cost-effective way to investigate the underlying physics of strain in semiconductors. Fabricating several wafers with varying amounts of process-induced stress is expensive. Furthermore, it can be difficult to accurately quantify the amount of stress present in the device. Also, modifying the process flow to alter the internal stresses can impact other characteristics of the device. Therefore, mechanical wafer bending is fundamental in performing controlled stress experiments.

Several methods have been used to externally apply mechanical stress to semiconductors. The first piezoresistance measurements by Smith [22] were achieved by hanging weights from slabs of semiconductors. This method requires large samples; therefore, only bulk measurements can be taken. The maximum stress achieved in Smith's study was 10 MPa. Bending of cantilevers or beams anchored at one end is another possible way to apply stress. While cantilevers are often used in microelectromechanical systems (MEMS) for transducer and resonator applications, the stress profile is nonuniform along the length of the beam, making specification of the applied stress difficult. Three-point bending can also be used. However, estimation of stress is difficult since stress varies between the three-point loads. The external mechanical-bending stress in this book chapter are obtained by using a flexure-based four-point wafer bending system [23], which achieves significantly improved uniformity of stress in the region between the inner load points.

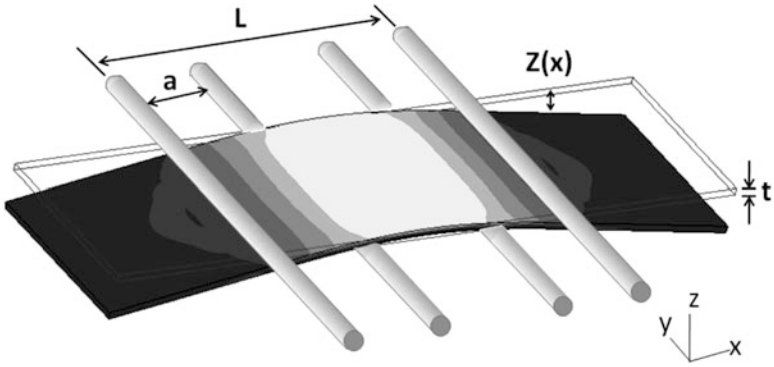



**Fig. 12.2** Illustrations of the four-point wafer bending setup for applying uniaxial stress. © [2008] American Institute of Physics (Reprinted with permission from Chu et al. [39], Fig. 3 and Fig. 5

#### 12.2.3.1 Four-Point Bending

In the four-point bending system, a beam is supported by two anchored points while being deformed by two driving loads as shown in Fig. 12.2. Between the center two rods, the sample is bent with a constant radius of curvature resulting in uniform stress. Therefore, unlike cantilevers and three-point bending, variation of device position does not affect the accuracy of the measurements. The magnitude of uniaxial stress on the top surface of a homogenous material sample between the center two rods can be calculated by [24]

$$\sigma = \frac{EZt}{2a\left(\frac{L}{2} - \frac{2a}{3}\right)} \tag{12.1}$$

where $E$ is Young's modulus along the stress direction, $Z$ is the sample vertical displacement, $t$ is the thickness of the sample, $a$ is the distance between the inner and outer rods, and $L$ is the distance between the two outer rods.

The magnitude of applied stress was calibrated by comparing the calculated stress to readings from a noncontact fiber optical displacement system and strain gauge measurements [25].

#### 12.2.3.2 Bending Measurements of Small Samples

In this work, AlGaN/GaN HEMT samples were diced into ~1 $cm^2$ size to maximize the number of measureable samples. However, the samples are smaller than the minimum size that can be directly bent in the four-point wafer bending setup. To apply stress to these small samples, we developed a technique to bend small wafer samples in the standard flexure-based wafer bending setup.

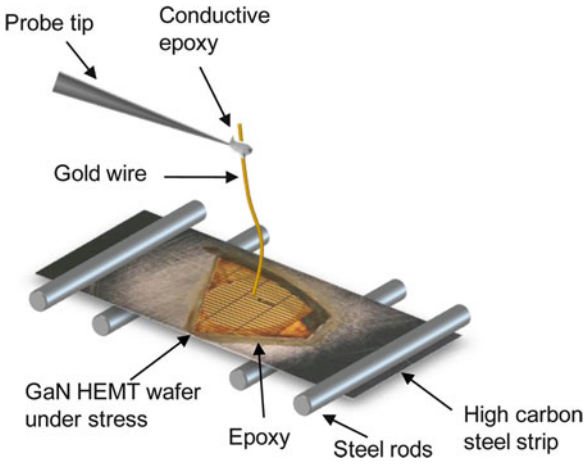



**Fig. 12.3** GaN wafer sample attached to heat-treated high-carbon stainless steel inserted in a four-point bending setup

This technique of applying mechanical stress to a small AlGaN/GaN HEMT wafer sample starts by attaching the wafer sample to a heat-treated high-carbon stainless steel plate (Fig. 12.3). First, the steel strip was sanded with fine grit sandpaper to remove oxidation and to provide a rough surface for adhesion. A thin layer of Epoxy Technology H74 two part epoxy was then applied to the middle of the steel strip. The wafer sample was placed on top of the epoxy and pressed down, and excess epoxy was wiped away. To eliminate air pocket formation during curing of the epoxy, a metal washer was placed on top of the wafer sample, and the sample was clamped with a metal binder clip. Then, the sample was inserted into a 100 °C oven for 5 min. The washer and metal binder were removed, and the sample was placed on a 150 °C hotplate for 5 min to complete the curing process. The sample attached to the steel plate was then inserted into the wafer bending setup.

Under the maximum amount of stress applied (360 MPa) in the experiments, the stainless steel plate does not permanently deform. A strain gauge is mounted on the top of the device with epoxy to calibrate the stress. As shown in Fig. 12.4, stress is applied and released from the sample. The amount of stress read from the strain gauge returns to the starting point, verifying that the stress applied to the sample is elastic.

To characterize the impact of mechanical stress on the AlGaN/GaN HEMT devices, electrical measurements are taken while simultaneously varying the amount of applied stress. In order to achieve this, wires were attached to the device bond pads. Standard ball- and wedge wire-bonding techniques resulted in delamination of bond pads destroying the device. Therefore, a technique was developed to attach wires to the bond pads without the use of heat or ultrasonic energy. First, a ball was formed on the end of a 1-mil Au wire in a ball-bonding machine. Then, the wire was cut to approximately 1 cm length and removed from the ball bonder. A small amount of electrically conductive Epoxy Technology EE129-4 two part epoxy was placed on the end of a probe tip. The probe tip was brought into contact with the end of the wire without the ball and cured for 24 h at room temperature.

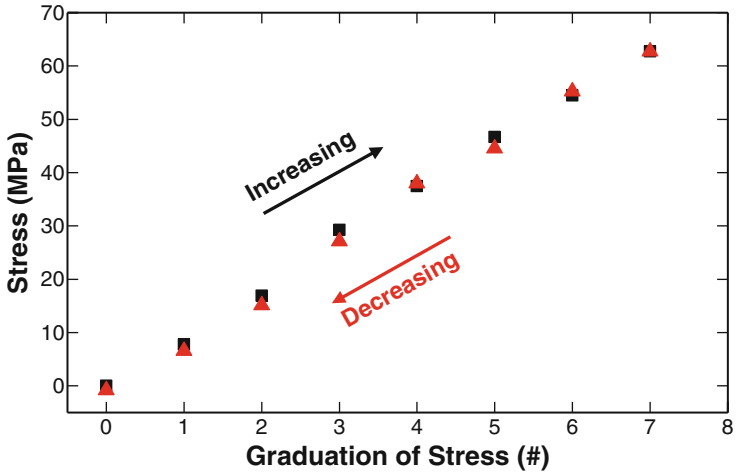



**Fig. 12.4** Increasing and decreasing stress applied to a wafer mounted on a stainless steel strip. A strain gauge is used to determine the amount of applied stress

The ball end of the wire was dipped in conductive epoxy, and the probe tip was inserted into the micropositioner. Using the micropositioner, the ball end of the wire with conductive epoxy was landed on the device's bond pad and left to cure for another 24 h at room temperature. After the epoxy cured, the micropositioner was lowered to allow slack on the wire for displacement of the wafer while applying stress.

### *12.2.4 Gauge Factor*

The gauge factor gives straightforward experimental information about strain-enhanced carrier mobility in semiconductors. This coefficient is defined as the normalized change in resistivity per unit strain

$$\mathrm{GF} = \frac{\Delta\rho}{\rho \cdot \varepsilon} \tag{12.2}$$

$$\rho = \frac{1}{q\mu_n n + q\mu_p p} \tag{12.3}$$

where $\varepsilon$ is the strain and $\rho$ is the resistivity, which can be calculated by Eq. 12.3.

For Si MOSFET under steady state, the electron and hole densities are approximately constant; thus, its GF is determined by the change in carrier mobility with stress. For GaN HEMT, however, the 2DEG density may vary under mechanical stress, due to a change in polarization. As a result, its GF is determined by both the

change in carrier mobility and carrier density. The gauge factor is a measure of the drive current enhancement under a particular stress and, therefore, has been widely used to estimate strained device performance [26–28].

## 12.3 Strain Effects on GaN HEMT Drain Current

### *12.3.1 Background*

The effect of stress on channel resistance has been studied, due to its relevance for reliability issues such as hot electron injection as well as for strain sensors [29]. Gaska et al. [30] applied biaxial compressive stress to study its effect on channel electron density. Their work observed a decreasing electron density with compressive stress. Eichhoff et al. [31], Kang et al. [32, 33], Zimmermann et al. [34], and Yilmazoglu et al. [35] applied mechanical stress to investigate the piezoresistive property of GaN HEMTs. A large variation in gauge factors has been reported in the literature, ranging from −4 to −40,000 [30, 31, 33–37]. This large disagreement likely results from inaccuracies in resolving the applied stress and changes in the trapped charge density over the time elapsed during measurement. These past studies used three-point bending [30, 31, 34, 36], cantilevers [34, 37], complex lever mass system [35], and circular membranes [33] to apply stress, which can be difficult to accurately quantify the amount of stress applied.

In our experiment, we use four-point bending while mitigating the effects of trap charging/discharging to experimentally characterize the effect of stress on AlGaN/GaN HEMTs. To better understand the stress effect on HEMT device channel resistance and accurately model the gauge factor, we also developed a simulation model by considering the stress-altered two-dimensional electron gas (2DEG) sheet carrier density and electron mobility. For the first time, uniaxial stress is incorporated into a $sp^3d^5$ empirical tight-binding model to investigate the stress effect on electron effective mass in wurtzite GaN.

### *12.3.2 Measurement*

Wafer samples were attached to heat-treated high-carbon steel plates with epoxy and stressed in a four-point wafer bending setup. Compressive and tensile uniaxial stress up to 360 MPa was applied longitudinal to the channel direction. To obtain an accurate measurement of the AlGaN/GaN HEMT gauge factor, parasitic charge-trapping transients and external resistances were addressed. After the effects of charge trapping were eliminated, and external resistances were accounted for, an accurate gauge factor measurement was performed [38].

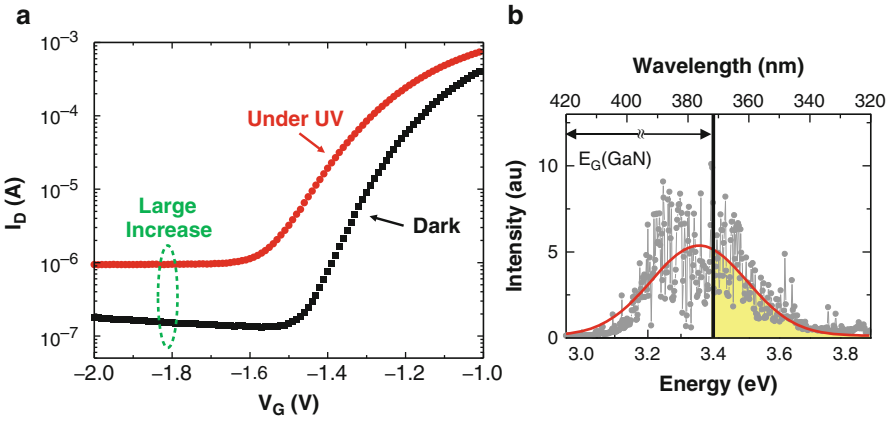



**Fig. 12.5** Unfiltered UV measurement. (**a**) $I_D$-$V_G$ measurements in dark and under unfiltered UV light with a large increase in off-state current and a decrease in subthreshold slope. (**b**) The spectral output of the unfiltered UV light

### 12.3.2.1 Experiment Setup for Stabilized Measurements

To combat the instability issue associated with trapped charges, the HEMT device was exposed to light with photon energy near, but below, the bandgap of GaN (~3.4 eV or 365 nm wavelength) to photoionize all trapped electrons influencing the resistance measurement without band-to-band generation of electron-hole pairs. Initially, a mercury arc ultraviolet (UV) spotlight with peak wavelength of 377.7 nm or 3.284 eV was chosen to illuminate the device under test. A sweep of $I_D - V_G$ under UV spotlight illumination compared to dark (Fig. 12.5a) showed a large increase in off-state drain current and a decrease in subthreshold slope. The spectral intensity of the light source was measured in a spectrometer. A significant portion of the photon energy was above the bandgap of GaN (3.39 eV ~ 365 nm) as shown in Fig. 12.5b. Under illumination of above bandgap light, mobile electron-hole pairs are photogenerated. An increase in off-state current and a decrease in subthreshold slope are consistent with carrier photogeneration. A 380-nm band-pass filter was then applied to filter out wavelengths below 365 nm (Fig. 12.6b). A horizontal shift in subthreshold slope and similar off-state leakage current (Fig. 12.6a) verified a decrease in the effect of trapped charge without photo-generation of electron-hole pairs.

A schematic of the experimental setup is shown in Fig. 12.7. The standard wafer bending setup is illuminated by the UV light source. The band-pass filter is mounted in a 4-in.-thick polystyrene heat shield to block ambient heat from the mercury arc lamp and block nonfiltered light from illuminating the device. As shown in Fig. 12.8, over 1,500 s, the change of the measured channel resistance is less than 0.02%. Since the resistance measurement has been stabilized, it is now appropriate to apply stress to extract the gauge factor.

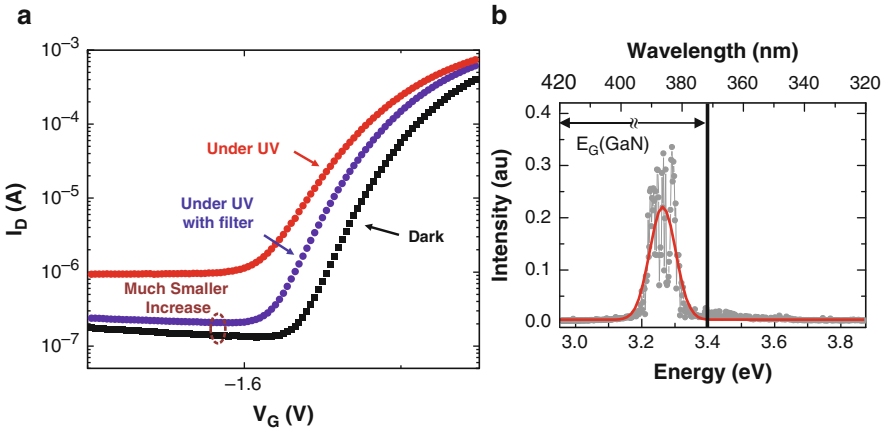



**Fig. 12.6** Filtered UV measurement. (**a**) $I_D$-$V_G$ measurements in dark and UV light filtered by a 380-nm band-pass filter with a much smaller increase in off-state current and no subthreshold slope change. (**b**) The spectral output of UV light with 380-nm band-pass filter

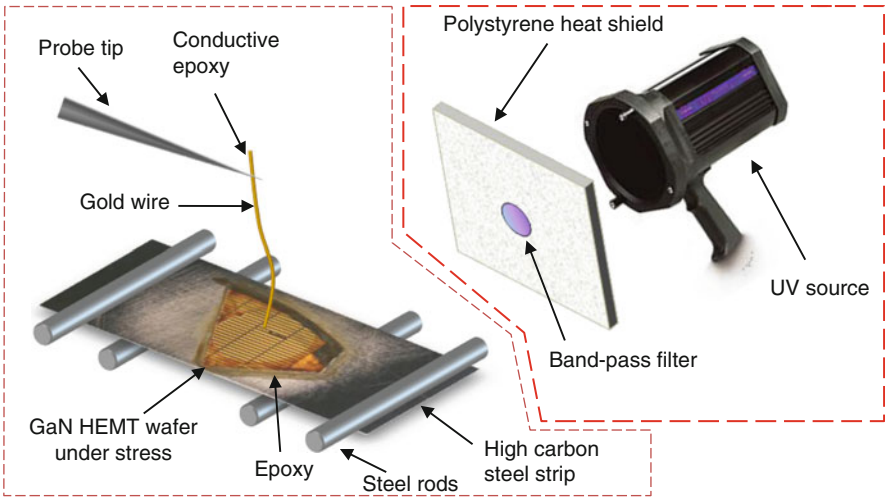



**Fig. 12.7** Experimental setup for photoionizing trapped charge to measure the gauge factor

### 12.3.2.2 External Resistance Consideration

The stress dependence of the channel resistance ($R_{CH}$) was measured at $V_{GS} = -1$ V and $V_{DS} = 0.1$ V, by excluding source/drain contact resistances. The high conductivity of GaN 2DEG results in a small channel resistance, especially for devices characterized with W/L ratio of 25. The measured resistance ($R_{meas}$) is the

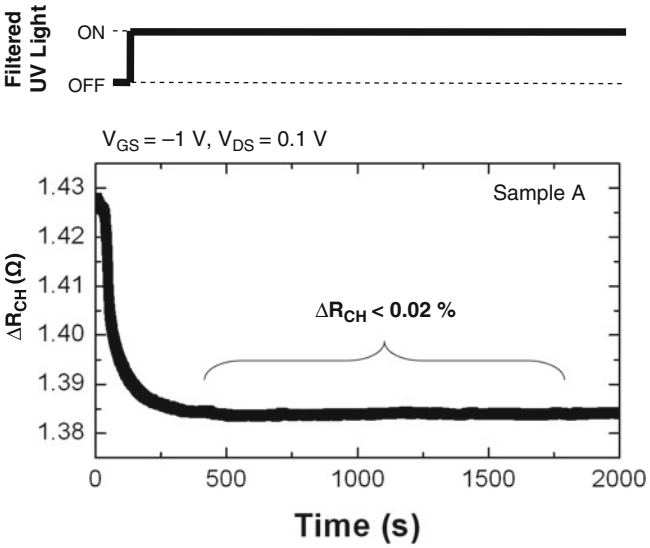



**Fig. 12.8** Illuminating the device under test with UV light-stabilized $R_{CH}$ to less than 0.02% variation for 1,200 s of measurement

sum of the channel resistance ($R_{CH}$), source contact resistance ($R_S$), drain contact resistance ($R_D$), and external parasitic resistances ($R_{ext}$) and was on the order of 100 Ω:

$$R_{meas} = R_{CH} + R_S + R_D + R_{ext}. \tag{12.4}$$

The source and drain contact resistances ($R_S = R_D = 5\ \Omega$), measured by transmission line measurements, are subtracted from the measured resistance and are assumed to have a negligible stress dependence. A four-point Kelvin measurement is used to eliminate the effect of external resistances. Two wires were bonded to both the source and drain pad and one to the gate. One pair of source and drain contacts are used to supply a dc current via the force connections on the semiconductor parameter analyzer. The other pair of connections are used to sense the voltage drop across the source and drain pads.

#### 12.3.2.3 Experimentally Obtained Gauge Factor

Longitudinal uniaxial stress was varied in 60 MPa increments and held for 100 s at each interval. The normalized change in $R_{CH}$ was measured for incrementally applied compressive and tensile stress up to 360 MPa, which was then released incrementally to zero as shown by the dotted lines of Fig. 12.9. Tensile stress

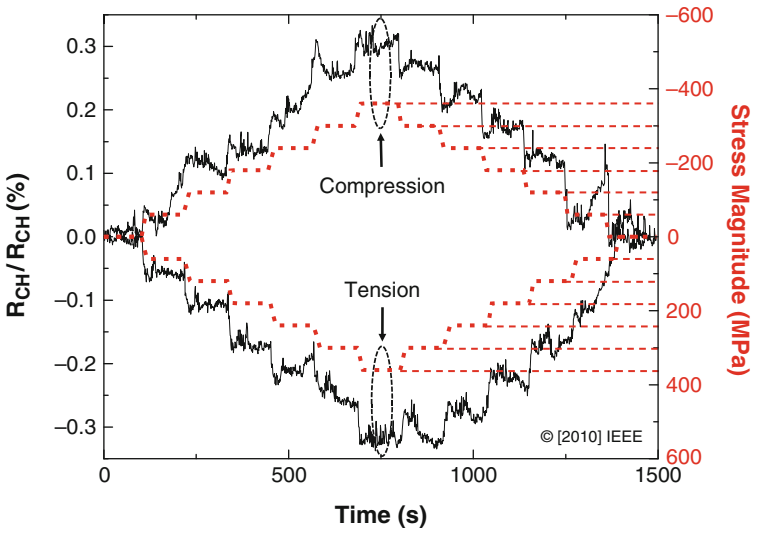



**Fig. 12.9** Normalized change in channel resistance with incrementally increasing and decreasing uniaxial stress (Reprinted, with permission, from Koehler et al. [38], Fig. 2)

decreases $R_{CH}$, while compressive stress increases in $R_{CH}$ are seen by the solid experimental lines of Fig. 12.9. At the maximum applied stress (360 MPa), the normalized resistance change was ~0.83%/100 MPa, which is much smaller than what is observed in (001)/<110> silicon nMOSFETs of 3.2%/100 MPa [39]. The resistance returned to the initial unstressed value after increasing and decreasing the compressive and tensile stress. This demonstrates that the change in resistance observed is due to a reversible strain effect, opposed to charge-trapping/detrapping transients.

The gauge factor was determined by averaging the $R_{CH}$ measurements over each time interval during which the stress was held constant (Fig. 12.10). Error bars represent a three standard-deviation confidence interval for the measurement. The slope of a total least squares linear fit of the averaged $R_{CH}$ versus strain curve was obtained to determine a gauge factor of $-2.5 \pm 0.4$. Total least squares analysis included uncertainty of the measurements. The determined gauge factor ($-2.5 \pm 0.4$) is small relative to values in literature ranging from $-4$ to $-40{,}000$.

### *12.3.3 Theory and Modeling*

The channel resistance of the AlGaN/GaN HEMT device is inversely proportional to the 2DEG sheet carrier density ($n_S$) and channel electron mobility ($\mu_e$) [40].

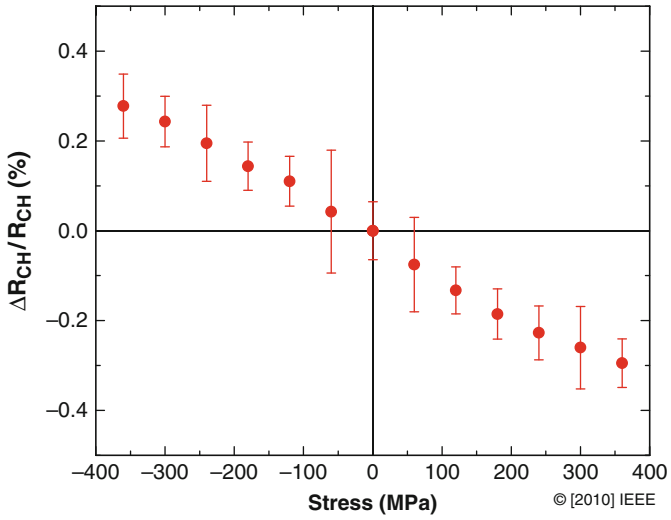



**Fig. 12.10** $R_{CH}$ measurements at each time interval stress was held constant. *Error bars* represent a three standard deviation confidence interval for the measurement (Reprinted, with permission, from Koehler et al. [38], Fig. 4)

#### 12.3.3.1 Stress Dependence of 2DEG Sheet Carrier Density

The 2DEG forms at the interface of the AlGaN and GaN layers, arising from the total polarization difference between them [20]. There are two types of polarization: spontaneous polarization ($P^{SP}$) and piezoelectric polarization ($P^{PE}$). Spontaneous polarization exists in both AlGaN and GaN layers, since their $\sqrt{c/a}$ ratio differs from the ideal wurtzite crystal value ($\sqrt{8/3}$). Piezoelectric polarization arises from the piezoelectric effect, which is proportional to the strain. In the "as-fabricated" AlGaN/GaN HEMT structure, a thin layer of strained AlGaN due to lattice mismatch is on top of a thick layer of relaxed GaN. As a result, piezoelectric polarization exists only in the AlGaN layer without external stress. Figure 12.11a schematically shows the total polarizations in the "as-fabricated" device. From the total polarization difference between the AlGaN and GaN layers, $\sigma = P(\text{AlGaN}) - P(\text{GaN})$, the 2DEG sheet carrier density can be calculated using Eq. 12.5 [20]

$$n_S(x) = \frac{+\sigma(x)}{q} - \left(\frac{\varepsilon_0\varepsilon(x)}{dq^2}\right)[q\phi_b(x) + E_F(x) - \Delta E_C(x)] \tag{12.5}$$

where $x$ is the Al content, $\sigma$ is the total polarization difference between AlGaN and GaN layers, $\varepsilon$ is the dielectric constant, $d$ is the depth of the AlGaN layer, $q\phi_b$ is the Schottky barrier of a gate contact, $E_F$ is the Fermi level with respect to the GaN

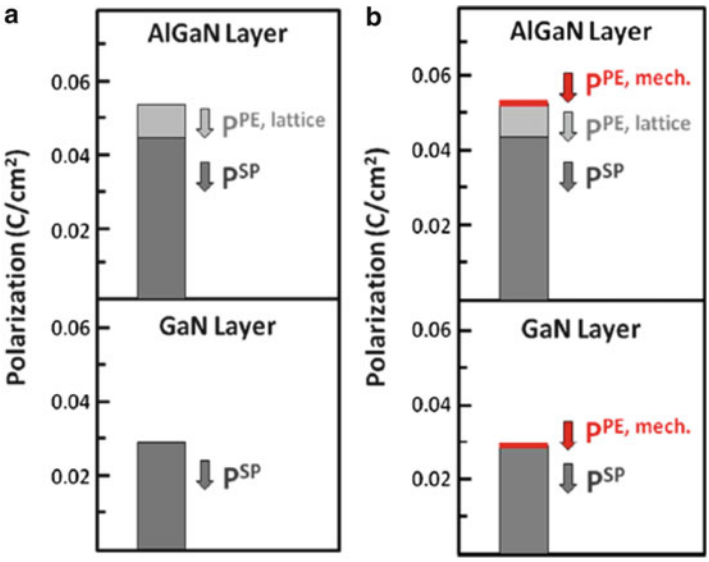



**Fig. 12.11** Illustrations of polarizations in the AlGaN and GaN layers. (**a**) For the as-fabricated GaN HEMTs. (**b**) For the GaN HEMTs under mechanical-bending stress. The mechanically applied stress generates additional piezoelectric polarization of similar magnitude in both the AlGaN and GaN layers. © [2010] American Institute of Physics (Reprint with permission from Chu et al. [40], Fig.3)

conduction-band-edge energy, and $\Delta E_C$ is the conduction-band offset at the AlGaN/GaN interface. This work uses an Al content of 0.26.

External mechanical stress affects the 2DEG density by generating additional piezoelectric polarization along the [0001] direction. Spontaneous polarization remains constant since it is an intrinsic material quality. When external stress is applied, additional piezoelectric polarization arises in both layers as shown in Fig. 12.11b. The amount is proportional to the strain and piezoelectric coefficients ($P^{\mathrm{PE,mech}} = e_{ij} \cdot \varepsilon$). In this work, the AlGaN and GaN layers are assumed to have the same level of strain, due to the fact that in most mechanical-bending experiment, both layers of the AlGaN/GaN HEMT are significantly thinner than the substrate and therefore are located near the top surface of the wafer. In addition, the piezoelectric coefficients of AlGaN and GaN are similar. As a result, the difference between the strain-induced piezoelectric polarizations of these two layers is close to zero. Therefore, we expect that external stress has little effect on the 2DEG sheet carrier density.

In this work, the simulation uncertainty is investigated by considering various elastic stiffness constants and piezoelectric coefficients of GaN and AlN listed in literature. The uncertainty arising from the tight-binding parameters is neglected since these parameters are achieved from an empirical tight-binding method that is considered to be accurate.

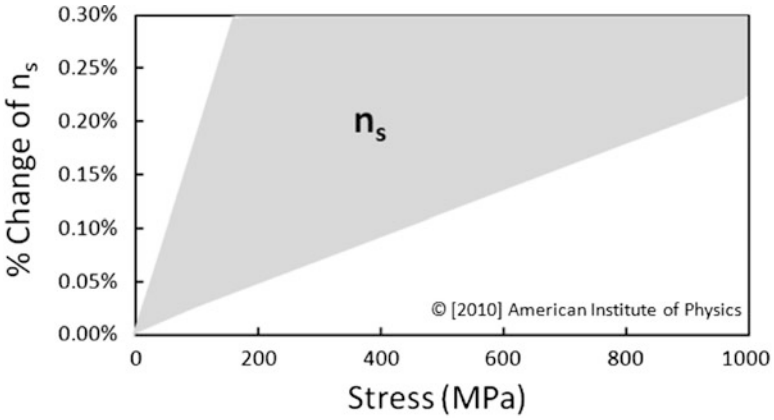



**Fig. 12.12** Percent change of 2DEG sheet carrier density under $[11\bar{2}0]$ uniaxial tension. Simulation uncertainty is shown as the shaded area by considering various elastic stiffness constants and piezoelectric coefficients. © [2010] American Institute of Physics (Reprint with permission from Chu et al. [40], Fig.4)

Figure 12.12 shows the percentage change of 2DEG sheet carrier density versus $[11\bar{2}0]$ tensile stress. An increase in $n_S$ ranging from 0.09 to 1.4% for 500 MPa can be achieved depending on the coefficients' value used in the simulation. This small enhancement matches our expectation due to the fact that the additional piezoelectric polarizations in AlGaN and GaN layers mostly cancel each other out.

#### 12.3.3.2 Stress Dependence of Channel Electron Mobility

Similar to the strained-Si analysis, the band structure of wurtzite GaN is the starting point of investigating strain-related GaN HEMT behavior. Nido [41] and Jogai [42] incorporated external stress into an empirical $sp^3d^5$ tight-binding model [43] to study the consequence of strain effects on the GaN band structure. Their work, however, considered only biaxial stress and focused mainly on the bandgap and valence-band structure. Since the electron is the majority carrier in the GaN HEMT channel, it is essential to study the strain-altered conduction band. In addition, uniaxial stress, which is the most beneficial stress for Si devices, is worth considering in GaN.

Electron mobility enhancement for bulk GaN can result from the average conductivity effective mass reduction and a suppression of intravalley scattering. Stress affects electron effective mass through two factors: band-splitting-induced electron repopulation and band warping. Unlike Si, which has six degenerate conduction bands, or GaAs which has energetically adjacent conduction bands, GaN is a direct bandgap material with only one conduction band. As a result, no electron repopulation occurs under stress, and thus band warping is the only mechanism responsible for effective mass change. The absence of conduction

band splitting also results in negligible change of the acoustic phonon scattering. The polar optical phonon scattering also has negligible dependence on stress, due to the fact that the mechanical stress does not induce polarization change along the longitudinal direction. Therefore, the change in effective mass through band warping is the dominant mechanism for stress-dependent bulk GaN electron mobility variation.

In a conventional GaN HEMT structure, the 2DEG is confined at the AlGaN/GaN interface and electron energy is quantized. However, there will not be confinement-induced subband splitting in GaN HEMT, since GaN has only one single conduction band. Therefore, similar to bulk GaN, the scattering rate change in strained GaN HEMT remains negligible, and the band-warping-induced effective mass variation should dominate its stress behavior.

We use an $sp^3d^5$ empirical tight-binding method developed in [43] to calculate the GaN band structure and electron effective mass. The unit cell of wurtzite GaN contains four atoms, two anions (N), and two cations (Ga), as shown in Fig. 12.13a. All nearest-neighbor s, p, and d interactions, as well as second-nearest-neighbor s and p interactions, are included using two-center approximation [44], which results in a $26 \times 26$ Hamiltonian matrix. The tight-binding parameters used in this work, including five on-site, one-center, eight nearest-neighbor two-center, and eight second-nearest-neighbor two-center integrals, are listed in Table IV of [43]. The GaN band structure is obtained by solving the eigenvalue of the Hamiltonian matrix. The electron effective mass is then calculated along the Γ-M, Γ-K, and Γ-A directions in the reciprocal lattice as shown in Fig. 12.13b.

The effect of mechanical stress is incorporated into the tight-binding model by considering the strain-induced change in atom location, which results in varied bond length, bond angle, and reciprocal lattice. In this work, in-plane biaxial stress and uniaxial stress along the $x$ and $y$ axis as shown in Fig. 12.13a are considered.

Under biaxial stress, the hexagon in the c-plane does not deform but only varies in size as shown in Fig. 12.14a. The resulting strain in all three directions is related to the applied stress through Eqs. 12.6 and 12.7

$$\varepsilon_{xx} = \varepsilon_{yy} = \frac{C_{33}}{(C_{11} + C_{12})C_{33} - 2C_{13}^2} \cdot \sigma \tag{12.6}$$

$$\varepsilon_{zz} = \frac{-2C_{13}}{(C_{11} + C_{12})C_{33} - 2C_{13}^2} \cdot \sigma \tag{12.7}$$

where $\varepsilon_{xx} = \varepsilon_{yy}$ is the biaxial stress, and $C_{11}$, $C_{12}$, $C_{13}$, and $C_{33}$ refer to the GaN elastic stiffness constants. The primitive translation vectors in rectangular coordinates become $a_1 = [a(1 + \varepsilon_{xx}), 0, 0]$, $a_2 = [-a/2(1 + \varepsilon_{xx}), \sqrt{3}a/2(1 + \varepsilon_{yy}), 0]$, and $a_3 = [0, 0, c(1 + \varepsilon_{zz})]$, where $a$ is the length of a hexagon side distance and $c$ is the repeat z direction. The corresponding reciprocal lattice vectors are $b_1 = [2\pi/(a(1 + \varepsilon_{xx})), 2\pi/(\sqrt{3}a(1 + \varepsilon_{yy})), 0]$, $b_2 = [0, 4\pi/(\sqrt{3}a\ (1 + \varepsilon_{yy})), 0]$, and $b_3 = [0, 0, 2\pi/(c(1 + \varepsilon_{zz}))]$. The reciprocal lattice, under biaxial stress,

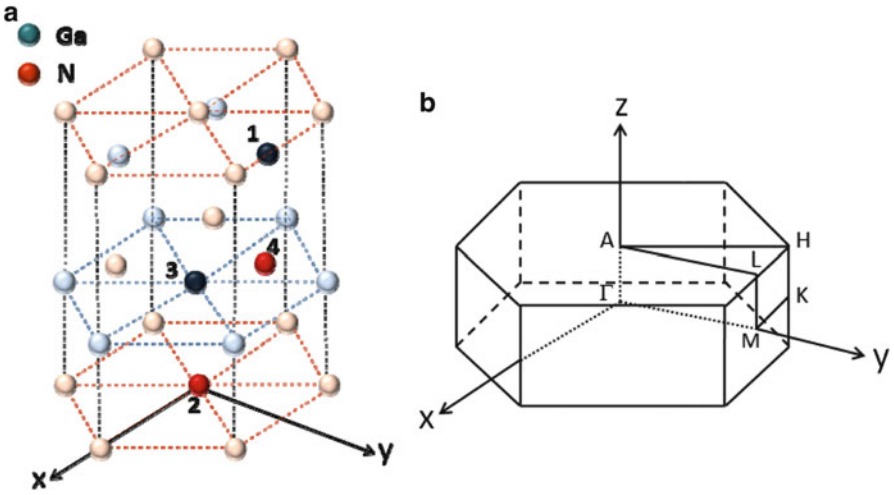



**Fig. 12.13** Wurtzite GaN structures in the real and reciprocal spaces. (**a**) Unit cell with 4 basis atoms, where 1 and 3 are Ga atoms, 2 and 4 are N atoms. (**b**) The reciprocal lattice for unstrained wurtzite GaN. Electron effective mass is calculated along Γ-K, Γ-M, and Γ-A directions. © [2010] American Institute of Physics (Reprint with permission from Chu et al. [40], Fig. 5)

remains a hexagonal shape. The basis vectors, where the atoms are located, are $t_1 = [0,\ \sqrt{3}a/3(1+\varepsilon_{yy}), 7c/8(1+\varepsilon_{zz})]$, $t_2 = [0,0,0]$, $t_3 = [0,0,3c/8(1+\varepsilon_{zz})]$, and $t_4 = [0,\ \sqrt{3}a/3(1+\varepsilon_{yy}), c/2(1+\varepsilon_{zz})]$. The strain-varied bond length and bond angle between all nearest and second-nearest-neighbors can then be calculated based on the new atom locations. Since the atomic layout in the c-plane remains hexagonal, the expressions of elements in the $26 \times 26$ Hamiltonian matrix do not change. The stress-dependent band structure calculation is straightforward, with the bond length and bond angle the only parameters needed to be changed.

Under uniaxial stress, however, the in-plane hexagonal shape is deformed as shown in Fig. 12.14b. The resulting strain in all three directions is related to the applied stress through Eqs. 3.8, 3.9) and 3.10.

$$\varepsilon_{xx} = \frac{C_{11}C_{33} - C_{13}^2}{C_{11}\left(C_{11}C_{33} - 2C_{13}^2\right) - C_{12}\left(C_{12}C_{33} - 2C_{13}^2\right)} \cdot \sigma \tag{12.8}$$

$$\varepsilon_{yy} = \frac{-\left(C_{12}C_{33} - C_{13}^2\right)}{C_{11}\left(C_{11}C_{33} - 2C_{13}^2\right) - C_{12}\left(C_{12}C_{33} - 2C_{13}^2\right)} \cdot \sigma \tag{12.9}$$

$$\varepsilon_{zz} = \frac{C_{13}(C_{12} - C_{11})}{C_{11}\left(C_{11}C_{33} - 2C_{13}^2\right) - C_{12}\left(C_{12}C_{33} - 2C_{13}^2\right)} \cdot \sigma \tag{12.10}$$

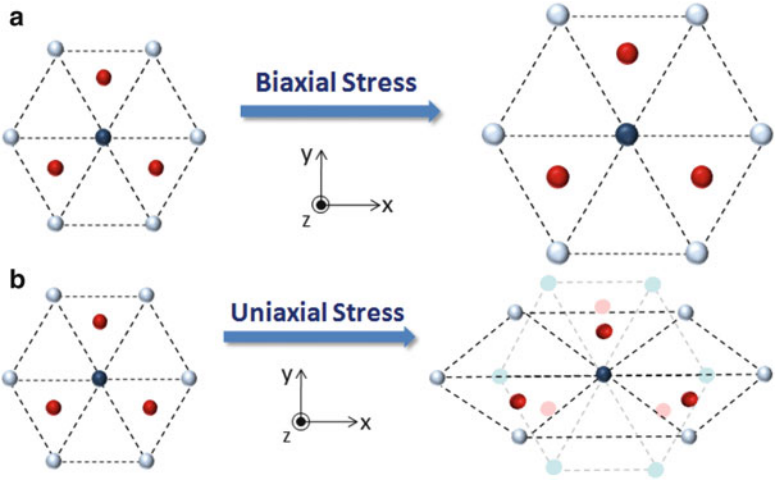



**Fig. 12.14** Lattice projection on the c-plane of GaN under externally applied mechanical stress. (**a**) Under biaxial stress, the atomic layout remains a hexagonal shape. (**b**) Under uniaxial stress, the hexagon deforms, and the strain in the *x*-direction differs from the strain in the y-direction. © [2010] American Institute of Physics (Reprint with permission from Chu et al. [40], Fig. 6)

The primitive translation vectors and basis vectors in rectangular coordinates are calculated the same way as under biaxial stress. Figure 12.14b shows that under uniaxial stress, the bond length between the central atom to atom 1 differs from the bond length between the central atom to atom 2 or atom 3. The original symmetry is broken, and therefore, the expressions of Hamiltonian matrix elements derived in [43] need to be regenerated, following a standard tight-binding method with the two-center approximation as described in [44]. All of atoms 1, 2, and 3, as shown in Fig. 12.14b, are still considered to be the nearest neighbors to the central atom. During our derivation, it is found that besides the bond length and bond angle, the expressions of $g_0$, $g_1$, $g_2$, and $g_3$ in [43] also change. The reciprocal lattice, under uniaxial stress, is not hexagonal in shape. The effective mass should be calculated along directions based on the new reciprocal lattice vectors.

Stress-induced band warping is determined by, first, whether the applied stress is along a high-symmetry direction, and second, whether there is strong subband interaction between the conduction band and any other subband. In this work, we study the effect of biaxial stress and uniaxial stress along the $[11\bar{2}0]$ and $[10\bar{1}0]$ directions. Although these types of stresses slightly alter the wurtzite crystal symmetry, we expect little change in electron effective mass and thus electron mobility. This is because GaN has a large bandgap and has no energetically adjacent conduction bands.

The stress effect on the electron effective mass along the longitudinal, transverse, and out-of-plane directions is calculated, and the results are plotted in Fig. 12.15a–c, respectively. Here we consider the $[11\bar{2}0]$ direction to be the channel direction. Therefore, the longitudinal, transverse, and out-of-plane directions refer

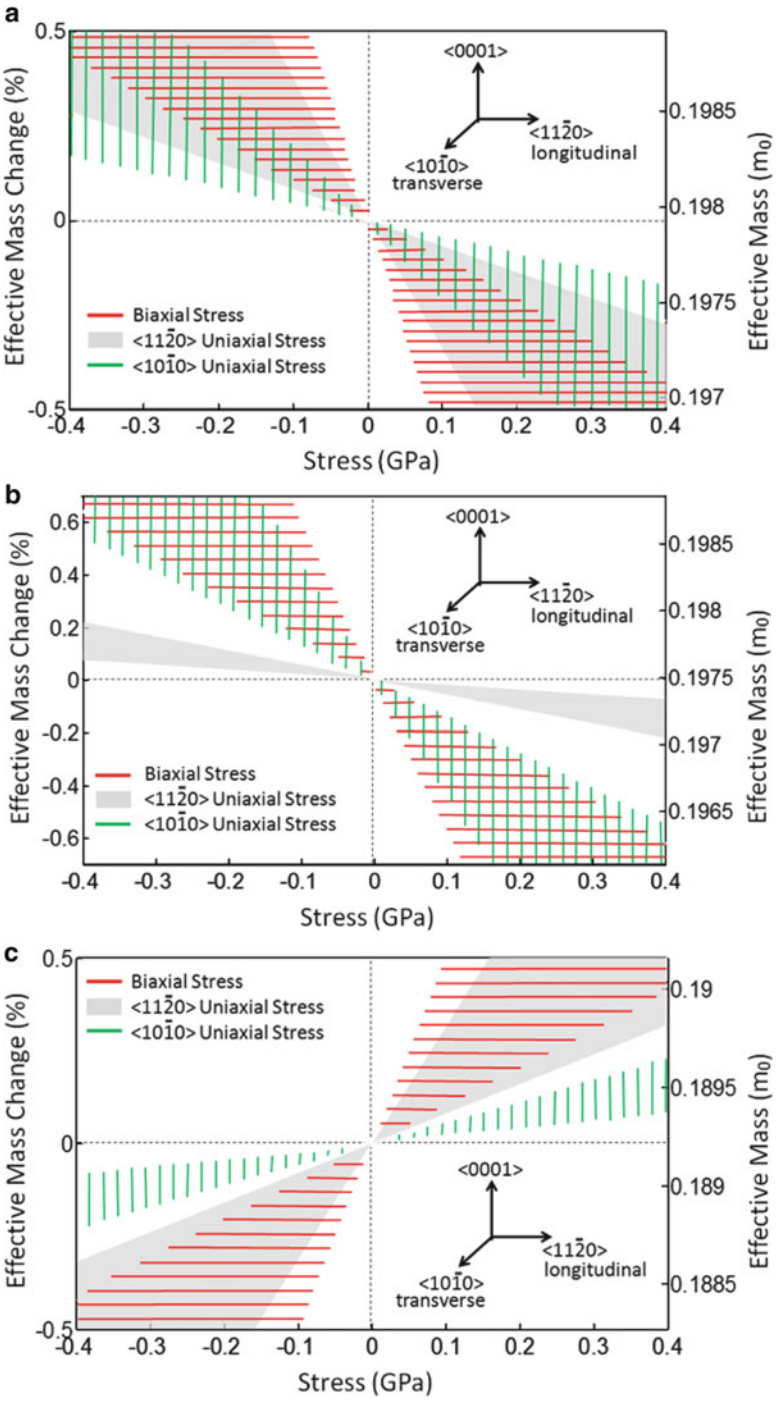



**Fig. 12.15** Change of electron effective mass under externally applied mechanical stress. (**a**) Longitudinal stress, (**b**) transverse stress, (**c**) biaxial stress © [2010] American Institute of Physics (Reprint with permission from Chu et al. [40], Fig. 7a)

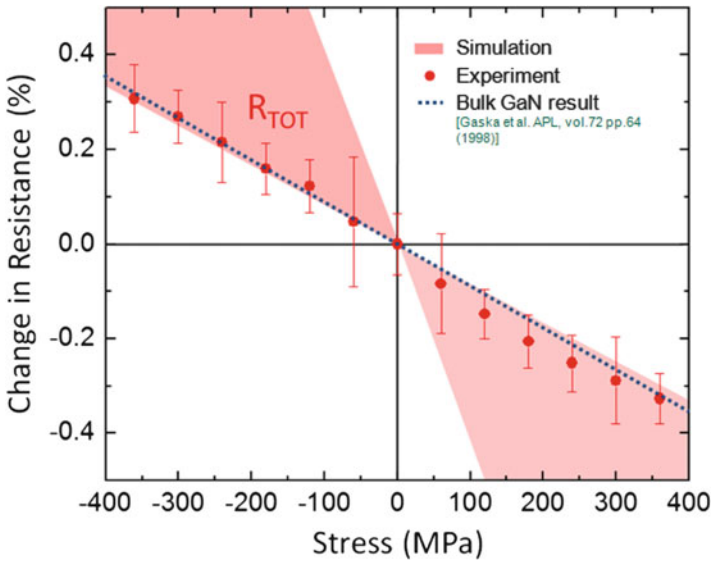



**Fig. 12.16** Change of GaN HEMT resistance ($R_{TOT}$) under longitudinal stress. *Symbols* represent experimental change in $R_{TOT}$ [38] with uniaxial stress. © [2010] American Institute of Physics (Reprint with permission from Chu et al. [40], Fig. 8)

to the $[11\bar{2}0]$, $[10\bar{1}0]$, and $[0001]$ direction, respectively. The shaded areas in the plots include all possible simulation results considering simulation uncertainties. Without stress, the longitudinal, transverse, and out-of-plane effective masses are 0.198, 0.197, and 0.189 $m_0$, respectively, which agree with the results in [43]. For longitudinal effective mass, biaxial stress has a slightly larger effect than uniaxial stress (~3%/500 MPa under biaxial stress comparing to ~1.5%/500 MPa under longitudinal stress). Similarly, transverse and out-of-plane effective mass also have a larger change under biaxial stress. For all types of stress, however, the changes in electron effective mass of GaN are much smaller than those of Si (~15%/500 MPa under <110> uniaxial stress [45]). This is because the subband splitting, which is an important factor affecting electron effective mass in Si, does not exist in GaN. The absence of band splitting and carrier repopulation in GaN causes the electron effective mass and thus mobility to only depend on conduction band warping, which is proved to be small through tight-binding calculations.

#### 12.3.3.3 Modeling Results for Gauge Factor

Combining the variation of 2DEG sheet carrier density and electron mobility, the stress-induced change in channel resistance of an AlGaN/GaN HEMT is plotted in Fig. 12.16. Simulation uncertainties are included. The simulation result is compared to our previous experimental result [38], in which repeatable gauge factors were

**Table 12.2** Gauge factors of GaN HEMTs and bulk GaN published in literature

| | Ref. | GF | $\varepsilon$ (%) | $\sigma$ (MPa) | Method of stressing |
|---|---|---|---|---|---|
| GaN HEMT | [38] | −2.6 | 0.114 | 360 | 4-point bending |
| | [30] | −4 | 0.03 | 95 | 3-point bending |
| | [31] | −42 | 0.005 | 15 | 3-point bending |
| | [36] | −75 | 0.04 | 126 | 3-point bending |
| | [34] | −90 | 0.14 | 442 | 3-point bending |
| | [32] | −1,259 | $1.35 \times 10^{-4}$ | 0.42 | Cantilever |
| | [33] | −38,889 | $3.85 \times 10^{-4}$ | 1.2 | Circular membrane |
| Bulk GaN | [31] | −3.5 | 0.005 | 15 | 3-point bending |
| | [46] | 1 ~ 4 | $4.25 \times 10^{-4}$ | 1.34 | 3-point bending |
| | [47] | −3.6 | | | 3-point bending |

**Table 12.3** The best fit set of stiffness constants of GaN and AlN © [2010] American Institute of Physics (Reprint with permission from Chu et al. [40], Table I)

| | $C_{11}$ | $C_{12}$ | $C_{13}$ | $C_{33}$ | $C_{44}$ | $C_{66}$ |
|---|---|---|---|---|---|---|
| GaN | 37.3 | 14.1 | 8.0 | 38.7 | 9.4 | 11.8 |
| AlN | 41 | 14 | 10 | 39 | 12 | 11.8 |

**Table 12.4** The best fit set of piezoelectric coefficients of GaN and AlN © [2010] American Institute of Physics (Reprint with permission from Chu et al. [40], Table II)

| | $e_{13}$ | $e_{33}$ | $e_{15}$ |
|---|---|---|---|
| GaN | −0.22 | −0.22 | 0.44 |
| AlN | −0.48 | −0.58 | 1.55 |

obtained after eliminating parasitic charge-trapping effects. The simulated gauge factor is determined to be $-7.9 \pm 5.2$, compared to $-2.5 \pm 0.4$ in experiment [38] and the values published in literature ranging from −4 to −40,000, as listed in Table 12.2. The wide range of gauge factors listed in literature is considered to be a result of the trapping effect occurring over the elapsed time of measurements. Our result also agrees with the gauge factors of bulk GaN reported in [31, 46, 47] with little variation (−1 to −3.6) that indicates negligible trapping effect. Comparing the simulation result with the experimental result of [38], the best fit set of elastic stiffness constants and piezoelectric coefficients used in simulation were determined to be $C_{ij}$(GaN) [48], $C_{ij}$(AlN) [49], $e_{ij}$(GaN) [46], and $e_{ij}$(AlN) [50], as listed in Tables 12.3 and 12.4.

## 12.4 Strain Effects on GaN HEMT Gate Current

### *12.4.1 Background*

The physical breakdown of the AlGaN barrier has been shown to occur at voltages beyond the critical voltage ($V_{crit}$), creating an irreversible increase in $J_G$ [13]. This type of degradation has been hypothesized to be related to the generation of crystallographic defects via the inverse-piezoelectric effect [13]. This creates additional tensile stress in the AlGaN barrier, which adds to the preexisting built-in tensile stress resulting from lattice mismatch between the AlGaN barrier and the GaN layer. It has been suggested that when stress reaches the material critical limit, defects form in order to relax the internal elastic energy [13]. These defects cause a low-resistance leakage path through the AlGaN barrier resulting in a sudden increase in $J_G$. A mechanical wafer bending experiment in literature showed a reduction of $V_{crit}$ with applied tensile mechanical stress; however, only five pairs of devices were investigated in this study, and a more comprehensive study is required [51].

Since stress is inherent in AlGaN/GaN HEMTs, understanding the role of stress on $J_G$ is essential to improve reliability. However, there are discrepancies in the published literature explaining the gate leakage mechanism of unstressed devices. Several leakage mechanisms have been proposed in the literature to explain the GaN HEMT gate current at various temperature and field strength conditions. Zhang et al. [52] noted that direct tunneling of electrons from the metal gate into the GaN bulk dominates the gate current measured at low temperature (<130 K), while the gate current measured at room temperature or above follows the Poole-Frenkel emission (PFE) trend. Mitrofanov et al. [53] also brought up a similar conclusion that PFE dominates the room-temperature gate leakage. Yan et al. [54] proposed a defect band-induced gate leakage mechanism which has similar form as the Poole-Frenkel emission expression. Karmalkar et al. [55] developed a bulk defect-assisted tunneling gate leakage model. In their model, the electrons undergo two thermal-assisted direct tunneling processes: from the gate to the trap and from the trap to the bulk GaN. They were able to reproduce their experimental results on various devices by using different sets of parameters including the defect level, defect density, and Schottky barrier height. Sathaiya et al. [56] proposed a similar two-step thermal-assisted tunneling model to be the dominant leakage mechanism in the GaN HEMT. Despite many models, only a few of them [54] considered the backward current which is essential in order to balance the forward current to properly satisfy the zero net current at zero-bias equilibrium condition. Summarizing the literature findings, the dominant leakage mechanism depends on temperature, gate bias, and device quality.

## 12.4.2 Measurement

### 12.4.2.1 Experiment Setup

Wafer samples were attached to heat-treated high-carbon steel plates with epoxy and stressed in a four-point wafer bending setup. Compressive and tensile uniaxial stress up to 360 MPa was applied longitudinal to the channel direction. The stress dependence of the AlGaN/GaN HEMT $J_G$ was characterized at the $V_{DS} = 0$ state, isolating the effect of electric field induced by the gate. Various reverse biases ($-0.1$ to $-4$ V) were applied to the gate and held constant until $J_G$ reached steady state to eliminate transient trapping effects before each measurement. At each bias, mechanical stress was incrementally applied then released while simultaneously measuring $J_G$. High-temperature measurements were taken on a temperature-controlled probe station using sample stage heaters.

### 12.4.2.2 Experimental Observations

Longitudinal stress, up to 360 MPa, was incrementally applied to the AlGaN/GaN HEMT device, while $J_G$ was simultaneously measured over a period of 1,800 s. Averaging $J_G$ over 1,800 s was done to account for random fluctuations in the measured current. It was found that a time duration of 1,800 s while the stress was held constant was adequate to obtain a reasonable statistical confidence in the measured $J_G$. The normalized change in $J_G$ due to applied stress ($\Delta J_G(\sigma)/J_G(0)$) was measured for several constant applied gate biases ($V_G = -0.1$, $-0.25$, $-0.5$, $-1$, $-2$, and $-4$ V) using the above mentioned procedure. Figure 12.17 shows results of $\Delta J_G(\sigma)/J_G(0)$ for $V_G = -0.25$ and $-4$ V. The stress dependence of $J_G$ was weaker at $V_G = -4$ V than at $V_G = -0.25$ V. Tensile (compressive) stress increased (decreased) $J_G$ for all applied gate biases. At each bias, after the compressive or tensile stress was applied to its maximum magnitude of 360 MPa, then the stress was incrementally released. Upon releasing the stress to zero, $J_G$ returned to its initial, unstressed value. This demonstrates that the measured change in $J_G$ is purely a reversible consequence of the applied stress and not a transient effect. Also, it demonstrates that the applied mechanical stress (up to 360 MPa) does not induce permanent damage to the device.

To quantify the magnitude of the normalized change in gate current density, $\Delta J_G(\sigma)/J_G(0)$, for each applied gate bias and stress level, $J_G$ is averaged over the duration of time the stress was held constant. Figure 12.18 shows $\Delta J_G(\sigma)/J_G(0)$ averaged for all levels of compressive and tensile stress at $V_G = -0.25$ and $-4$ V. Error bars representing the uncertainty in the measurement of $\Delta J_G(\sigma)/J_G(0)$ are three times the standard deviation of the $J_G$ measurements over the duration when the stress was held constant.

The sensitivity of $J_G$ to stress defined as the normalized change in $J_G$ per stress, $[\Delta J_G(\sigma)/J_G(0)]/\sigma$, is calculated from the slope of the weighted total least squares

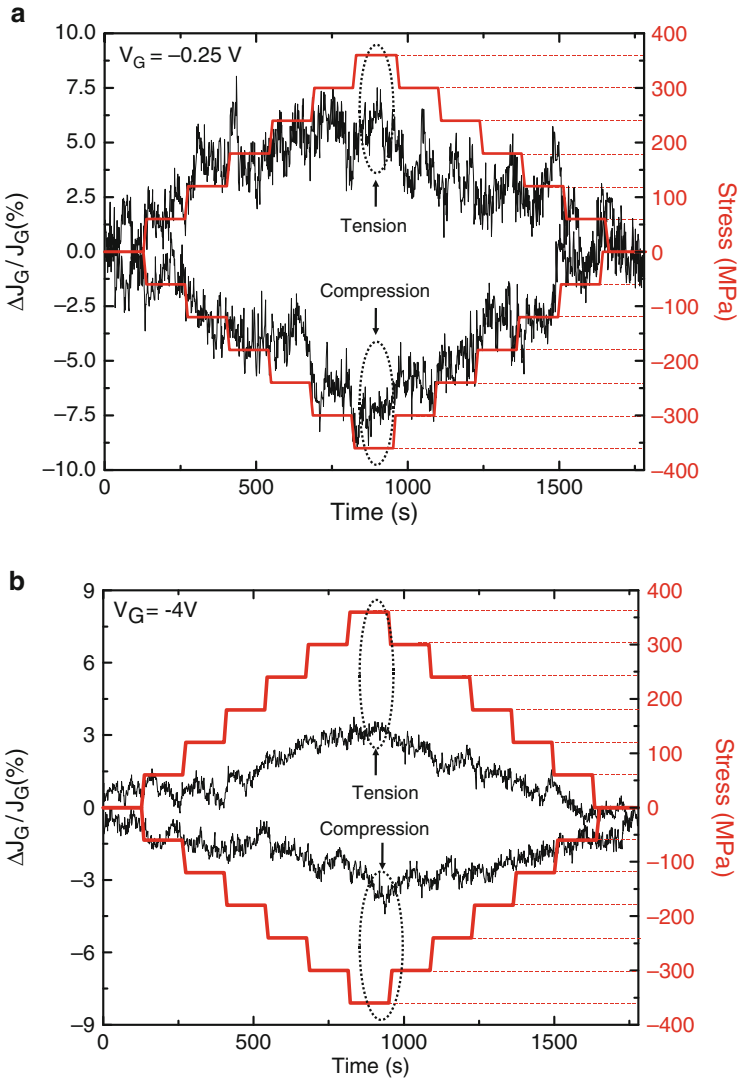



**Fig. 12.17** Normalized change in $J_G$ for incrementally increasing and decreasing uniaxial stress for (**a**) $V_G = -0.25$ V and (**b**) $V_G = -4$ V

linear fit of $\Delta J_G(\sigma)/J_G(0)$ versus stress, including uncertainty in $J_G$ at each stress increment. The stress sensitivity of $J_G$ is plotted as a function of reverse gate bias in Fig. 12.19. Increasing the reverse gate bias is observed to decrease the sensitivity of $J_G$ to stress. The sensitivity of $J_G$ to stress decreased from 1.7 $\pm$0.3%/100 MPa at $V_G = -0.25$ V to 0.6 $\pm$0.1%/100 MPa at $V_G = -4$ V as shown in Fig. 12.19. To interpret the decreasing sensitivity for increasing reverse gate bias, the dominant gate leakage transport mechanism needs to be analyzed.

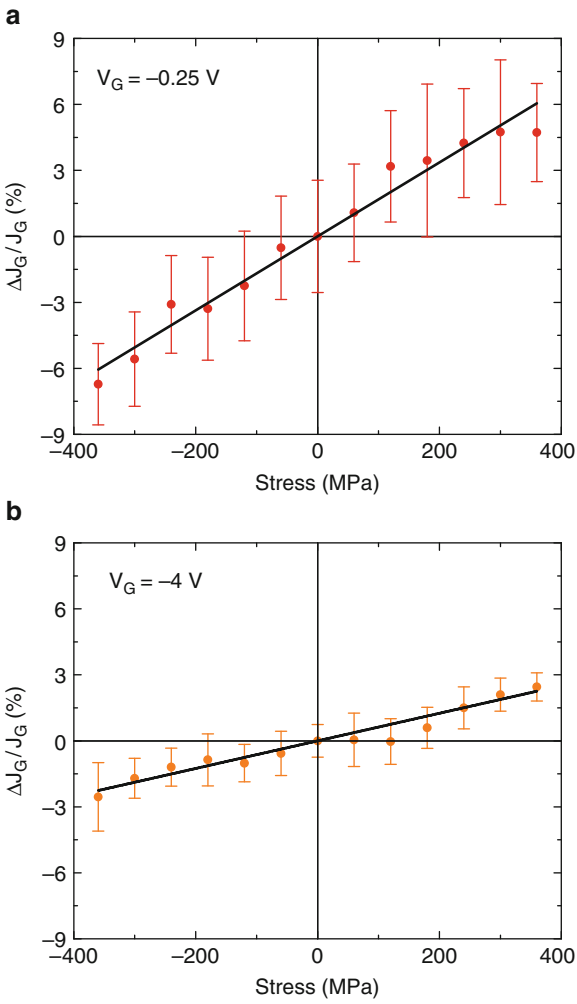



**Fig. 12.18** $\Delta J_G(\sigma)/J_G(0)$ averaged for all levels of compressive and tensile stress at (**a**) $V_G = -0.25$ V and (**b**) $-4$ V. Uncertainty comes from three standard deviation from the measurement of $\Delta J_G(\sigma)/J_G(0)$ over the duration stress was held constant

### *12.4.3 Determining the Vertical Electric Field Under the Gate*

The dominant leakage mechanism is strongly dependent on the materials and processing conditions and the electric field in the AlGaN Barrier, $E_{AlGaN}$. To characterize the effect of mechanical stress on AlGaN/GaN HEMT devices, an accurate model for the leakage mechanism is necessary. To compare different leakage models to the experimental measurements, an accurate calculation of $E_{AlGaN}$ is needed. In past works, $E_{AlGaN}$ has been simplified to a linear relationship with the gate voltage [53, 57], simulated using Medici 2D simulations [58], and experimentally measured [54]. The $V_{DS} = 0$ state is of particular interest for exploring reliability since both the source and drain sides of the device gate are electrically stressed simultaneously.

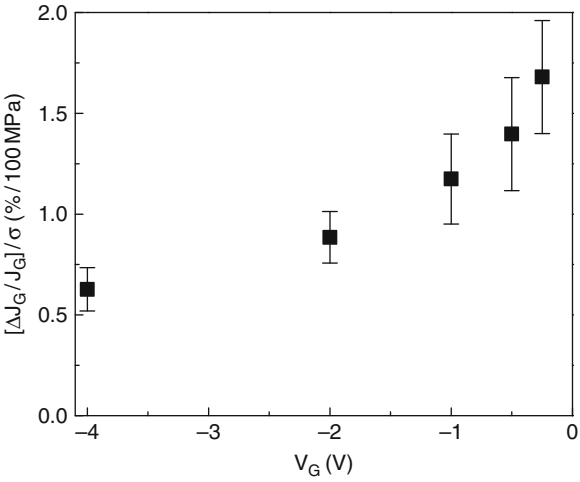



**Fig. 12.19** Experimentally measured stress sensitivity of $J_G$ per 100 MPa of stress, as a function of reverse gate bias

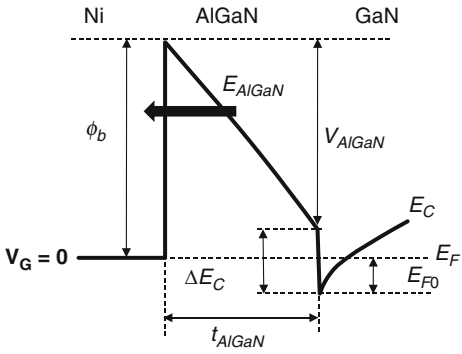



**Fig. 12.20** Energy band diagram schematic showing Ni/AlGaN/GaN interface

### 12.4.3.1 1D Electric Field Model

A one-dimensional (1D) calculation of $E_{\mathrm{AlGaN}}$ provides insight into the general relationship between voltage and field. In this condition, the gate is assumed to be infinitely wide. Also, for simplicity, $E_{\mathrm{AlGaN}}$ is assumed to be a constant throughout the entire thickness of the AlGaN barrier. Based on these assumptions, a simple expression for the electric field in the AlGaN barrier can be derived from the voltage drop across the AlGaN barrier ($V_{\mathrm{AlGaN}}$) from inspection of the energy band diagram (Fig. 12.20). $E_{\mathrm{AlGaN}}$ can be written as

$$E_{\mathrm{AlGaN}} = \frac{V_{\mathrm{AlGaN}}}{t_{\mathrm{AlGaN}}} = \frac{\phi_{\mathrm{b}} - (\Delta E_{\mathrm{C}} - E_{\mathrm{F}})}{q t_{\mathrm{AlGaN}}} \tag{12.11}$$

Recalling Eq. 12.5, $n_s$ can be rewritten in terms of $E_{AlGaN}$

$$n_s = \frac{\sigma_{int}}{q} - \left(\frac{\varepsilon_0 \varepsilon_r}{q^2 t_{AlGaN}}\right)[\phi_b + E_F - \Delta E_C] = \frac{\sigma_{int} - \varepsilon_0 \varepsilon_r E_{AlGaN}}{q} \tag{12.12}$$

Then, the simple expression for $E_{AlGaN}$ is

$$E_{AlGaN} = \frac{\sigma_{int} - q n_s}{\varepsilon_0 \varepsilon_r} \tag{12.13}$$

In the expression for $E_{AlGaN}$ (12.13), only the 2DEG sheet carrier density ($n_s$) is a function of gate voltage. The total fixed charge density at the AlGaN/GaN interface ($\sigma_{int}$) is assumed to be independent of bias because it is based on the polarization. Charge trapped at the AlGaN/GaN interface is estimated to be constant with bias. To analyze the dependence of $n_s$ with gate voltage, $n_s$ is rewritten to include the contribution of $V_G$ in terms of the equilibrium 2DEG concentration $n_{s0}$

$$n_s = \frac{\sigma_{int}}{q} - \left(\frac{\varepsilon_0 \varepsilon_r}{q^2 t_{AlGaN}}\right)[\phi_b + E_{F0} - \Delta E_C + q V_G] = n_{s0} + \left(\frac{C_{AlGaN}}{q}\right) V_G \tag{12.14}$$

where $C_{AlGaN} = \varepsilon_0 \varepsilon_r / t_{AlGaN}$ is the AlGaN capacitance per unit area. At equilibrium, $V_G = 0$ V and $n_s = n_{s0}$, which is the maximum induced 2DEG sheet carrier density. The fixed positive charge at the AlGaN/GaN interface induces accumulation of electrons at the interface in the GaN. The electrons that form the 2DEG originate from the AlGaN surface [59], not from the source, drain, and substrate as in a Si MOSFET inversion layer. A negative bias applied to the gate depletes the 2DEG, decreasing $n_s$ linearly, as shown by Eq. 12.14. The threshold voltage for the depletion mode AlGaN/GaN HEMT is defined as the voltage required on the gate to entirely deplete the 2DEG ($n_s = 0$). Since the AlGaN/GaN HEMT is a depletion mode device with a negative $V_T$, the device is considered to be turned off below threshold ($|V_G| > |V_T|$). Above threshold, the 2DEG is formed ($|V_G| < |V_T|$). Below threshold, the 2DEG remains depleted. Since the intrinsic carrier concentration of GaN is extremely low ($n_i \sim 1 \times 10^{-10}$ cm$^{-3}$ at 300 K), hole accumulation at the surface is negligible. Figure 12.21 shows for an ideal device, without interface-trapped charge, the 2DEG sheet carrier density plotted versus $V_G$. The fixed charge density at the AlGaN/GaN interface ($\sigma_{int} = 2.2 \times 10^{-6}$ C/cm$^2$) is a result of the polarization differences between AlGaN and GaN. Parameters given in Table 12.5 were used for the calculation.

To express $E_{AlGaN}$ as a function of gate voltage, the voltage-dependent 2DEG equation (Eq. 12.14) was incorporated into the expression for $E_{AlGaN}$ (Eq. 12.13) to give

$$E_{AlGaN}(V_G) = \frac{-\left(\frac{\varepsilon_0 \varepsilon_r}{q t_{AlGaN}}\right)[q\phi_b + E_{F0} - \Delta E_C + q V_G]}{\varepsilon_0 \varepsilon_r} \qquad for\ V_G > V_T \tag{12.15}$$

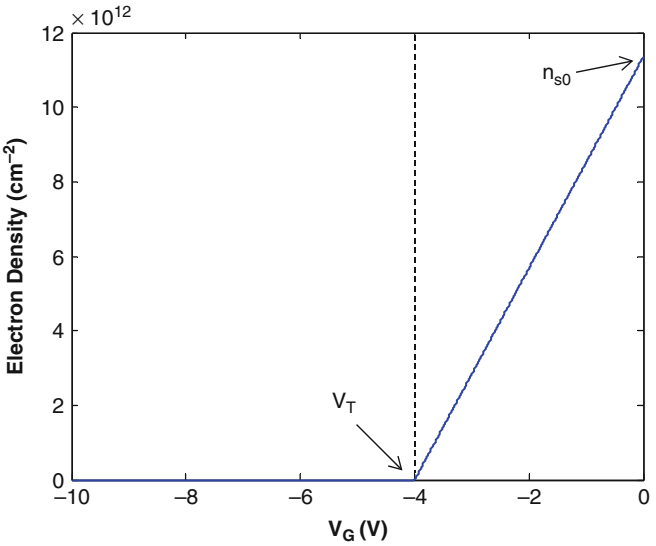



**Fig. 12.21** Dependence of the 2DEG density ($n_s$) with gate bias for the 1D case, with no interface-trapped charge. Threshold is defined when $n_s$ is entirely depleted ($n_s = 0$)

**Table 12.5** List of parameters used in 1D $E_{AlGaN}$ calculation

| Parameter | Symbol | Value | Units |
|---|---|---|---|
| Temperature | $T$ | 300 | K |
| Intrinsic carrier concentration | $n_i$ | $10^{-10}$ | $cm^{-3}$ |
| AlGaN thickness | $t_{AlGaN}$ | 18 | nm |
| Permittivity of GaN | $E_{GaN}$ | 9.5 | – |
| Permittivity of AlN | $E_{AlN}$ | 8.5 | – |
| Bandgap of GaN | $E_G$,GaN | 3.4 | eV |
| Bandgap of AlN | $E_G$,AlN | 6.13 | eV |
| Electron affinity of GaN | $\chi_{GaN}$ | 4.1 | eV |
| Ni work function | $\phi_m$ | 5.27 | eV |
| Mass of GaN | m*$G_a$N | $0.2m_0$ | – |
| Mass of AlN | m*$A_l$N | $0.48m_0$ | – |
| Fixed charge at AlGaN/GaN interface | $\sigma_{int}$ | $2.2 \times 10^{-6}$ | $C/cm^2$ |

For gate bias below threshold, $E_{AlGaN}$ saturates at $E_{AlGaN} = -\sigma_{int}/(\varepsilon_0\varepsilon_r)$. Figure 12.22 shows $E_{AlGaN}$ increasing linearly with bias for gate biases above threshold.

Estimation of $E_{AlGaN}$ for an actual AlGaN/GaN HEMT device based on the 1D model requires both $\sigma_{int}$ and $n_s$ to be experimentally measured. The threshold

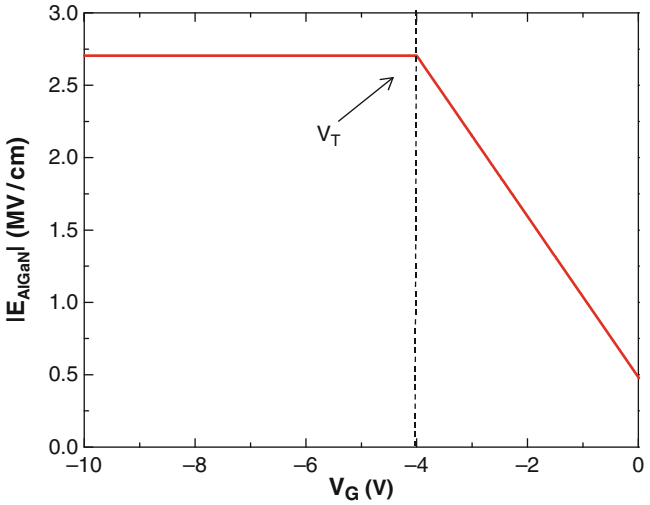



**Fig. 12.22** Ideal 1D calculation of $E_{\mathrm{AlGaN}}$, assuming no interface-trapped charge. $|E_{\mathrm{AlGaN}}|$ increases linearly until the threshold voltage, then saturates

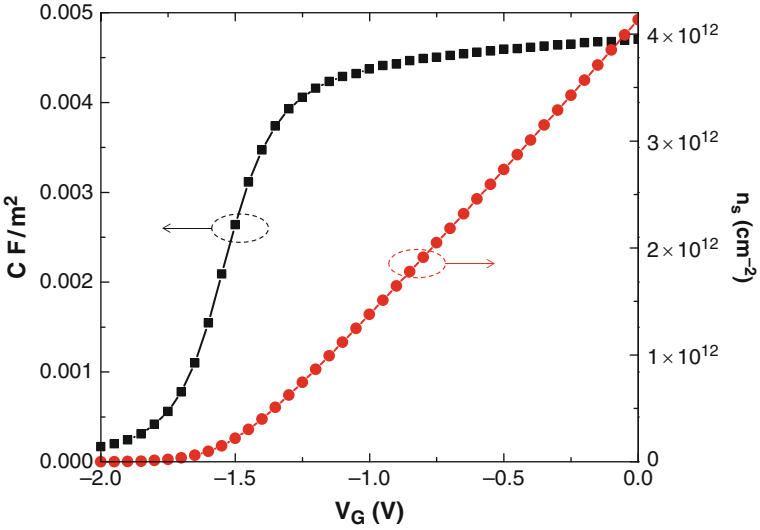



**Fig. 12.23** Capacitance-voltage measurement, which is integrated in order to determine $n_{\mathrm{s}}(V)$

voltage determined by the standard linear extrapolation method ($V_{\mathrm{T}} = -4$ V) is not consistent with the definition of $V_{\mathrm{G}} = V_{\mathrm{T}}$ when $n_{\mathrm{s}} = 0$. A more accurate threshold of $-1.9$ V is obtained based on the initial increase of the capacitance-voltage curve (Fig. 12.23). From Eq. 12.14 and the definition of threshold ($V_{\mathrm{G}} = V_{\mathrm{T}}$ when

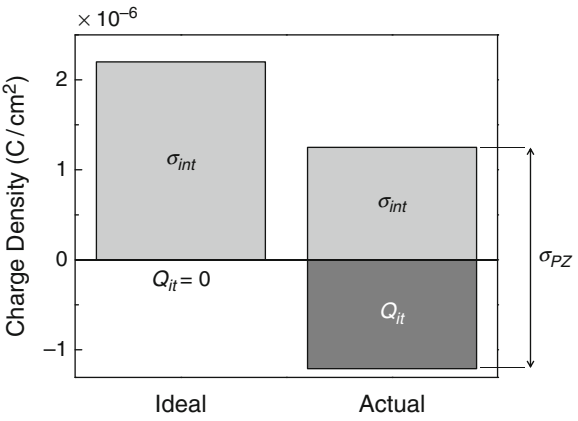



**Fig. 12.24** Bar diagram of the AlGaN/GaN interface charge for an ideal device with no trapped charge and an actual device. Trapped charge reduces the positive fixed sheet charge density at the AlGaN/GaN interface

$n_s = 0$), the 2DEG sheet carrier density at $V_G = 0$ can be calculated from the measured $V_T$

$$n_{s0} = \frac{-C_{AlGaN} V_T}{q} \tag{12.16}$$

which gives $n_{s0} = 5.3 \times 10^{12}$ cm$^{-3}$ for $V_T = -1.9$ V. Then, the fixed charge at the AlGaN/GaN interface can be estimated by solving for $\sigma_{int}$ in Eq. 12.17

$$n_{s0} = \frac{\sigma_{int}}{q} - \left(\frac{\varepsilon_0 \varepsilon_r}{q^2 t_{AlGaN}}\right)[\phi_b + E_{F0} - \Delta E_C] \tag{12.17}$$

which results in $\sigma_{int} = 1.25 \times 10^{-6}$ C/cm$^2$ that is nearly half the ideal value ($\sigma_{int} = 2.2\ 10^{-6}$ C/cm$^2$) using the constants in Table 12.5. The interface-trapped charge density is estimated from the subthreshold slope (SS) measurements [60, 61]

$$SS \approx \frac{kT}{q} \ln 10 \left[1 + \frac{C_{it}}{C_{AlGaN}}\right] \tag{12.18}$$

The capacitance associated with the interface-trapped charge ($C_{it}$) can be related to the density of interface traps by $C_{it} = qD_{it}$. For the measured device with SS = 100 mV/dec, the interface trap density corresponds to $D_{it} = 1.92 \times 10^{12}$ cm$^{-2}$ eV$^{-1}$. Assuming all traps are full, integrating over the bandgap of AlGaN gives an interface-trapped charge of $Q_{it} = 1.2 \times 10^{-6}$ C/cm$^2$. Figure 12.24 shows a bar chart of the charge density in the ideal device compared to the actual device. In the ideal device, $Q_{it} = 0$ and $\sigma_{pz} = \sigma_{int}$. However, in the actual device, the fixed charge at the AlGaN/GaN interface induced by polarization is reduced by negative trapped charge.

To obtain $E_{AlGaN}$ for the actual device, the $n_s$ versus $V_G$ relationship needs to be obtained. A high-frequency (1 MHz) capacitance-voltage curve (Fig. 12.23) was

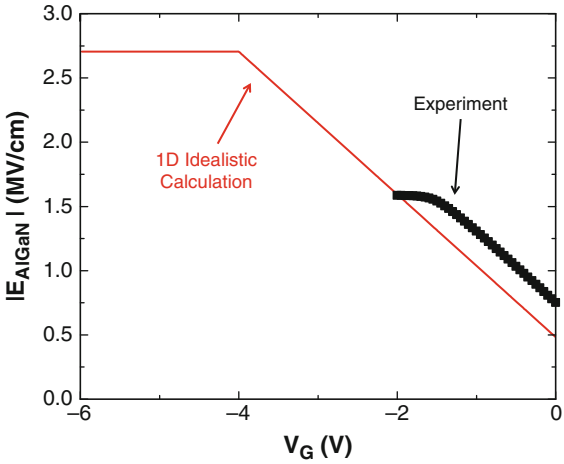



**Fig. 12.25** 1D ideal (no interface-trapped charge) calculation and the experimental result from experimental parameters obtained by adjusting $\sigma_{int}$ and obtaining ns from C-V

integrated from pinch-off to voltage $V_G$. The experimental bias range was limited to $V_G = -2$ to 0 V to avoid additional charge trapping from larger applied biases. The value of $n_s$ ($V_G$) is also shown in Fig. 12.23. The experimentally obtained $n_s$ versus $V_G$ relationship is used in Eq. 4.13 to calculate the experimentally obtained $E_{AlGaN}$ versus $V_G$ relationship.

Implementing the experimentally determined $\sigma_{int}$ and $n_s$ versus $V_G$ relationship provides the experimentally determined 1D calculation of $E_{AlGaN}$ versus $V_G$. Figure 12.25 shows a comparison between the experimental and ideal 1D calculation for $E_{AlGaN}$ versus $V_G$. Since both methods are based on the 1D expression, they have similar trends. Above threshold, $|E_{AlGaN}|$ increases linearly with increasing reverse bias, and below threshold $E_{AlGaN}$ saturates at $E_{AlGaN} = -\sigma_{int}/(\varepsilon_0\varepsilon_r)$. The presence of $Q_{it}$ reduces $\sigma_{int}$ and the magnitude of the saturated value of $E_{AlGaN}$ below threshold in the actual device compared to the ideal 1D calculation. In addition, $V_T$ is shifted toward the positive direction by the negative $Q_{it}$. At $V_G = 0$, the experimental $|E_{AlGaN}|$ is larger than the ideal 1D case because the experimentally extracted $n_{s0}$ is lower than what is theoretically predicted, making the numerator of Eq. 12.13 larger.

#### 12.4.3.2 2D TCAD Simulation

The AlGaN/GaN HEMT structure was simulated using the Sentaurus device simulator. A two-dimensional (2D) simulation was performed to gain a physical understanding of 2D effects on $E_{AlGaN}$. The thin (~1.5 nm) GaN cap layer was neglected in the simulation for simplicity. The device structure is quantized into a mesh or grid of discrete elements. A rectangular gate is assumed for simplicity. The grid was condensed in areas where large variations of carrier concentration are expected at short distances to optimize computational accuracy (Fig. 12.26).

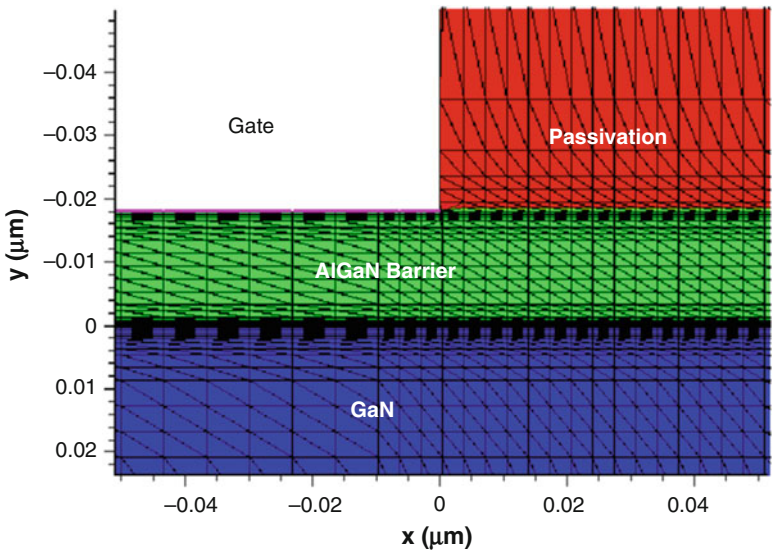



**Fig. 12.26** Optimized grid for Sentaurus simulation of the AlGaN/GaN HEMT device

At each point on the grid, three variables are solved for simultaneously. The electrostatic potential, electron concentration, and hole concentration are solved for through Poisson's equation and the electron and hole continuity equations, respectively. The boundary conditions used to solve for the three variables are the metal-contact work function difference, the ohmic contacts at the source and drain, and local conservation of charge. In the actual device, the high-resistivity Si layer isolates the GaN layer from the back side of the wafer, leaving it electrically floating. The bulk is left floating in simulation to model this. Shockley-Read-Hall recombination and generation were included in the calculation. Also, a hydrodynamic model was used where the carrier temperature is not assumed to be equal to the lattice temperature. Electron mobility was modeled including doping dependence, high field saturation, and temperature. Doping was introduced under the source and drain contacts to emulate metal spikes to provide an ohmic contact to the 2DEG [62]. The piezoelectric effect is modeled by incorporating a fixed charge at the AlGaN/GaN interface equal to $\sigma_{\text{int}}$ and at the AlGaN surface equal to $\sigma_{\text{PZ}}$ to match experiment. The parameters used in the simulation are given in Table 12.5.

The simulated dependence of $E_{\text{AlGaN}}$ versus $V_{\text{G}}$ at the center of the gate is compared to the 1D model with experimentally determined parameters (Fig. 12.27). In the simulation, $\sigma_{\text{int}}$ was matched to the experimentally obtained value of $1.25 \times 10^{-6}$ C/cm$^2$. The trend of the 2D simulated $E_{\text{AlGaN}}$ is similar to the 1D calculation. Above threshold, $|E_{\text{AlGaN}}|$ increases until threshold with decreasing reverse gate bias. Below threshold, $|E_{\text{AlGaN}}|$ tends to saturate. Although $\sigma_{\text{int}}$ is

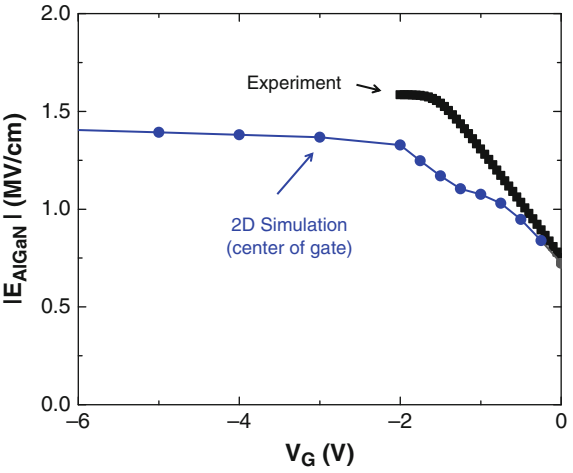



**Fig. 12.27** Experimental calculation of $E_{AlGaN}$ versus $V_G$ compared to 2D simulation results

matched, the saturation value of $E_{AlGaN}$ in the 2D simulation at the center of the gate is slightly lower than the experimentally determined curve. Since the simulated device has a finite gate width (1 μm), it is important to analyze the edges of the gate for a comprehensive understanding of the $E_{AlGaN}$ versus $V_G$ relationship.

Vertical cross-section plots of $E_{AlGaN}$ through the middle of the gate and drain edge of the gate are shown in Fig. 12.28. The AlGaN/GaN interface is referenced at $y = 0$ μm and the AlGaN surface is at $y = -0.018$ μm. In the middle of the gate, $E_{AlGaN}$ remains essentially constant throughout the depth of the AlGaN layer for $V_G = -2$ to $-10$ V. As shown in Fig. 12.22, the magnitude of $E_{AlGaN}$ increases more above $V_T$ than below $V_T$ as a function of $V_G$, demonstrating the saturation of $E_{AlGaN}$ below $V_T$. However, at the gate edge, $E_{AlGaN}$ varies with depth in the AlGaN barrier, and $E_{AlGaN}$ continues to increase with increasing reverse gate bias.

A horizontal cross section 1 nm below the gate contact in the AlGaN barrier shows a large increase in $E_{AlGaN}$ at the gate edges (Fig. 12.29). At $V_G = 0$ V, $|E_{AlGaN}|$ is lower directly under the Schottky gate than under the passivation layer. This is a result of the negative polarization charge at the top surface of the AlGaN. Under the nitride, the surface polarization exists; however, the gate has control of the potential under the gate electrode.

To explain the large increase in $E_{AlGaN}$ at the gate edges, a plot of the potential at the GaN surface, 0.5 nm below the AlGaN/GaN interface, is shown (Fig. 12.30). At large reverse gate biases, there is a large horizontal potential drop in the GaN surface near the gate edges. This potential drop results in a large horizontal electric field at the gate edge. The horizontal field at the gate edge is much larger than in the center of the gate (Fig. 12.31). The contribution of the horizontal electric field at the gate edge adds to the vertical electric field increasing the magnitude of $E_{AlGaN}$ increasing at the gate edges. It should be noted that a different gate structure such as a T-gate will result in different edge effects including a smaller increase in $|E_{AlGaN}|$ with increasing reverse gate bias.

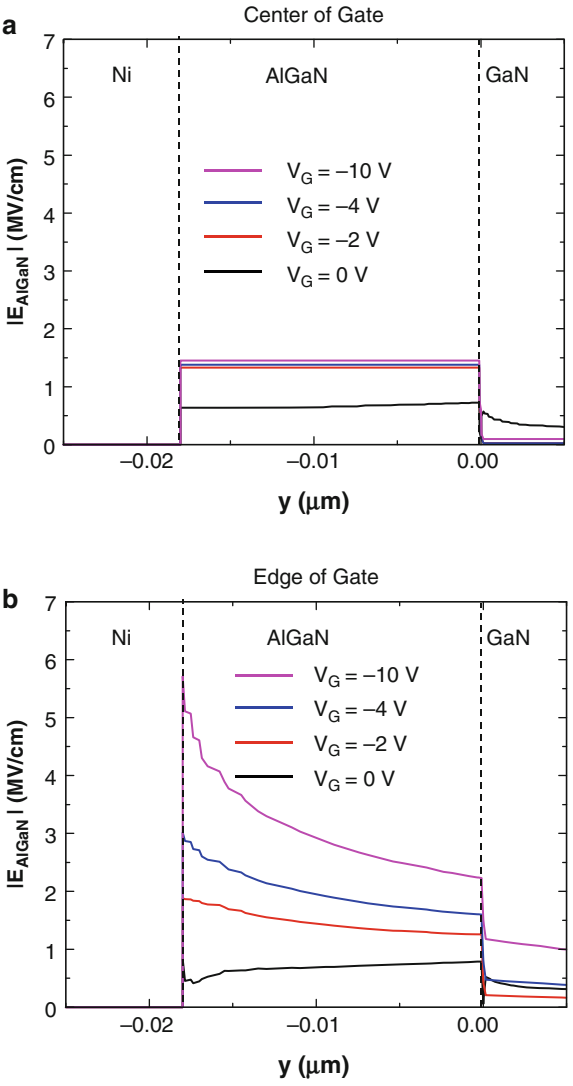



**Fig. 12.28** Vertical cross section of electric field of the AlGaN/GaN HEMT at the (**a**) *center* of the gate and (**b**) drain edge of the gate

## *12.4.4 Gate Leakage Mechanisms*

Several GaN HEMT gate leakage mechanisms have been proposed in the literature, such as the direct tunneling [52], Poole-Frenkel emission (PFE) [53, 54], bulk defect-assisted tunneling [55, 56], and defect band-induced leakage [54]. Despite many models, the dominant leakage mechanism depends on temperature, gate bias, and device quality. Figure 12.32 schematically illustrates these leakage processes.

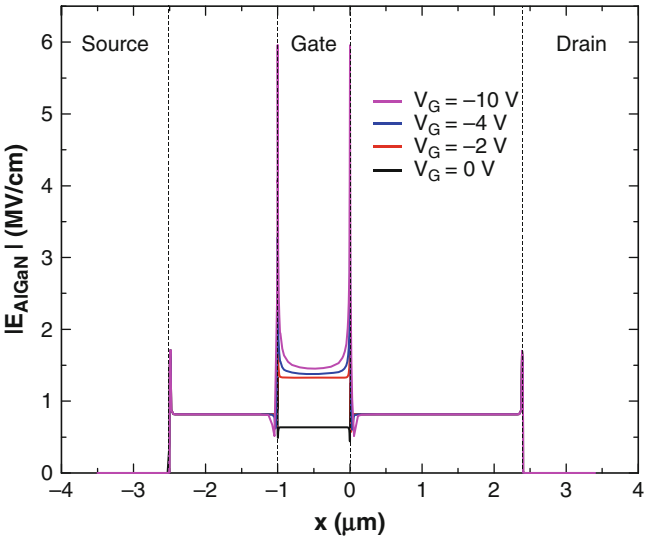



**Fig. 12.29** Horizontal cross section of $E_{AlGaN}$ taken near the top surface of the AlGaN barrier 1 nm below the gate contact

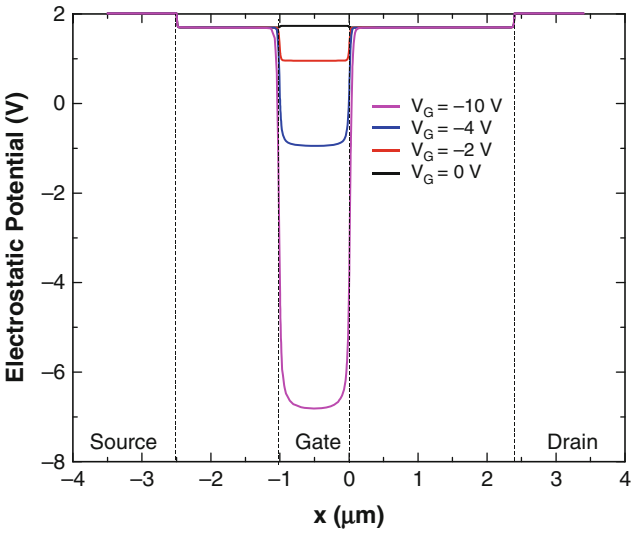



**Fig. 12.30** Horizontal cross section of the electrostatic potential taken near the *top* surface of the GaN 0.5 nm below the AlGaN/GaN interface

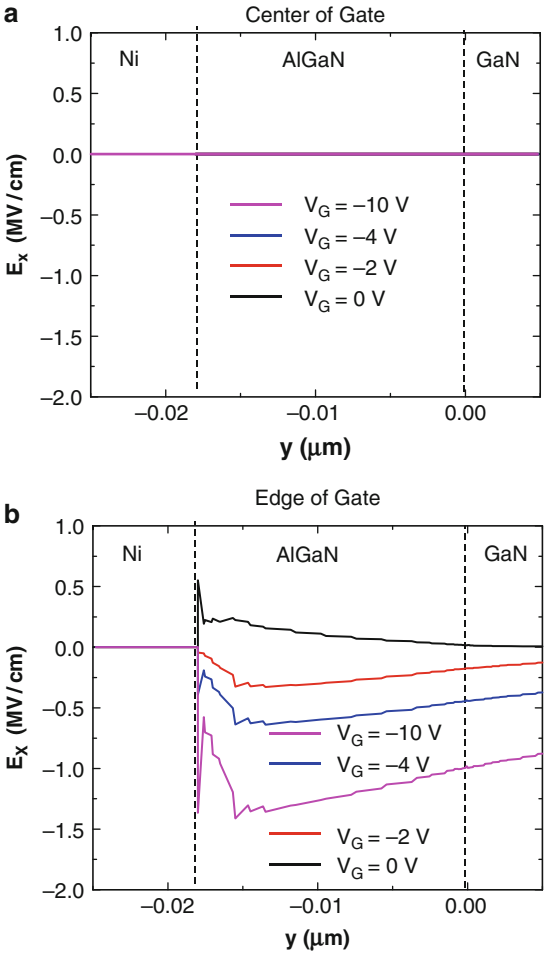



**Fig. 12.31** Horizontal component of $E_{AlGaN}$ at the (**a**) *center* and (**b**) *edge* of the gate

#### 12.4.4.1 Direct Tunneling

Due to the large electric field across the AlGaN barrier, the direct tunneling in GaN HEMT mainly includes the Fowler-Nordheim tunneling (FN) and thermionic field emission (TFE). Both the FN and TFE models are well established in literature and can be expressed analytically as Eq. 12.19 for the FN tunneling [63] and Eqs. 12.20, 12.21, and 12.22 for the TFE [55], respectively.

$$J = \frac{q^2(m_e/m^*)}{8\pi h\phi_B} E^2 \exp\left(-\frac{8\pi\sqrt{2m^*(q\phi_B)^3}}{3qhE}\right) \tag{12.19}$$

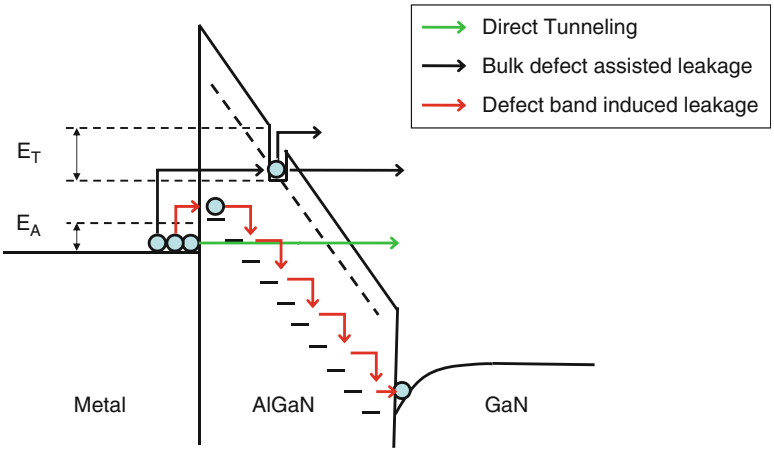



**Fig. 12.32** Schematic illustrations of the gate leakage processes in a reverse-biased GaN HEMT

$$J = \frac{qA^*T}{k} \int_0^{\phi_B} f_{FD}(\phi) P(\phi)\, d\phi \tag{12.20}$$

$$f_{FD} = \frac{1}{1 + \exp\left(\frac{\phi_B - \phi}{kT/q}\right)} \tag{12.21}$$

$$P = \exp\left(-\frac{8\pi\sqrt{2m^*q}}{3hE}\ \phi^{3/2}\right) \tag{12.22}$$

Here, $q$, $m_e$, $h$, $k$, and $A^*$ are the electron charge, the free electron mass, the Planck constant, the Boltzmann constant, and the effective Richardson constant, respectively. $E$ is the electric field strength in the AlGaN layer. $m^*$ is the out-of-plane effective mass of electrons inside AlGaN. $f_{FD}$ is the Fermi-Dirac distribution function of electrons on the metal side. $\phi_B$ is the Schottky barrier height between the metal gate and the AlGaN, obtained from Eq. 12.23 considering the barrier lowering due to image force and bandgap narrowing [55]. The values of $\phi_{B0}$, $\gamma_1$, and $\gamma_T$ are 1.4 eV, 0.4, and 2.4 $\times$ $10^{-4}$ V/k for AlGaN [55].

$$\phi_B = \phi_{B0} - \gamma_1 \sqrt{\frac{q}{4\pi\varepsilon}}\sqrt{E} - \gamma_T T \tag{12.23}$$

Figure 12.33 shows the simulated direct tunneling current and the experimental measured gate current at room temperature. The result shows that the direct tunneling currents are several orders smaller than the experimental observation until at least 2.5 MV/cm. This is expected since the AlGaN layer (18 nm) is too thick for the

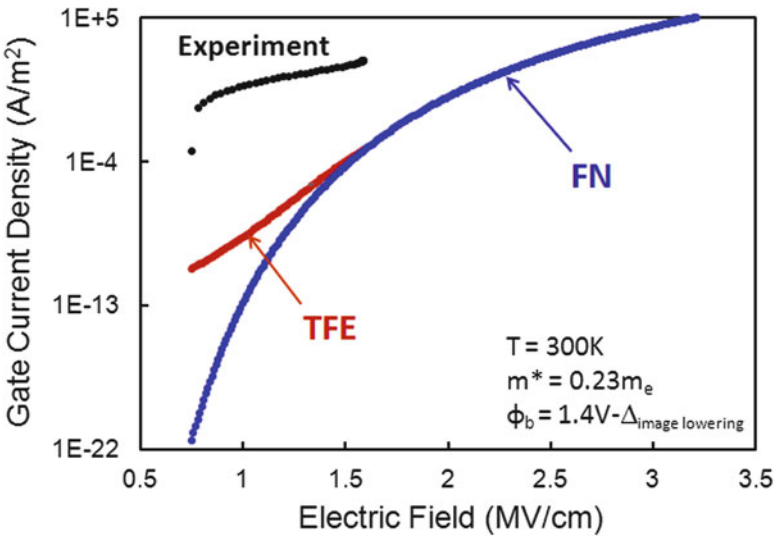



**Fig. 12.33** Simulation results of the FN tunneling current and the TFE current at room temperature

electrons to have a high tunneling rate. In addition, the Schottky barrier height (~1.4 eV) is unlikely to result in a high thermal emission rate at room temperature.

As a result, direct tunneling is not the dominant leakage mechanism in these devices when biased above the threshold ($V_G > -1.9$ V). At below threshold, there can be a direct tunneling path at the edge of the gate once the edge electric field is high enough.

### 12.4.4.2 Bulk Trap-Assisted Leakage

To consider the possibility of GaN HEMT gate leakage following the bulk trap-assisted leakage mechanism, this work adopts Fleischer's two-step leakage model [64] and develops it to incorporate the thermal contribution in step 2. In the first step of our model, the electrons tunnel from the metal gate to the trap level through direct tunneling or thermal-assisted direct tunneling. In step 2, electrons escape from the trap into bulk GaN through direct tunneling, Poole-Frenkel emission [65], or phonon-assisted tunneling [66, 67]. The complete leakage process is shown schematically in Fig. 12.32. The tunneling (emission) rates of steps 1 and 2 are calculated from Eqs. 4.14)and 4.15.

$$R_1 = C_t N_t f_{FD}(1-f)P_1 \tag{12.24}$$

$$R_2 = C_t N_t f P_2 \tag{12.25}$$

Here, $N_t$ is the bulk defect density, $f$ is the probability that the defect level is filled with an electron, $P_1$ and $P_2$ are the tunneling (emission) probability, and $C_t$ is given by

$$C_t = \left(\frac{m^*}{m_e}\right)^{5/2} \left(\frac{8E_1^{3/2}}{3\hbar\sqrt{E_T - E_1}}\right) \tag{12.26}$$

where $E_T$ is the defect level referring to the conduction band edge, and $E_1$ is the total energy of an electron (0.2 eV [64]). $P_1$ is calculated following the standard direct tunneling expression

$$J = \frac{q^2(m_e/m^*)}{8\pi h\phi_B} E^2 \exp\left(-\frac{8\pi\sqrt{2m^*q}}{3hE}\left(\phi_B^{3/2} - \phi^{3/2}\right)\right) \tag{12.27}$$

and $P_2$ is calculated using either the FN tunneling model, the classical Poole-Frenkel emission model, or the phonon-assisted tunneling model developed by Pipinys et al. [66]. At steady state, $R_1$ should equal $R_2$. After a straightforward mathematical derivation, the tunneling rate $R$ is obtained.

$$R = N_t \cdot \left(\frac{1}{C_t f_{FD} P_1} + \frac{1}{C_t P_2}\right)^{-1} \tag{12.28}$$

The overall gate current density is then calculated by integrating $R$ over the thickness of the entire AlGaN barrier.

$$J = \frac{q}{E} \int_{\phi_t}^{E \cdot d} R(\phi) d\phi \tag{12.29}$$

In the current model, a uniform defect distribution is assumed across the AlGaN barrier. The AlGaN potential profile is assumed to be triangular, which is reasonable for devices biased above threshold. We also considered only one single effective defect level as a fitting parameter. Figure 12.34 shows the calculated and experimental gate current density at room temperature, with $N_t = 5 \times 10^{17}/\text{cm}^3$ and $E_T = 0.82\,\text{eV}$. The value of $E_T$ significantly affects the resulting current magnitude. In fact, $E_T$ has opposite effects on $P_1$ and $P_2$. With small $E_T$, the energy difference between the metal gate Fermi level and the defect level is large, and the first step of electrons tunneling from the metal gate into the defect level is unlikely to happen. With large $E_T$, the energy difference between the defect level and the AlGaN conduction band is large, and the step two probability of electrons undergoing PFE or PAT becomes smaller. The value of $E_T = 0.82$ eV was estimated by trying to fit the two-step model and the

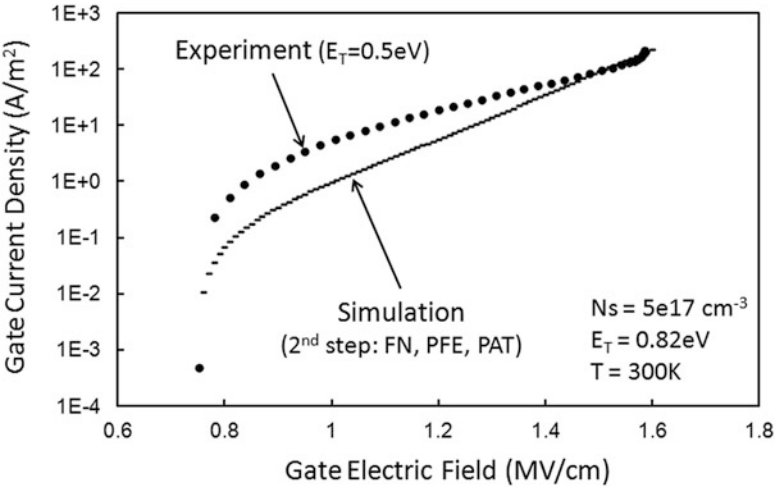



**Fig. 12.34** Simulation result of the two-step bulk trap-assisted leakage current at room temperature

experimental results within the same order of magnitude, under a reasonable defect density $N_t = 5 \times 10^{17}/\text{cm}^3$. Under this condition, it is found that the calculated current density does not depend on the chosen mechanism for the $P_2$ step. This indicates that the first step of electrons tunneling from the metal gate into the defect level limits the bulk trap-assisted leakage current.

Regardless of what leakage mechanism for step two and fitting parameters are chosen, a good match between the model and experimental results is unachievable. This suggests that the bulk trap-assisted tunneling mechanism may not be the dominant leakage mechanism in these GaN HEMT devices.

#### 12.4.4.3 Defect Band-Induced Gate Leakage

It has been proposed in literature that defects with a large dispersion in energy exist throughout the AlGaN barrier layer [68–73]. These defects can form a conductive path which has an effective energy $E_A$ above the metal gate Fermi level, as shown in Fig. 12.32. At room temperature or at elevated temperature, some electrons from the metal gate may gain sufficient thermal energy to overcome the barrier between the metal gate and the defect band and emit into the defect band states. Once the electrons reach the defect band, they may undergo trap-to-trap transitions toward the GaN substrate by high vertical electric field-induced elastic and inelastic trap-to-trap tunneling due to the high density of traps.

The accurate modeling of the defect band induced gate leakage is complex. It highly depends on the defect distribution in both energy and real space. Qualitatively speaking, we expect this leakage current to increase with increasing vertical electric field, larger defect density, lower gate-to-defect barrier height ($E_A$), higher

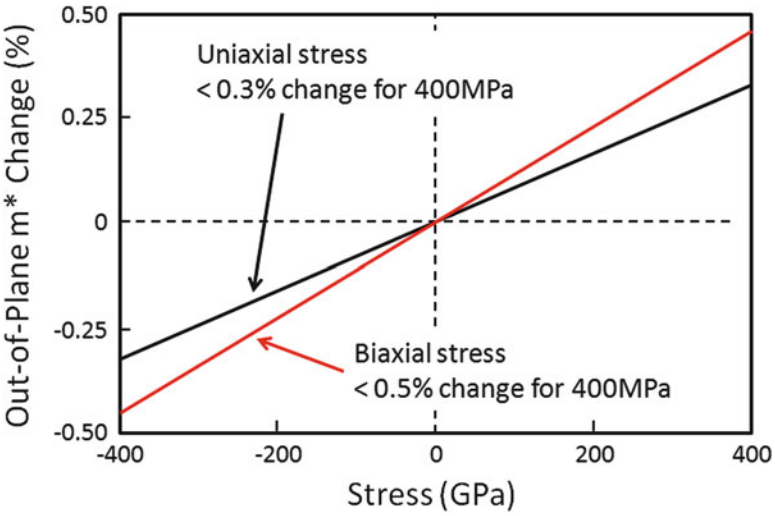



**Fig. 12.35** Change of the electron out-of-plane effective mass under uniaxial and biaxial stress

temperature, and smaller electron out-of-plane effective mass. This mechanism may explain our experimental results since the energy barrier is reduced compared to the bulk trap-assisted, Poole-Frenkel, and direct tunneling leakage mechanisms. The details of model development for the defect band induced gate leakage require further investigation.

### *12.4.5 Stress-Dependent Parameters*

This section qualitatively discusses the effects of external mechanical stress on the GaN HEMT gate leakage, by considering the stress-altered electron out-of-plane effective mass, the AlGaN electric field, and the defect energy level.

We used the $sp^3d^5$-$sp^3$ tight-binding model to calculate the change of electron out-of-plane effective mass under both uniaxial and biaxial stresses. In the experiment, the uniaxial stress is induced by wafer bending, and the biaxial stress is induced by the inverse-piezoelectric effect when gate bias is present. Figure 12.35 shows the tight-binding calculation results for both uniaxial and biaxial stresses. It is observed that the electron out-of-plane effective mass has negligible change under both stresses (<0.5% per 400 MPa).

Stress causes variation in the AlGaN polarization due to the piezoelectric effect, leading to a change in the electric field strength. As mentioned before, this additional piezoelectric polarization can be calculated by $P^{\mathrm{PE,mech}} = e_{ij} \cdot \varepsilon$, where $e_{ij}$ is the piezoelectric constants, and $\varepsilon$is the strain. The simulation results for both uniaxial and biaxial stresses are shown in Fig. 12.36. Approximately 0.7 and 1.4% change in

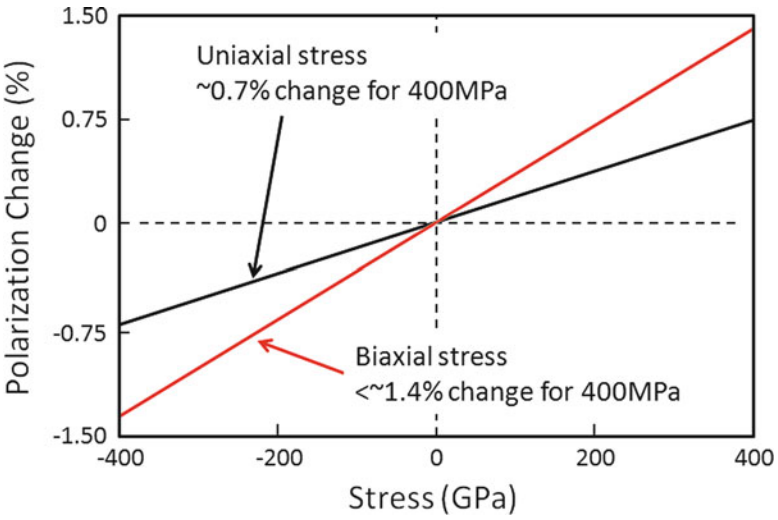



**Fig. 12.36** Change of polarization in the AlGaN layer under uniaxial and biaxial stresses

polarization is predicted for 400 MPa uniaxial and biaxial stresses, respectively. This observation indicates a weak stress dependence of the AlGaN electric field.

Stress has been proposed to shift the defect energy level by changing the atom-defect bond angle and bond length [74]. It was observed that both tensile and compressive stresses decrease $Pb_1$ (or $Pb_2$) defect levels for Si MOSFETs, resulting in an increase of gate leakage current [74]. For GaN HEMTs, our experimental results in Fig. 12.17 show that tensile stress increases gate current, while compressive stress decreases gate current. This can be explained qualitatively with Fig. 12.37, taking the N-vacancy as an example. As mentioned before, approximately 2.8-GPa biaxial stress is present in the AlGaN layer due to lattice mismatch. Compared to the unstrained lattice, the bond angle $\Theta_1$ decreases and $\Theta_2$ increases, and the bond length $L_1$ decreases and $L_2$ increases. The defect in the AlGaN becomes less stable under this large stress, and its energy level shifts up toward the conduction band. When a tensile wafer bending stress is applied, a similar trend of bond angle and bond length variation is expected. It adds to the effect of biaxial stress and shifts the defect level further up. In contrast, a compressive wafer bending stress tends to relax the strain from the lattice mismatch, therefore resulting in a downward shift of the defect level toward its original state. Since the gate current exponentially depends on the defect level, a decrease (or increase) of the defect level from tensile (or compressive) stress increases (or decreases) the gate current.

The overall effect of mechanical stress on the GaN HEMT gate leakage is a combination of the above discussed factors. The shift of the defect activation energy (or defect band energy) is expected to dominate the leakage variation, due to the negligible change of electric field and effective mass, as well as the exponential dependence of gate leakage on the defect level. A rigorous modeling of stress-altered defect level is yet to be developed.

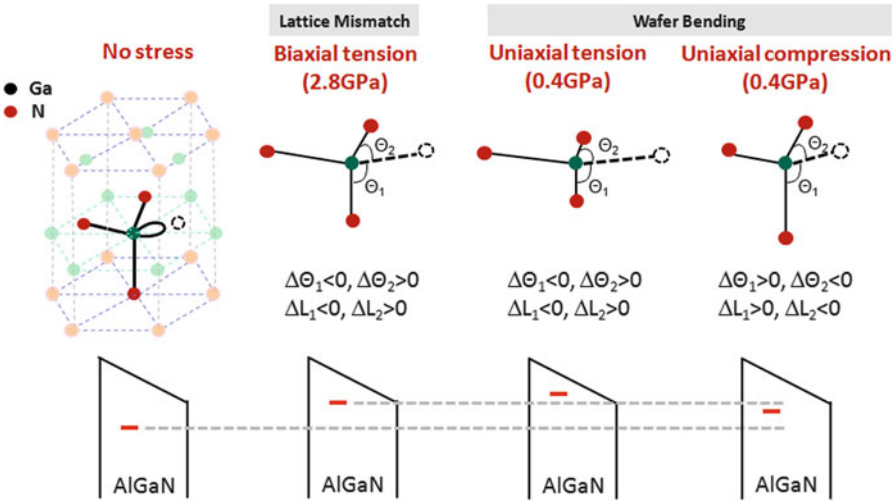



**Fig. 12.37** Schematic illustrations of the defect bond angle and bond length variation and the defect level shift under lattice mismatch and wafer bending stress

## 12.5 Conclusion

Stress is inherent to AlGaN/GaN HEMTs and can benefit or degrade performance and reliability depending on the operating condition. Lattice mismatch stress between AlGaN and GaN benefits the device by inducing polarization and creating the 2DEG, while additional stress generated via the inverse-piezoelectric effect can induce degradation. Mechanical stress applied by four-point wafer bending is used to isolate the effect of stress on the AlGaN/GaN HEMT devices in order to study the impact of stress on performance and reliability.

The effect of strain on AlGaN/GaN HEMT drain current and gauge factor was investigated experimentally and theoretically. A novel technique was developed to apply external stress to small (~1 $cm^2$) wafer samples while simultaneously taking electrical measurements. The enormous variation in AlGaN/GaN HEMT gauge factor measurements (4 to −40,000) reported in literature [30, 31, 33–37] was explained as a likely result of charge-trapping effects. After eliminating the charge-trapping effects, the measured gauge factor of the AlGaN/GaN HEMT was determined to be −2.8 ±0.4. This gauge factor indicates a small stress dependence on the device resistivity. The strain-altered 2DEG sheet carrier density and electron mobility were considered. It was found that externally applied mechanical stress has negligible effect on the electron density, due to the near cancellation of the stress-induced piezoelectric polarization in both the AlGaN and GaN layers. Strain is incorporated into an $sp^3d^5$ tight-binding model to calculate the mobility change under uniaxial and biaxial stresses. The simulation result suggests negligible mobility change due to the fact that the single GaN conduction band barely warps.

Comparing simulation with the experimental results, the best fit set of material parameters was determined.

To investigate the dominant gate leakage mechanism in reverse-biased GaN HEMT, a thorough investigation of the relationship between $E_{AlGaN}$ and $V_G$ is provided. In a simple model of an ideal 1D device, $E_{AlGaN}$ has a linear relationship to $V_G$ above $V_T$ and saturates below $V_T$. Adjusting $\sigma_{int}$ and $n_s$ based on experimental measurements provided an accurate $V_G$ versus $E_{AlGaN}$ relationship above threshold. Also, a 2D simulation was performed to provide insight on $E_{AlGaN}$ for large reverse biases much below $V_T$.

Several possible gate leakage mechanisms were investigated by comparing the experimental observations and theoretical modeling. The direct tunneling and bulk trap-assisted leakage mechanisms are unlikely to dominate the gate leakage in reverse-biased GaN HEMT. The theory of defect band-induced leakage is feasible to explain the experimental measurement, though it requires further investigation. The gate leakage increases (decreases) with tensile (compressive) stress, and its stress sensitivity decreases at larger reverse bias. To fully understand the stress-altered gate leakage, the effects of mechanical stress on individual contributing factors need to be further clarified and properly combined.

## References

1. R.J. Trew, High-frequency solid-state electronic devices. IEEE Trans. Electron Devices **52**, 638–649 (2005)
2. A. Sozza, A. Curutchet, C. Dua, N. Malbert, N. Labat, A. Touboul, AlGaN/GaN HEMT reliability assessment by means of low frequency noise measurements. Microelectron. Reliab. **46**, 1725–1730 (2006)
3. N. Sghaier, M. Trabelsi, N. Yacoubi, J.M. Bluet, A. Souifi, G. Guillot, C. Gaquière, J.C. DeJaeger, Traps centers and deep defects contribution in current instabilities for AlGaN/GaN HEMT's on silicon and sapphire substrates. Microelectron. J. **37**, 363–370 (2006)
4. H. Rao, G. Bosman, Device reliability study of high gate electric field effects in AlGaN/GaN high electron mobility transistors using low frequency noise spectroscopy. J. Appl. Phys. **108**, 053707-5 (2010)
5. W. Kruppa, S.C. Binari, K. Doverspike, Low-frequency dispersion characteristics of GaN HFETs. Electron. Lett. **31**, 1951–1952 (1995)
6. M.A. Khan, M.S. Shur, Q.C. Chen, J.N. Kuznia, Current/voltage characteristic collapse in AlGaN/GaN heterostructure insulated gate field effect transistors at high drain bias. Electron. Lett. **30**, 2175–2176 (1994)
7. J. Joh, J.A. del Alamo, J. Jimenez, A simple current collapse measurement technique for GaN high-electron mobility transistors. IEEE Electron Device Lett. **29**, 665–667 (2008)
8. S.C. Binari, W. Kruppa, H.B. Dietrich, G. Kelner, A.E. Wickenden, J.A. Freitas Jr., Fabrication and characterization of GaN FETs. Solid State Electron. **41**, 1549–1554 (1997)
9. S. Trassaert, B. Boudart, C. Gaquiere, D. Theron, Y. Crosnier, F. Huet, M.A. Poisson, Trap effects studies in GaN MESFETs by pulsed measurements. Electron. Lett. **35**, 1386–1388 (1999)
10. S.C. Binari, K. Ikossi, J.A. Roussos, W. Kruppa, P. Doewon, H.B. Dietrich, D.D. Koleske, A.E. Wickenden, R.L. Henry, Trapping effects and microwave power performance in AlGaN/GaN HEMTs. IEEE Trans. Electron Devices **48**, 465–471 (2001)

11. G. Meneghesso, A. Chini, E. Zanoni, M. Manfredi, M. Pavesi, B. Boudart, C. Gaquiere, Diagnosis of trapping phenomena in GaN MESFETs, in *Electron Devices Meeting, 2000. IEDM Technical Digest. International*, 2000, pp. 389–392
12. H. Kim, J. Lee, D. Liu, W. Lu, Gate current leakage and breakdown mechanism in unpassivated AlGaN/GaN high electron mobility transistors by post-gate annealing. Appl. Phys. Lett. **86**, 143505-3 (2005)
13. J. Jungwoo, J.A. del Alamo, Critical voltage for electrical degradation of GaN high-electron mobility transistors. IEEE Electron Device Lett. **29**, 287–289 (2008)
14. J. Joh, J.A. del Alamo, Mechanisms for electrical degradation of GaN high-electron mobility transistors, in *Electron Devices Meeting, 2006. IEDM '06. International*, 2006, pp. 1–4
15. Y. Sun, S.E. Thompson, T. Nishida, Physics of strain effects in semiconductors and metal-oxide-semiconductor field-effect transistors. J. Appl. Phys. **101** (2007)
16. S.Y. Son, Y.S. Choi, P. Kumar, H. Park, T. Nishida, R.K. Singh, S.E. Thompson, Strained induced changes in gate leakage current and dielectric constant nitrided Hf-silicate dielectric silicon MOS capacitors. Appl. Phys. Lett. **93**, 153505 (2008)
17. S. Dey, M. Agostinelli, C. Prasad, X. Wang, L. Shifren, Effects of hot carrier stress on reliability of strained-Si mosfets, in *Reliability Physics Symposium Proceedings, 2006. 44th Annual., IEEE International*, 2006, pp. 461–464
18. P.Y. Yu, M. Cardona, *Fundamentals of Semiconductors*, 3rd edn. (Springer, New York, 2003)
19. T.G.P. Rajagopal, J. Roberts, Large-area, device quality GaN on Si using a novel transition layer scheme. Mater. Res. Soc. Symp. Proc. **1**, 6–11 (2003)
20. O. Ambacher, J. Smart, J.R. Shealy, N.G. Weimann, K. Chu, M. Murphy, W.J. Schaff, L.F. Eastman, R. Dimitrov, L. Wittmer, M. Stutzmann, W. Rieger, J. Hilsenbeck, Two-dimensional electron gases induced by spontaneous and piezoelectric polarization charges in N- and Ga-face AlGaN/GaN heterostructures. J. Appl. Phys. **85**, 3222–3233 (1999)
21. D. Gregušová, J. Bernát, M. Držík, M. Marso, J. Novák, F. Uherek, P. Kordoš, Influence of passivation induced stress on the performance of AlGaN/GaN HEMTs. Phys. Status Solidi (c) **2**, 2619–2622 (2005)
22. C.S. Smith, Piezoresistance effect in germanium and silicon. Phys. Rev. **94**, 42 (1954)
23. S. Suthram, J.C. Ziegert, T. Nishida, S.E. Thompson, Piezoresistance coefficients of (100) silicon nMOSFETs measured at low and high (similar to 1.5 GPa) channel stress. IEEE Electron Device Lett. **28**, 58 (2007)
24. S. Timoshenko, *Strength of Materials*, 3rd edn. (R. E. Krieger Pub. Co, Huntington, 1976)
25. K. Wu, Strain effects on the valence band of silicon: piezoresistance in p-type silicon and mobility enhancement in strained silicon PMOSFET, PhD Electrical and Computer Engineering, University of Florida, Gainesville, 2005
26. S.E. Thompson, M. Armstrong, C. Auth, S. Cea, R. Chau, G. Glass, T. Hoffman, J. Klaus, Z.Y. Ma, B. Mcintyre, A. Murthy, B. Obradovic, L. Shifren, S. Sivakumar, S. Tyagi, T. Ghani, K. Mistry, M. Bohr, Y. El-Mansy, A logic nanotechnology featuring strained-silicon. IEEE Electron Device Lett. **25**, 191–193 (2004)
27. A.T. Bradley, R.C. Jaeger, J.C. Suhling, K.J. O'Connor, Piezoresistive characteristics of short-channel MOSFETs on (100) silicon. IEEE Trans. Electron Devices **48**, 2009 (2001)
28. A. Steegen, M. Stucchi, A. Lauwers, K. Maex, Silicide induced pattern density and orientation dependent transconductance in MOS transistors, in *IEDM Technical Digest*, 1999, p. 497
29. G. Meneghesso, G. Verzellesi, F. Danesin, F. Rampazzo, F. Zanon, A. Tazzoli, M. Meneghini, E. Zanoni, Reliability of GaN high-electron-mobility transistors: state of the art and perspectives. IEEE Trans. Device Mater. Reliab. **8**, 332–343 (2008)
30. R. Gaska, J.W. Yang, A.D. Bykhovski, M.S. Shur, V.V. Kaminski, S.M. Soloviov, The influence of the deformation on the two-dimensional electron gas density in GaN–AlGaN heterostructures. Appl. Phys. Lett. **72**, 64–66 (1998)
31. M. Eickhoff, O. Ambacher, G. Krotz, M. Stutzmann, Piezoresistivity of AlxGa1-xN layers and AlxGa1-xN/GaN heterostructures. J. Appl. Phys. **90**, 3383–3386 (2001)

32. B.S. Kang, S. Kim, J. Kim, F. Ren, K. Baik, S.J. Pearton, B.P. Gila, C.R. Abernathy, C.C. Pan, G.T. Chen, J.I. Chyi, V. Chandrasekaran, M. Sheplak, T. Nishida, S.N.G. Chu, Effect of external strain on the conductivity of AlGaN/GaN high-electron-mobility transistors. Appl. Phys. Lett. **83**, 4845–4847 (2003)
33. B.S. Kang, S. Kim, F. Ren, J.W. Johnson, R.J. Therrien, P. Rajagopal, J.C. Roberts, E.L. Piner, K.J. Linthicum, S.N.G. Chu, K. Baik, B.P. Gila, C.R. Abernathy, S.J. Pearton, Pressure-induced changes in the conductivity of AlGaN/GaN high-electron mobility-transistor membranes. Appl. Phys. Lett. **85**, 2962–2964 (2004)
34. T. Zimmermann, M. Neuburger, P. Benkart, F.J. Hernandez-Guillen, C. Pietzka, M. Kunze, I. Daumiller, A. Dadgar, A. Krost, E. Kohn, Piezoelectric GaN sensor structures. IEEE Electron Device Lett. **27**, 309–312 (2006)
35. O. Yilmazoglu, K. Mutamba, D. Pavlidis, M.R. Mbarga, Strain sensitivity of AlGaN/GaN HEMT structures for sensing applications. IEICE Trans. Electron. **E89c**, 1037–1041 (2006)
36. C. Chia-Ta, H. Shih-Kuang, E.Y. Chang, L. Chung-Yu, H. Jui-Chien, L. Ching-Ting, Changes of electrical characteristics for AlGaN/GaN HEMTs under uniaxial tensile strain. IEEE Electron Device Lett. **30**, 213–215 (2009)
37. B.S. Kang, S. Kim, J. Kim, R. Mehandru, F. Ren, K. Baik, S.J. Pearton, B.P. Gila, C.R. Abernathy, C.C. Pan, G.T. Chen, J.I. Chyi, V. Chandrasekaran, M. Sheplak, T. Nishida, S.N.G. Chu, AlGaN/GaN high electron mobility transistor structures for pressure and pH sensing. Phys. Status Solidi (c) **2**, 2684–2687 (2005)
38. A.D. Koehler, A. Gupta, M. Chu, S. Parthasarathy, K.J. Linthicum, J.W. Johnson, T. Nishida, S.E. Thompson, Extraction of AlGaN/GaN HEMT Gauge factor in the presence of traps. IEEE Electron Device Lett. **31**, 665–667 (2010)
39. M. Chu, T. Nishida, X.L. Lv, N. Mohta, S.E. Thompson, Comparison between high-field piezoresistance coefficients of Si metal-oxide-semiconductor field-effect transistors and bulk Si under uniaxial and biaxial stress. J. Appl. Phys. **103** (2008)
40. M. Chu, A.D. Koehler, A. Gupta, T. Nishida, S.E. Thompson, Simulation of AlGaN/GaN high-electron-mobility transistor gauge factor based on two-dimensional electron gas density and electron mobility. J. Appl. Phys. **108**, 104502–104506 (2010)
41. M. Nido, Effect of biaxial strain on cubic and hexagonal gan analyzed by tight-binding method. Jpn. J. Appl. Phys. Part 2-Lett. **34**, 1513–1516 (1995)
42. B. Jogai, Effect of in-plane biaxial strains on the band structure of wurtzite GaN. Phys. Rev. B **57**, 2382–2386 (1998)
43. T. Yang, S. Nakajima, S. Sakai, Electronic structures of wurtzite GaN, InN and their alloy Ga1-xInxN calculated by the tight-binding method. Jpn. J. Appl. Phys. **34**, 5912–5921 (1995)
44. J.C. Slater, G.F. Koster, Simplified Lcao method for the periodic potential problem. Phys. Rev. **94**, 1498–1524 (1954)
45. M. Chu, Y.K. Sun, U. Aghoram, S.E. Thompson, Strain: a solution for higher carrier mobility in nanoscale MOSFETs. Ann. Rev. Mat. Res. **39**, 203–229 (2009)
46. A.D. Bykhovski, V.V. Kaminski, M.S. Shur, Q.C. Chen, M.A. Khan, Piezoresistive effect in wurtzite n-type GaN. Appl. Phys. Lett. **68**, 818–819 (1996)
47. S. Mingiacchi, P. Lugli, A. Bonfiglio, G. Conte, M. Eickhoff, O. Ambacher, A. Rizzi, A. Passaseo, P. Visconti, R. Cingolani, Thermoresistive and piezoresistive properties of wurtzite N-GaN. Phys. Stat. Soli. A **190**, 281–286 (2002)
48. M. Yamaguchi, T. Yagi, T. Sota, T. Deguchi, K. Shimada, S. Nakamura, Brillouin scattering study of bulk GaN. J. Appl. Phys. **85**, 8502–8504 (1999)
49. C. Deger, E. Born, H. Angerer, O. Ambacher, M. Stutzmann, J. Hornsteiner, E. Riha, G. Fischerauer, Sound velocity of AlxGa1-xN thin films obtained by surface acoustic-wave measurements. Appl. Phys. Lett. **72**, 2400–2402 (1998)
50. K. Tsubouchi, N. Mikoshiba, Zero-temperature-coefficient saw devices on Aln epitaxial-films. IEEE Trans. Sonics Ultrason. **32**, 634–644 (1985)

51. J. Jungwoo, X. Ling, J.A. del Alamo, Gate current degradation mechanisms of GaN high electron mobility transistors, in *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, 2007, pp. 385–388
52. H. Zhang, E.J. Miller, E.T. Yu, Analysis of leakage current mechanisms in Schottky contacts to GaN and Al0.25 Ga0.75N/GaN grown by molecular-beam epitaxy. J. Appl. Phys. **99**, 023703 (2006)
53. O. Mitrofanov, M. Manfra, Poole-Frenkel electron emission from the traps in AlGaN/GaN transistors. J. Appl. Phys. **95**, 6414–6419 (2004)
54. D. Yan, H. Lu, D. Cao, D. Chen, R. Zhang, Y. Zheng, On the reverse gate leakage current of AlGaN/GaN high electron mobility transistors. Appl. Phys. Lett. **97**, 153503 (2010)
55. S. Karmalkar, D.M. Sathaiya, M.S. Shur, Mechanism of the reverse gate leakage in AlGaN/GaN high electron mobility transistors. Appl. Phys. Lett. **82**, 3976–3978 (2003)
56. D.M. Sathaiya, S. Karmalkar, Thermionic trap-assisted tunneling model and its application to leakage current in nitrided oxides and AlGaN/GaN high electron mobility transistors. J. Appl. Phys. **99**, 093701 (2006)
57. W. Chikhaoui, J.M. Bluet, P. Girard, G. Bremond, C. Bru-Chevallier, C. Dua, R. Aubry, Deep levels investigation of AlGaN/GaN heterostructure transistors. Phys. B: Condens. Matter **404**, 4877–4879 (2009)
58. E.J. Miller, X.Z. Dang, E.T. Yu, Gate leakage current mechanisms in AlGaN/GaN heterostructure field-effect transistors. J. Appl. Phys. **88**, 5951–5958 (2000)
59. J.P. Ibbetson, P.T. Fini, K.D. Ness, S.P. DenBaars, J.S. Speck, U.K. Mishra, Polarization effects, surface states, and the source of electrons in AlGaN/GaN heterostructure field effect transistors. Appl. Phys. Lett. **77**, 250–252 (2000)
60. J.W. Chung, J.C. Roberts, E.L. Piner, T. Palacios, Effect of gate leakage in the subthreshold characteristics of AlGaN/GaN HEMTs. Electron Device Lett. IEEE **29**, 1196–1198 (2008)
61. J. W. Chung, X. Zhao, T. Palacios, Estimation of trap density in AlGaN/GaN HEMTs from subthreshold slope study, in *Device Research Conference, 2007 65th Annual*, 2007, pp. 111–112
62. L.K.A.R. Kakanakov, Ohmic contacts for high power and high temperature microelectronics, *Micro Electronic and Mechanical Systems,* 2009
63. K.L. Jensen, Electron emission theory and its application: Fowler–Nordheim equation and beyond, Lyon, France, 2003, pp. 1528–1544
64. S. Fleischer, P.T. Lai, Y.C. Cheng, Simplified closed-form trap-assisted tunneling model applied to nitrided oxide dielectric capacitors. J. Appl. Phys. **72**, 5711–5715 (1992)
65. J. Frenkel, On pre-breakdown phenomena in insulators and electronic semi-conductors. Phys. Rev. **54**, 647 (1938)
66. P. Pipinys, V. Lapeika, Temperature dependence of reverse-bias leakage current in GaN Schottky diodes as a consequence of phonon-assisted tunneling. J. Appl. Phys. **99**, 093709 (2006)
67. P. Ohlckers, P. Pipinys, Phonon-assisted tunneling process in amorphous silicon nanostructures and GaAs nanowires. Physica E **40**, 2859–2861 (2008)
68. J.-G. Tartarin, G. Soubercaze-Pun, J.-L. Grondin, L. Bary, J. Mimila-Arroyo, J. Chevallier, Generation-recombination defects in AlGaN/GaN HEMT on SiC substrate, evidenced by low frequency noise measurements and SIMS characterization. AIP Conf. Proc. **922**, 163–166 (2007)
69. S. Bradley, A. Young, L. Brillson, M. Murphy, W. Schaff, Role of barrier and buffer layer defect states in AlGaN/GaN HEMT structures. J. Electron. Mater. **30**, 123–128 (2001)
70. M. Wolter, P. Javorka, M. Marso, A. Fox, R. Carius, A. Alam, M. Heuken, P. Kordoš, H. Lüth, Photoionization spectroscopy of traps in doped and undoped AlGaN/GaN HEMTs. Phys. Status Solidi (c) **0**, 82–85 (2003)
71. T. Okino, M. Ochiai, Y. Ohno, S. Kishimoto, K. Maezawa, T. Mizutani, Drain current DLTS of AlGaN-GaN MIS-HEMTs. IEEE Electron Device Lett. **25**, 523–525 (2004)

72. N. Sghaier, N. Yacoubi, J.M. Bluet, A. Souifi, G. Guillot, C. Gaquiere, J.C. De Jaeger, Current instabilities and deep level investigation on AlGaN/GaN HEMT's on silicon and sapphire substrates, in *Microelectronics, 2004. ICM 2004 Proceedings. The 16th International Conference on*, 2004, pp. 672–675
73. A.R. Arehart, Investigation of electrically active defects in GaN, AlGaN, and AlGaN/GaN high electron mobility transistors, Ohio State University, 2009
74. Y.S. Choi, T. Nishida, S.E. Thompson, Impact of mechanical stress on direct and trap-assisted gate leakage currents in p-journal article silicon metal-oxide-semiconductor capacitors. Appl. Phys. Lett. **92**, 173507 (2008)
75. S. Adachi, *Handbook on Physical Properties of Semiconductors*, vol. 1 (Kluwer, Boston, 2004)
76. S.R.M. Levinstein, M. Shur, *Handbook Series on Semiconductor Parameters* (World Scientific, London, 1999)
77. S.M.A.F.M. Anwar, S. Wu, R.T. Webster, Temperature dependent transport properties in GaN, AlxGa1-xN semiconductors. Transport **48**, 567 (2001)
78. R. Quay, A temperature dependent model for the saturation velocity in semiconductor materials. Mater. Sci. Semicond. Process. **3**, 149 (2000)

# Chapter 13
# Reliability Issues in AlGaN/GaN High Electron Mobility Transistors

**E.A. Douglas, L. Liu, C.F. Lo, B.P. Gila, F. Ren, and Stephen J. Pearton**

**Abstract** As AlGaN/GaN high electron mobility transistors gain commercial acceptance for use in high-power and high-frequency applications, it is becoming ever important to understand the degradation mechanisms that drive failure in the field. Because of the complex nature and multifaceted operation modes of these devices, reliability studies must go beyond the typical Arrhenius accelerated life tests. To do this, we investigated the electric field-driven degradation in devices with different gate metallization, device dimensions, and electric field mitigation techniques (such as source field plate) and the effect of device fabrication processes for both dc and RF stress conditions.

## 13.1 Introduction

Gallium nitride-based high electron mobility transistors (HEMTs) have become attractive for use in high-power and high-frequency applications due to large energy band gaps, great electron mobility, high breakdown voltages, and considerable 2-D electron gas densities as compared to its GaAs counterpart [1–31]. As AlGaN/GaN HEMTs gain traction for use in commercial and military applications, it has become increasingly critical for reliability studies to extend beyond the normal temperature-accelerated life tests. The failure rate at operating temperatures is extrapolated from the failure rate under thermally accelerated stress conditions through the Arrhenius equation. This equation is limited to expressing acceleration of failure mechanisms that are due to physiochemical reactions.

E.A. Douglas • B.P. Gila • S.J. Pearton (✉)
Department of Materials Science and Engineering, University of Florida, Gainesville, FL 32611, USA
e-mail: spear@mse.ufl.edu

L. Liu • C.F. Lo • F. Ren
Department of Chemical Engineering, University of Florida, Gainesville, FL 32611, USA

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_13,
© Springer Science+Business Media New York 2013

With the bulk of reliability predictions focusing on Arrhenius extrapolations of device lifetime, they do not capture failure mechanisms that govern degradation under dc or RF operation that cannot be accelerated by elevated temperatures.

Employment of GaN HEMTs for high-power radar systems will require devices to be subjected to large-signal RF while being driven into saturation, resulting in devices experiencing high electric fields and high current densities. Impressive mean-time-to-failure values of greater than $10^7$ h have been reported at operating temperatures below 200°C, with activation energies ranging from 0.18 to 2 eV [1–4]. However, these studies neglect the effect of high electric field and current on device lifetime. Additionally as the demand increases for faster data communication at higher frequencies, scaling of the gate will be driven below 0.2 μm [4]. In particular, we have investigated field-driven failure mechanisms that may limit the device lifetime at normal operating temperatures.

As the device design and material processing technology for AlGaN/GaN HEMTs has matured over the years, several failure mechanisms that limited device lifetime have been addressed and improved. However, there are still numerous degradation mechanisms that have been reported to plague AlGaN/GaN HEMTs. These mechanisms can be grouped together into three main categories that affect lifetime: contact degradation, hot-electron effect, and inverse piezoelectric effect. Both Schottky and ohmic contacts have shown excellent stability below 300°C. Piazza et al. have reported an increase in contact resistance and passivation cracking due to Ga out-diffusion and Au interdiffusion after a 100-h thermal storage test stress at 340°C [8]. Nickel-based Schottky contacts have been shown to form nickel nitrides on GaN at annealing temperatures as low as 200°C, resulting in a significant decrease in Schottky barrier height [9]. Future improvement of the stability of contacts will allow for high-power operation of AlGaN/GaN HEMTs. The observed current collapse and gate lag under high-voltage and high-current operation have been attributed to hot electrons. Hot electrons are electrons that have been accelerated in a large electric field, resulting in very high kinetic energy, which can result in trap formation. Creation of traps can occur in both the AlGaN layer and the buffer, leading to reversible degradation of transconductance and saturated drain current [2, 5, 10]. GaN is a piezoelectric material, which has been exploited for a wide range of applications from gas detection to mechanical pressure sensors. However, under high bias conditions, the electric field induces additional tensile stress to the already strained AlGaN layer. Several authors have shown that upon reaching a "critical voltage," irreversible damage to the device occurs resulting in defect formation through which electron leakage can occur [9–13].

Degradation of AlGaN/GaN HEMTs with micron and submicron gate lengths was studied under both dc and RF operation. In order to gain further understanding of the mechanisms driving device degradation, several different device designs have been investigated, including different gate metallization, device dimensions, electric field mitigation techniques (such as source field plate), and the effect of device fabrication processes. Due to the complex nature of AlGaN/GaN HEMTs and the multifaceted operation modes of these devices, isolating fundamental failure mechanisms has been anything but trivial.

## 13.2 GaN Degradation Mechanisms

### *13.2.1 Hot-Carriers and Trap Generation*

Permanent device degradation after high $V_{DG}$ stress under on-state conditions has been attributed to the presence of hot electrons. In GaAs-based devices, hot electrons generate holes which are accumulated by the gate and result in a negative shift in $V_T$ [14–16]. Typically, $I_G$ is used to derive the field-acceleration laws for failure. Impact ionization, however, is negligible in GaN HEMTs. This is due to the fact that tunneling injection dominates gate current, preventing gate current from being used as an indicator for hot-electron degradation [1, 4]. However, these hot electrons likely lead to trap generation at the AlGaN/GaN interface and/or at the passivation GaN cap interface. As in GaAs- and InP-based HEMTs, traps lead to an increase in the depletion region between the gate and the drain, ultimately resulting in an increase in drain resistance and subsequently a decrease in IDSS. Comparatively, under off-state conditions the degradation is greatly reduced due to the reduction of electrons present in the channel. Sozza et al. showed that GaN/AlGaN/GaN HEMTs that underwent a 3,000-h on-state stress resulted in an increase in surface traps with an activation energy of about 0.55 eV [3]. On the other hand, devices stressed under off-state conditions saw a very small increase in traps.

Meneghesso et al. have employed the use of electroluminescence (EL) to study the effect of hot-carriers and its dependence on stress conditions [1]. Uniform EL emission was observed along the channel for devices stressed at VGS = 0 V and VDS = 20 V, which is due to hot electrons. However, there is no presence of hot spots or current crowding. On the other hand, under off-state conditions with VGS = −6 V and VDS = 20 V (resulting in a VGD = −26 V), the EL emission from the channel is not uniform. These hot spots may be due to injection of electrons from the gate into the channel. Due to the high bias conditions, the electrons acquire enough energy to give rise to photon emission.

### *13.2.2 Contact Degradation*

As described above, contact degradation and gate sinking are significant degradation mechanisms at elevated temperatures in GaAs- and InP-based HEMTs. This has not yet proven to be a significant issue with AlGaN/GaN HEMTs at temperatures below 400°C for Pt/Au Schottky contacts and Ti/Al/Pt/Au annealed ohmic contacts [1, 10]. An increase in Schottky barrier height was observed for Ni/Au Schottky contacts after dc stress at elevated junction temperatures (200°C) [1, 10, 32–35]. This was due to a consumption of an interfacial layer between the Schottky contact and the AlGaN layer. Though the resulting positive shift in the Schottky barrier height, and thus the pinch-off voltage, is ideal, the subsequent change in $I_{DSS}$ is not

favorable. Unstressed devices were subjected to an anneal after the Schottky contact was deposited in order to decrease the interfacial layer between the gate and semiconductor. Devices that underwent the gate anneal showed 50% less degradation during a 24-h stress test as opposed to devices that did not receive a gate anneal [1, 10]. Thermal storage tests up to 2,000 h on Ti/Al/Ni/Au ohmic contacts at and above 290°C showed an increase in contact resistance as well as surface roughness due to growth of Au-rich grains that ultimately led to cracks in passivation [32–39]. The two primary degradation mechanisms were determined to be Au interdiffusion within the metal layers and Ga out-diffusion from the semiconductor into the metallic compounds. Similar degradation was observed after dc stress tests that resulted in junction temperatures equivalent to the thermal storage tests. Due to the high-power capability of AlGaN/GaN HEMTs, proper temperature management is crucial in order to optimize device performance under high-current and high-voltage operation [40–47]. Self-heating of devices can ultimately result in poor device performance through contact degradation. As shown above, reliability of contacts is highly dependent upon both metal schemes and processing during fabrication.

### *13.2.3 Inverse Piezoelectric Effect*

Several research groups have shown that high reverse bias on the gate results in a defect-generated path of gate current leakage [1, 10, 32]. It has been determined that this defect formation mechanism is a result of the converse piezoelectric effect. Due to the fact that GaN and AlGaN are intrinsically piezoelectric materials, the presence of high electric fields will result in an increase in stress within the GaN and AlGaN layers. AlGaN is lattice mismatched to GaN, resulting in significant tensile strain, even in the absence of an electric field. If under electric stress the elastic energy within the AlGaN/GaN layers surpasses a critical value, the strained layer will relax through crystallographic defect formation. It is possible that the defects could be electrically active and result in device degradation [32].

J. Joh et al. have established that $I_D$ and $I_G$ degradation under high reverse gate bias occurs at a critical voltage, typically above 20 V on $V_{DG}$ [37]. This is also correlated with a sharp rise in both source and drain resistance as well as a positive shift in $V_T$. However, the critical voltage for devices can deviate substantially within one wafer, though adjoining devices appear to exhibit similar performance. This broad distribution of critical voltage, ranging from $V_{DG}$ of ~15 to ~30 V, has been attributed to slow changes within the substrate or epilayer growth over the wafer [35–40]. As mentioned above, hot electrons are generated exponentially as the field increases and only linearly with current [1, 4]. Though stress experiments in the high-power state have shown that increasing $I_{Dstress}$ does not significantly accelerate degradation, it was found that the critical voltage for reverse bias stress in which $I_{Goff}$ dramatically increases is dependent upon the $I_{Dstress}$, as $V_{crit}$ increases with increasing $I_{Dstress}$ [32, 37]. It is evident from this result that hot electrons are not the driving degradation mechanism for this stress condition. In order to verify

the converse piezoelectric effect, TEM cross sections were studied by Chowdhury et al. after stressing with $V_{DS} = 40$ V and $I_{D0} = 250$ mA/mm at various baseplate temperatures, which corresponded to a junction temperature of 250, 285, and 320°C based on device modeling [9]. Unstressed devices showed no evidence of pits or cracks near the edge of the Schottky contact. However, all stressed devices showed evidence of pit-like defects on the drain side of the gate. The depth of the pit was about 10 nm and remained within the AlGaN layer. Crack-like defects were observed in a few of the stressed devices and appeared to originate at the bottom of the pit defect, extending to the heterointerface of the AlGaN/GaN layer and occasionally into the GaN buffer. As the junction temperature increased, the time after which the crack appeared decreased, developing within 6 h at a temperature of 320°C. Gate metal was also observed to diffuse ~2 nm into the defect crack. The formation of the crack was hypothesized by Chowdhury et al. to originate in the deepest points in the defect pit and spread along the gate width, thus explaining the presence of cracks in very shallow defect pits [9].

Del Alamo et al. have postulated that the converse piezoelectric effect is solely an electric field-driven degradation mechanism due to the fact that it is the induced mechanical stress that results in the relaxation of the AlGaN layer [10, 38]. It has also been hypothesized by the del Alamo group that current should not drive this mechanism, except for indirect self-heating that would accelerate degradation of the device. Device design that affects the profile of the electric field on the drain side of the gate will also, in turn, impact the critical voltage [38].

Sarua et al. have investigated the effect of piezoelectric strain in AlGaN/GaN FETs under bias with micro-Raman spectroscopy [13]. Devices implemented 30 nm of $Al_{0.25}$ $Ga_{0.75}$ N on 1.2 μm of undoped, insulating GaN on insulating 4H-SiC with 2 × 50 μm (source-drain gap = 4.8 μm, gate length = 1.2 μm). It was confirmed with 2-D finite element simulations that a pinched-off device at $V_{DS} = 20$ V results in a peak electric field on the drain side of the gate within the AlGaN layer. However, the z component of the electric field extended down into the GaN layer [13]. Fe doping of the GaN buffer layer is often used in order to improve the control of short-channel effects in HFETs [16]. It was later shown by Sarua et al. that Fe-doped GaN, which raises the acceptor concentration and decreases the depletion width in the GaN buffer, confines the z component of the electric field to the AlGaN/GaN interface [13]. Self-heating will occur under high-power stress conditions, which results in a compressive thermal strain/stress [41, 42]. It is possible that due to the mitigation of the piezoelectric stress by the thermal stress, slower device degradation was seen in devices stressed under the high-power state as opposed to off-state condition. This is in contrast to the hypothesis of del Alamo et al. that higher temperatures will result in an acceleration of device degradation [10, 43, 44].

Other issues can lead to additional compressive and tensile strains on the underlying epitaxial layers, including SiN passivation, which is used extensively to minimize surface traps on the AlGaN surface. Mastro et al. reported the simulated effects of nonuniform strain due to SiN passivation [17]. Typically, SiN has a relatively small magnitude of stress as compared to the tensile strain

present in the AlGaN layer due to lattice mismatch. The strain in SiN is highly dependent on processing conditions, i.e., thickness, frequency of the plasma during PECVD, pressure, and temperature. When deposited on the device, variations and discontinuities can increase the stress fields. For instance, the opening at the edge of the gate metal will result in a force on the AlGaN which will be perpendicular to the gate edge and parallel to the surface of the AlGaN [17]. It was predicted by Mastro et al. that as the gate length decreases, the magnitude of the strain fields increases. This effect on gate length is of great importance due to the desire to continuously scale down the dimensions of the devices.

## 13.3 Device Structure

Structure "A" AlGaN/GaN HEMT heterostructures were grown by metal–organic chemical vapor deposition (MOCVD) on 6H SiC semi-insulating substrate [48–50], followed by a AlN nucleation layer, with a 2.25-μm Fe-doped GaN buffer layer grown on top. This was followed by 15 nm of Al0.28 Ga0.72 N, capped with 3 nm of unintentionally doped GaN. The HEMTs employed a Ti/Al/Ni/Au ohmic metallization and a Ni/Au double gate design with a gate periphery of 300 μm, source-to-gate and gate-to-drain distances of 2 μm with SiNx passivation. Structure "B" AlGaN/GaN HEMTs were grown on Si wafers by MOCVD. The growth process was nucleated with an AlN layer to avoid unwanted Ga–Si interactions. The epitaxial stack consisted of an AlGaN transition layer, an ~800-nm GaN buffer layer, and a 16-nm unintentionally doped Al0.26 Ga0.74 N barrier layer. HEMT fabrication began with Ti/Al/Ni/Au ohmic metallization and rapid thermal annealing in flowing $N_2$ at approximately 825°C. Device isolation was accomplished by using multiple-energy $N$ + implantation to produce significant lattice damage throughout the thickness of the GaN buffer layer. Immediately following implantation, the wafers were passivated with 70-nm-thick SiNx by PECVD at 300°C. Schottky gate definition was achieved by patterning the gate and selectively removing the SiNx passivation layer. The contact windows to the ohmic contact pads were also opened at the same time as defining the Schottky gate. After SiNx etching, wider gate patterns were redefined with another photolithography step, and contact windows to the ohmic contact pads were also opened. Ni-/Au-based gate metallization was deposited on the gate and ohmic contact pads simultaneously. There was an additional metal deposition for the HEMT with the source field plate. The field plate was connected to the source terminal and extended by 1 μm over the gate to the gate-to-drain region. The source to gate distance and channel length of the HEMTs with and without the source field plate were kept constant at 1 and 4.7 μm, respectively. Structure "C" AlGaN/GaN HEMT heterostructures were also grown by MOCVD on c-plane sapphire. The epilayers consisted of a 1-μm-thick carbon-doped GaN buffer layer followed by a 55-nm-thick undoped GaN channel layer, 21 nm of Al0.25 Ga0.75 N, and 2.2-nm GaN cap layer. The gate dimension was 1 × 200 μm, and the distances of

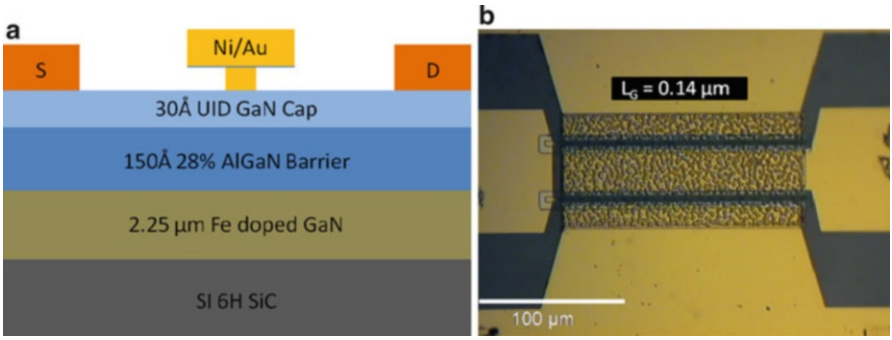



**Fig. 13.1** (*left*) Cross-section diagram of AlGaN/GaN HEMT. (*right*) Optical microscope image of device with gate length ($L_G$) of 0.14 µm (Reprinted with permission from Microelectronics Reliability 52, 23 (2012). Copyright 2012 Elsevier)

gate-to-source (Lgs) and gate-to-drain (Lgd) were 1 and 3 µm, respectively. The HEMTs were passivated using 400 nm of SiNx with a plasma-enhanced chemical vapor deposition (PECVD) system at 300°C. Structure "A" devices exhibited typical maximum drain currents of 1.1 A/mm, extrinsic transconductance of 250 mS/mm at VDS of 10 V, threshold voltage of −3.9 V, $f_T$ of 60 GHz, and $f_{MAX}$ of 90 GHz. With the devices biased at 28 V, the output power is about 5 W/mm, and the power-added efficiency was 30% at 10 GHz. Figure 13.1 shows the device cross section and an optical microscope image.

## 13.4 DC Stress Test Results

To study the effect of high electric field on GaN HEMTs, off-state step stresses were performed in the dark at room temperature with the gate biased up to −100 V reverse gate voltage at various fixed source-drain bias using an HP 4156 C semiconductor parameter analyzer for both structure A and B. The effect of gate length, gate-to-source, and gate-to-drain distance, as well as the effect of applied drain bias on breakdown, was investigated on structure A. As the gate voltage is stepped in 1 V increments for 1 min per step from −10 to −42 V, gate leakage current in AlGaN/GaN HEMTs with Ni-/Au-based gate metallization is seen to steadily increase until the critical voltage ($V_{CRI}$) is reached, resulting in a permanent increase in gate leakage current several orders of magnitude (Fig. 13.2). This sharp rise in current has been attributed to the inverse piezoelectric effect, in which the tensile strain in the AlGaN layer reaches a critical point whereby defect formation begins and subsequent lattice relaxation can occur. Even though a large increase in gate current is observed after stress, this is accompanied by a small decrease (~20%) in the drain current due to an increase in on-state resistance. Similarly, the transfer characteristics typically show a reduction in maximum transconductance of ~40 mS/mm as a result of the gate bias stress cycle and a shift in threshold voltage of ~0.6 V.

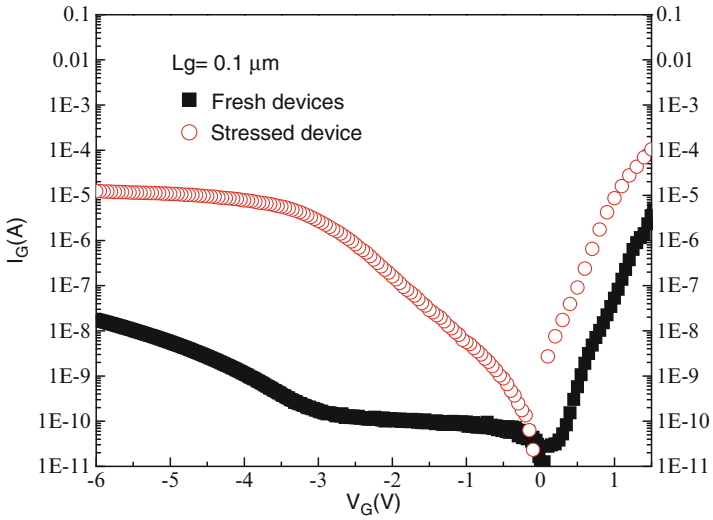



**Fig. 13.2** Gate current vs voltage before and after high reverse gate bias step stress

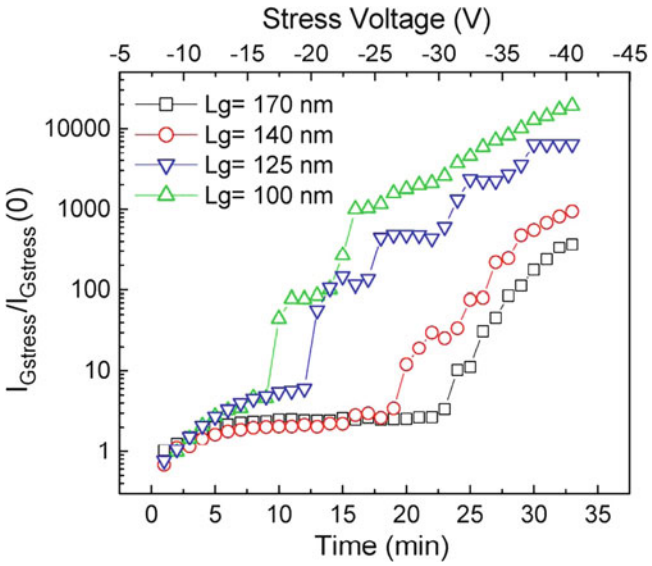



**Fig. 13.3** Gate stress current and stress voltage as a function of time for various gate lengths

Submicron gate lengths ranging from 0.1 to 0.17 µm were step-stressed from −10 to −42 V in the same manner. While a linear dependence of gate length on VCRI is observed (Fig. 13.3), ATLAS/Blaze simulations confirm that degradation occurs at a critical maximum electric field in the channel, ~1.8 MV/cm. This is consistent with previously published results indicating that failure is due to electric

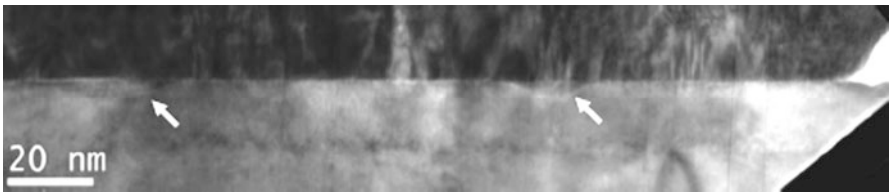



**Fig. 13.4** Cross-sectional TEM of drain edge of gate contact. *White arrows* indicate regions of Ni and oxygen diffusion with an associated threading dislocation

field-induced tensile strain. Photoluminescence (PL) was performed on the degraded devices. The PL spectrum of an ungated, open area of an unstressed HEMT was measured. GaN showed a peak centered at 366 nm, while the AlGaN peak was centered at 352 nm. After stressing the device, PL spectra showed a decrease in intensity in both peaks, but no additional defect peaks appeared. This degradation is indicative of the formation of non-radiative centers in both the GaN and AlGaN layers.

## 13.5 Field Plate

In order to improve the off-state breakdown voltage, a source field plate over the gate electrode was fabricated to reduce the peak electric field at the drain side of the gate edge with structure "B" devices. This resulted in a significant increase in $V_{CRI}$ from −40 to −65 V. Transmission electron microscopy (TEM) images indicated the presence of Ni from the gate metal stack interacting with the underlying nitride layer in close proximity to a threading dislocation (Fig. 13.4).

Figure 13.5 shows the typical step-stress results of more than 30 HEMTs with and without the source field plate; the gate current, Ig, has been plotted as a function of the stressed gate voltage. The devices were step-stressed in −1 V increments for 1 min at each increment, while grounding the source electrode and maintaining +5 V to the drain. The typical critical voltages of HEMTs without the source field plate were around −40 V [51]. This value increased to around −65 V for the HEMTs with the source field plate. This increase of the critical voltage for the HEMT with the source field plate was attributed to the reduction in electric field on the drain side of the gate edge, consistent with the simulated electric-field results [52].

## 13.6 Temperature Dependence

The interacting effects of temperature and electric field during off-state stress were studied for structure "A" devices. HEMTs with a gate length of 0.14 μm were step-stressed with an HP 4156 C semiconductor parameter analyzer from −10 to −42 V

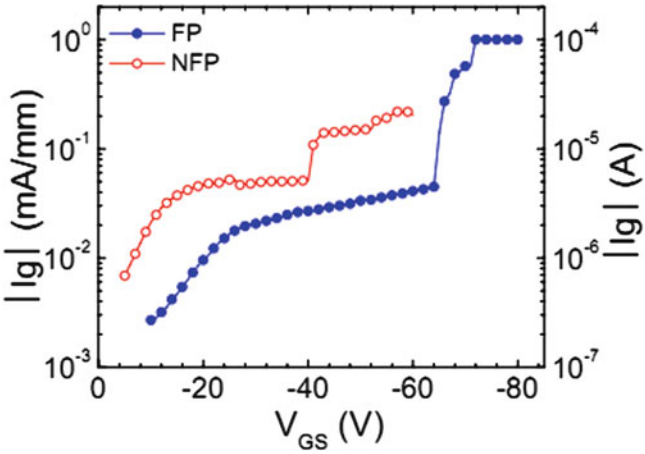



**Fig. 13.5** Off-state gate currents as a function of Vgs for HEMT with (FP) and without source field plate (NFP)

in the dark at various temperatures ranging from room temperature to 150°C, regulated by a heated chuck. Both source and drain were held at ground in order to symmetrically stress the gate contact. The voltage at which irreversible degradation occurred, $V_{CRI}$, followed a log relationship of the form

$$V_{CRI} = V_0 e^{-E_a/kT}, \tag{13.1}$$

where the activation energy is 38 meV and the value for $V_0$ is −28 V at room temperature (Fig. 13.6). The abrupt and permanent gate degradation at elevated temperatures was observed to occur at similar gate leakage currents regardless of critical voltage and stress temperature (~$10^{-7}$ A). Gate leakage currents immediately after critical voltage was reached also display similar values at all stress temperatures (~$10^{-6}$ A).

The critical voltage of our devices at 24°C is −30 V. As the temperature of the devices is increased to 150°C, the critical voltage is observed to linearly decrease (Fig. 13.7). Error bars shown in Fig. 13.6 indicate the small variations in critical voltage observed across the region of wafer in which the devices were located, likely due to slight variations in epitaxial layers [51]. Due to the fact that $V_{CRI}$ occurs at lower voltages as the temperature increases, the maximum electric field present at the edge of the gate at $V_{CRI}$ also decreases. ATLAS/Blaze electrical simulations indicate that the peak electric field decreases from 3.3 MV/cm at a critical voltage of −28 V with a stress temperature of 28°C to 2.6 MV/cm at a critical voltage of −18 V at 150°C. The gate leakage current measured during the step stress of four different devices stressed at four different temperatures ranging from 24 to 150°C further illustrates the negative temperature dependence. Previous reports in the literature attribute the sharp rise in $I_G$ at the critical voltage and resulting permanent $I_G$ degradation to the inverse piezoelectric effect.

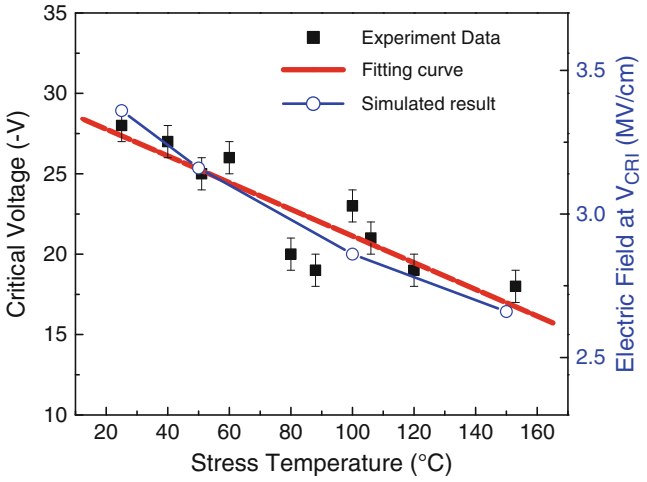



**Fig. 13.6** Stress temperature vs critical voltage of 0.14-μm gate length HEMT and stress temperature vs maximum electric field at critical voltage determined by ATLAS/Blaze 2-D simulations

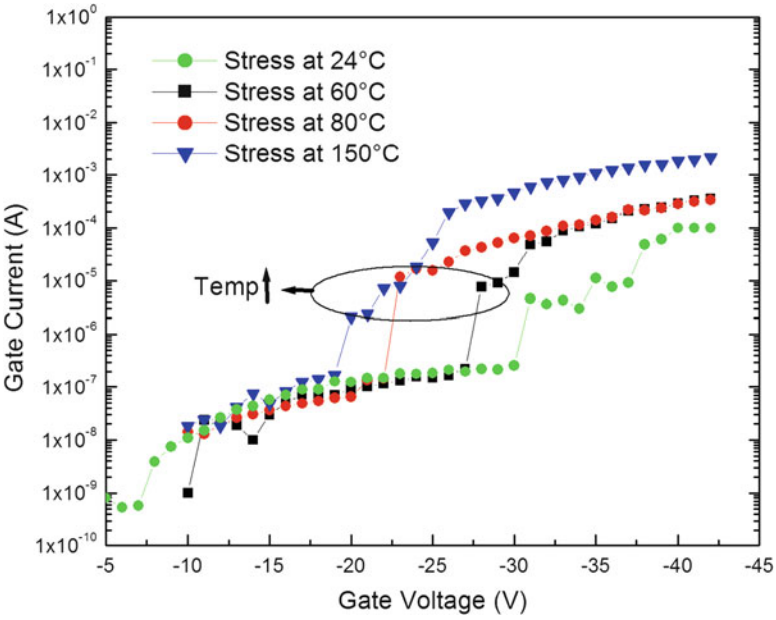



**Fig. 13.7** Gate current of four AlGaN/GaN HEMTs step-stressed from −10 to −42 V at 24, 60,80, and 150°C (Reprinted with permission from Microelectronics Reliability 52, 23 (2012). Copyright 2012 Elsevier)

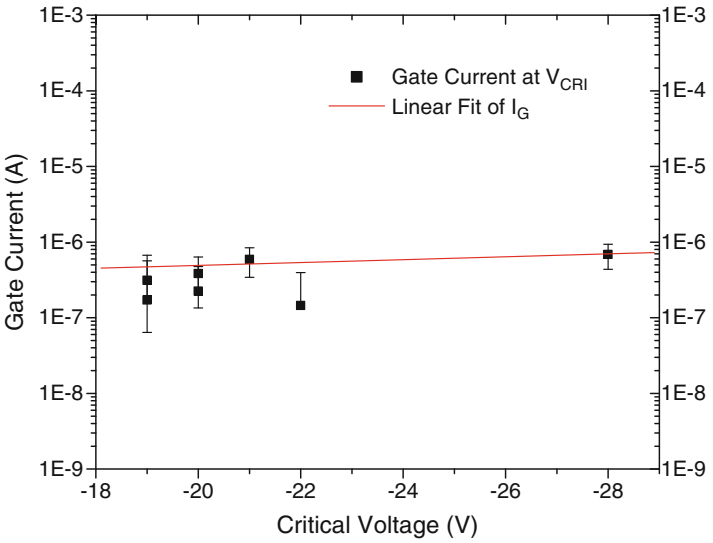



**Fig. 13.8** Gate current dependence on critical voltage of AlGaN/GaN HEMTs measured at $V_{CRI}$ (Reprinted with permission from Microelectronics Reliability 52, 23 (2012). Copyright 2012 Elsevier)

However, this result reveals that the breakdown which results in a sharp increase in gate leakage current does not occur at the same electric field and therefore does not occur at the same piezoelectric-induced stress.

Upon closer examination, one can see that the gate leakage current is independent of critical voltage, (Fig. 13.8). In addition, a linear fit of gate current indicates that there is no change in $I_G$ as a function of stress temperature either at the critical voltage or just above critical voltage ($V_{CRI}$–1 V). Error bars in Figs. 13.7 and 13.8 were calculated based on the variations of current measured during the step stress and variations across the region of the wafer in which the devices were located. Devices stressed at room temperature have been reported to display a significant dependence on gate voltage/electric field. Ni/Au metallization schemes for Schottky contacts on GaN have been determined to be thermally unstable above 400°C, with numerous nickel nitrides reported being formed at temperatures as low as 200°C [45–47]. Additionally, thermal instability of Ni-/Au-based ohmic contacts have been reported due to Ga out-diffusion and Au interdiffusion at elevated temperatures [46]. The voltage at which gate leakage current sharply increases (i.e., critical voltage) exhibits a negative temperature dependence with an activation energy of 41.6 meV. The abrupt and permanent gate degradation at elevated temperatures was observed to occur at similar gate leakage currents regardless of critical voltage and stress temperature (~$10^{-7}$ A). Gate leakage currents immediately after critical voltage was reached also displayed similar values at all stress temperatures (~$10^{-6}$ A), indicating similar leakage paths for all stress temperatures.

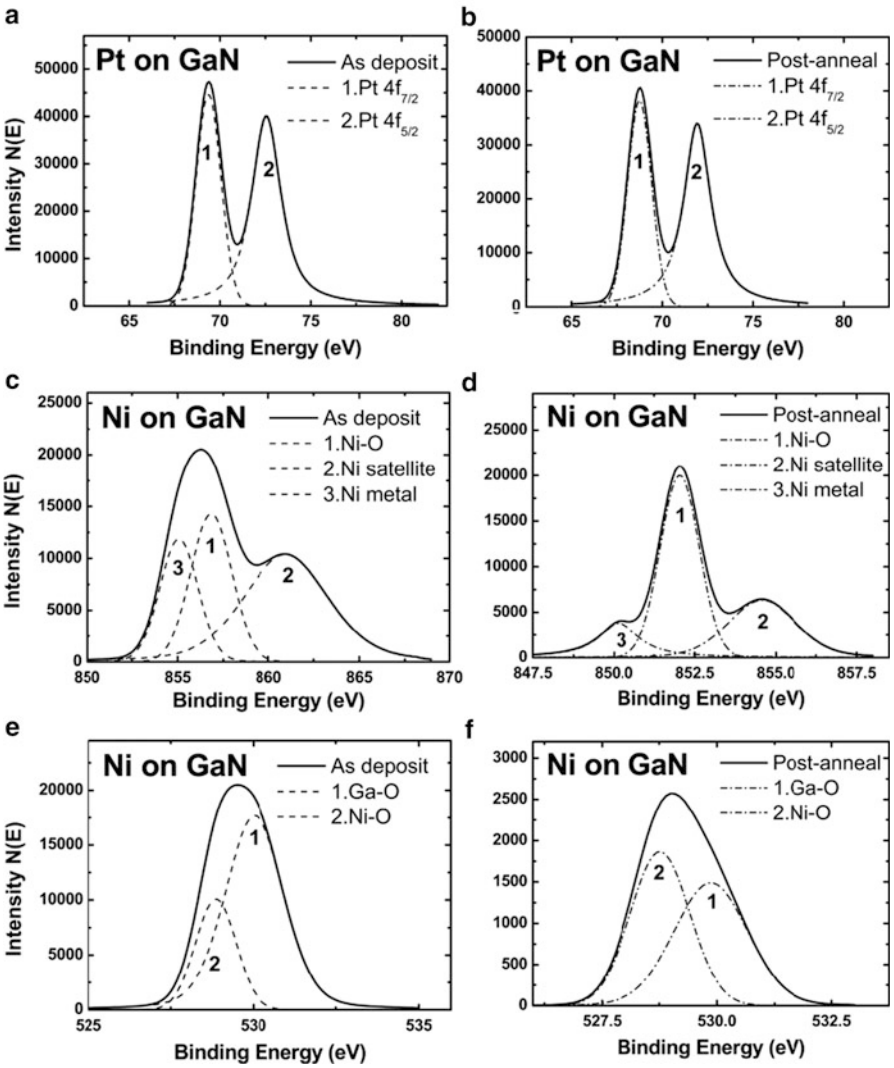



**Fig. 13.9** XPS spectra of (**a**) Pt 4f XPS spectra for as deposit Pt/GaN sample. (**b**) Pt 4f XPS spectra for the Pt/GaN sample annealed at 300°C for 30 min. (**c**) Ni 2p XPS spectra for as deposit Ni/GaN sample. (**d**) Ni 2p XPS spectra for the Ni/GaN sample annealed at 300°C for 30 min. (**e**) O 1 s XPS spectra for as deposit Ni/GaN sample. (**f**) O 1 s XPS spectra for the Ni/GaN sample annealed at 300°C for 30 min

Ni and Pt with a thickness of 1 nm were deposited on GaN in order to study the stability of the metals at elevated temperatures by X-ray photoelectron spectroscopy (XPS). It is necessary to keep the metal films thin, so the metal/GaN interface could be studied in XPS. As shown in Fig. 13.9, the spectra of Pt 4f XPS spectra for the as deposit Pt/GaN sample and post-annealed sample showed no significant

changes or shifts in the spectra. By contrast, by comparing the pre- and post-anneal spectra, the Ni-oxide peak increased, and the Ni metal peak decreased significantly. The discrepancy of the Ni-O peak position for the as deposit and annealed samples was due to charging effect. These results indicated the Ni oxidized during the anneal in the vacuum chamber. By observing the O 1 s peak and its changes between pre- and post-anneal, it was determined that the O-Ga bonding portion decreased and the O-Ni bonding portion increased. This led to the conclusion that the Ni film stripped oxygen from the native oxidized GaN surface and created a layer of NiO.

## 13.7 Gate Metal

Fifteen separate HEMTs with Ni/Au or Pt/Ti/Au gate metallization fabricated on structure "C" wafers were stressed for 60 s at each gate voltage step, while grounding the source electrode and maintaining +5 V on the drain. The stress started at −10 V of gate voltage, and the voltage step was kept at 1 V increments. As shown in Fig. 13.10, the HEMTs with Ni/Au gate metallization exhibited a critical voltage around −55 V. However, there was no critical voltage observed for the HEMT with the Pt/Ti/Au gate metallization up to −100 V, which was limited by the instrument used in this experiment. This suggested that the use of Pt-based gate metallization could extend the operating bias conditions and improve the device reliability. The Schottky barrier height and ideality of the Pt/Ti/Au were 1.23 and 1.21 V, respectively, which did not exhibit noticeable changes as a result of the bias stressing. On the contrary, the HEMTs with Ni/Au gate metallization showed significantly higher gate reverse bias leakage current and much lower breakdown

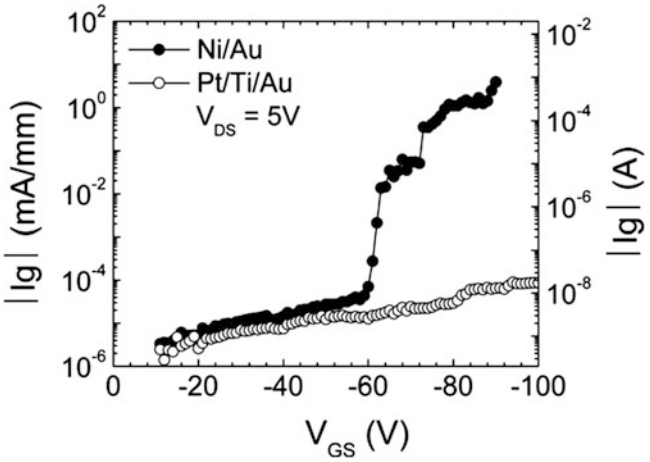



**Fig. 13.10** Off-state gate currents as a function of gate for the HEMTs fabricated with Ni/Au or Pt/Ti/Au gate metallization

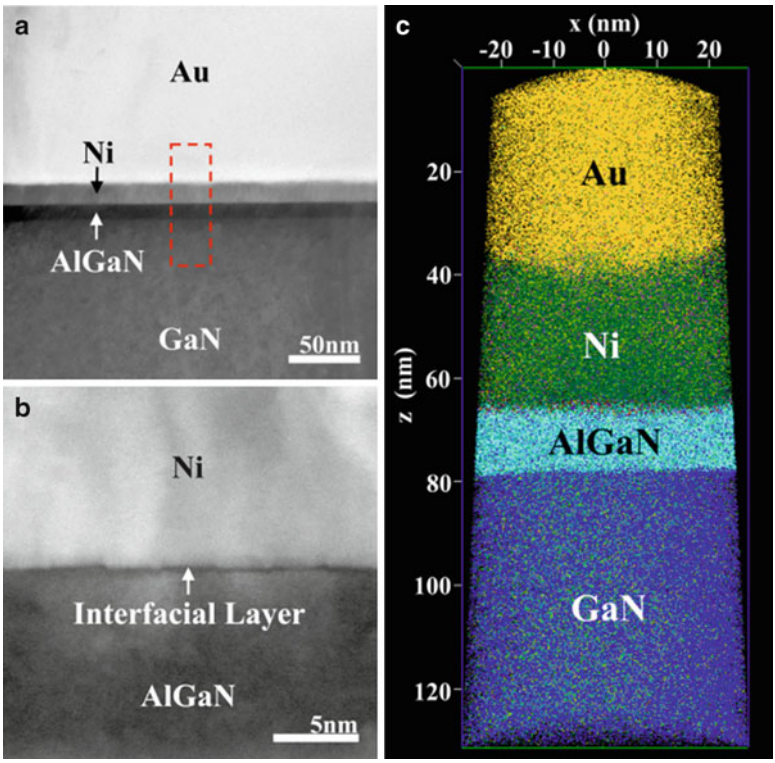



**Fig. 13.11** (**a**) HAADF-STEM image of the gate region of an AlGaN/GaN HEMT structure showing the distinct gate metal stack layers and semiconducting epilayers and (**b**) a high-magnification AADF-STEM image of the Ni/AlGaN interface, showing the presence of an interfacial layer. (**c**) Reconstructed APT data collected near gate metal stack/semiconducting epilayer interface displaying the distinct Au, Ni, AlGaN, and GaN layers. The boxed area in part (**a**) indicates the approximate area analyzed by APT as shown in (**c**)

voltage. The forward gate characteristics of the Ni/Au gate contact appeared very leaky after the stress, and the Schottky height reduced from 1.09 to 0.66 V after stress.

The interface between Ni/Au gate metal stack and the underlying semiconductor layer from devices fabricated on structure "B" wafers showed an oxide layer present prior to performing off-state stress. Focused ion beam milling, described elsewhere, was used to prepare samples in order to analyze the gate oxide of the devices by HAADF-STEM and laser-assisted atom probe tomography [51]. Laser-assisted APT was performed using an Imago local electrode atom probe (LEAP) 3000X-Si system. Semi-insulating materials can be examined through the use of a pulsed laser [51]. Figure 13.11 shows the Au, Ni, O, Al, N, and Ga atoms in a reconstruction of the sample. By analyzing the one-dimensional concentration profiles of these elements, it is possible to see that there is an O peak at the interface between the Ni and AlGaN layers. Complete mass-to-charge spectrum

indicates peaks at 21.5 and 74.0, which can be identified with $27AlO^{++}$ and $58NiO^{+}$. Since these peaks do not coincide with the $O^{+}$ profile, it can be concluded that AlOx and NiOx are present at the interface. Due to the instability of Ni with oxygen, consumption of this oxide layer, and even diffusion of oxygen, has been observed after off-state stress, leading to device degradation. Conversely, Pt as a gate metal has been shown to be stable and enhance device reliability due largely in part to its stability in the presence of oxygen.

## 13.8 RF Stress Tests

Devices with a 0.125-μm gate length fabricated on structure "A" wafers were biased under Class AB conditions with a quiescent drain current (IDQ) of 200 mA/mm at drain bias (VDS) of 10, 20, and 25 V. RF stress up to 350 h was performed with an AccelRF stress system at 10 GHz, driven into 3 dB compression and a baseplate temperature of 30°C. As can be seen in Fig. 13.12, both POUT and IDS showed little to no degradation for drain bias of 10 and 20 V. However, rapid permanent degradation is present at a drain bias of 25 V. A small recovery in both $P_{OUT}$ and $I_{DS}$ is present after pausing at ~25 h to obtain mid-stress device characteristics. Operating at a drain bias of 25 V is above the threshold necessary for the onset of degradation and is consistent with the observed critical voltage during dc stress. Small degradation in saturated drain current ($I_{DSS}$), less than 10%, occurred for all three drain bias conditions after stress and is consistent with prior reports on RF reliability [53]. For the lowest drain bias of 10 V, a small increase (~3%) in PAE and $I_{DS}$ (~1 mA) was observed during RF stressing for 350 h and may indicate that device burn-in is necessary in order to stabilize and improve RF device performance at lower bias conditions. Similarly, minimal degradation in PAE, $P_{OUT}$, gain, and transconductance occurred for a drain bias of 20 V. An increase in threshold voltage ($V_{TH}$) was observed for all stress conditions, with the largest $\Delta V_{TH}$ of 0.37 V at a drain bias of 25 V, indicating significant degradation of the Schottky contact. High dc bias stress of these devices, as well as others, has shown a consumption of the oxide layer present between the Schottky contact and underlying semiconductor layer and has been correlated to an increase in threshold voltage and decrease in $I_{DSS}$ [54].

Additionally, the Schottky barrier height reduced from 673 to 602 mV after stress at $V_{DS} = 25$ V. Even though device characteristics for drain bias at or below 20 V do not show significant RF degradation, the gate current–voltage sweeps with $V_{DS} = 0$ V indicate that the Schottky contact suffered degradation, with almost a 2 order of magnitude increase in gate leakage current. Even though the stress at a drain bias of 25 V was performed for a shorter duration, only ~85 h, gate leakage current shows the largest increase with almost a 3 order of magnitude rise in current.

Prior to stress, devices exhibited uniform EL emission under forward bias. However, points of gate leakage appeared with gate bias of −10 V. After RF stress, gate leakage paths were not present on the same channel as present prior to stress

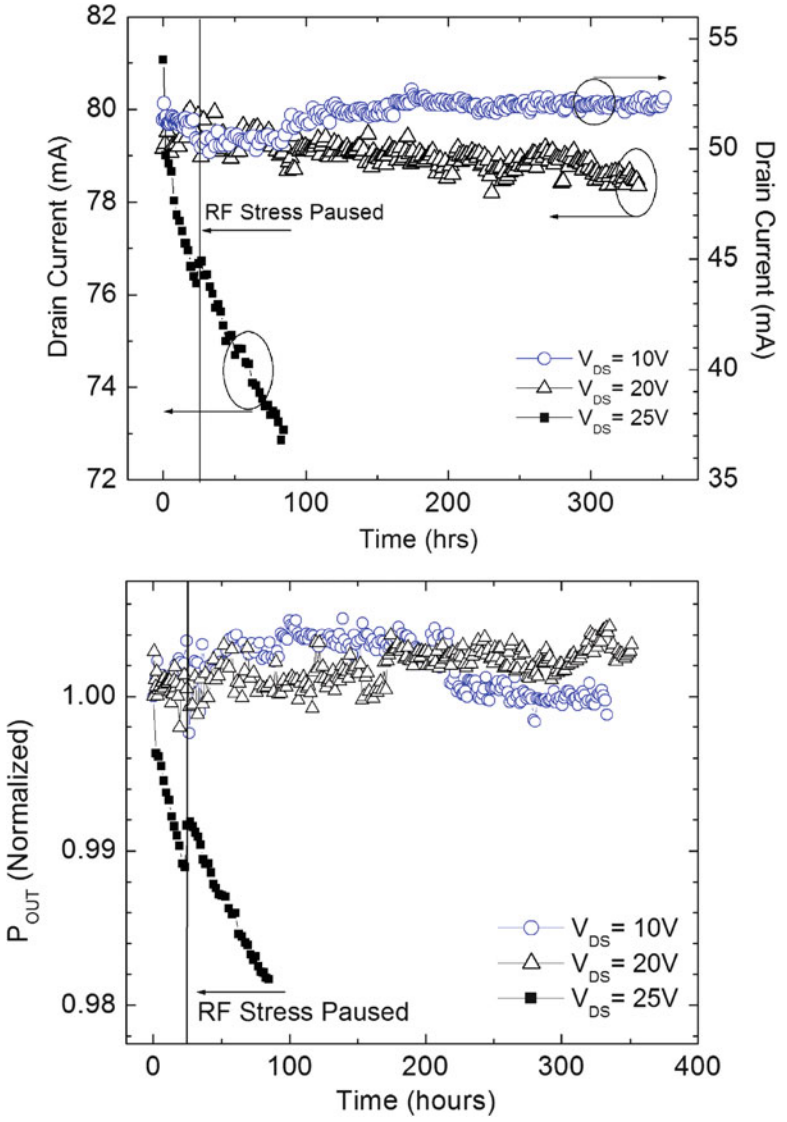



**Fig. 13.12** (***Top***) Drain current and (***bottom***) output power during RF stress at 10 GHz under 3-dB compression of 0.125-μm gate length AlGaN/GaN HEMT (Reprinted with permission from Electrochem and Solid State Lett 14, H464 (2011). Reproduced by permission of ECS-The Electrochemical Society)

(Fig. 13.13) [54]. It appears from EL emission that degradation due to RF stress may be localized along the channel on the right. Additionally, EL emission may not be a beneficial technique to prescreen devices in order to indicate possible regions of failure.

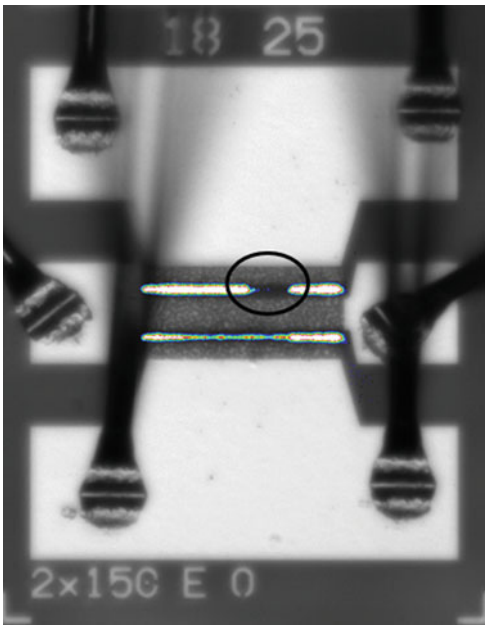



**Fig. 13.13** Electroluminescence of AlGaN/GaN HEMT after stress at $V_{DS} = 25$ V and Class AB operation. Circled region indicates area of channel that has an increase in non-radiative trap formation after stress. Uniform EL emission during forward bias was observed prior to stress. Photoluminescence spectra taken at circled region

In conclusion, submicron gate length AlGaN/GaN HEMTs show excellent reliability up to a drain bias of 20 V; however, the Schottky contact properties show substantial degradation in terms of threshold voltage shift, gate leakage current, and Schottky barrier height increase. This indicates that the gate contact is the cause for device failure under long-term life tests. Additionally, RF stressing at a drain bias of 25 V is above the threshold necessary for the onset of degradation and resulted in rapid, permanent degradation in $P_{OUT}$ and $I_{DS.}$ EL and PL indicate an increase in non-radiative trap formation after RF stress.

## 13.9 Conclusions

GaN-based HEMTs are being inserted into numerous applications ranging from communication base stations to satellites. Though they are an extremely attractive candidate to replace the present day leading materials for power applications, there is still significant uncertainty with respect to the reliability of GaN HEMTs while under high-power operation. The purpose of this research was to understand the driving degradation mechanisms in AlGaN/GaN HEMTs under electrical and thermal stress conditions for both dc and RF operation.

The effect of device design on self-heating was investigated for high-power applications. The maximum channel temperature in AlGaN/GaN HEMTs was shown to increase when the die size is reduced below a critical distance, $d$, from the gate finger. This distance is significantly dependent on the substrate material, with the largest critical distance of 300 μm for the sapphire substrate. Deviation of temperature between individual gate fingers and along the width of the gates was observed, especially as the number of gate fingers increased. A substantial variation in peak temperatures between 2-D and 3-D simulations was also observed, indicating the need for three-dimensional analysis for accurate reliability and MTTF predictions.

The effect of on-state, high-power step stress was investigated on several device structures, including 0.17-μm gate length HEMTs, gateless HEMT structures, and TLM structures with 5-μm spacing. Permanent degradation was observed in the 0.17-μm gate length HEMTs at relatively low drain bias voltages. However, temperature-dependent stress tests revealed that permanent degradation was dependent on channel temperatures reaching 195°C. Conversely, TLM structures, in which there is no Schottky contact, exhibited exceptional stability up to 25 V bias even though current densities, thus channel temperatures, reached much higher values. Therefore, the Schottky contact is the likely cause for permanent degradation. Breakdown voltages for all three structures indicated that catastrophic failure was not due to channel temperatures, as peak channel temperatures varied significantly (110–310°C) at breakdown. There are likely multiple degradation drivers present under these stress conditions, and future tests will be required to isolate the effect of current, voltage, and temperature under high-power conditions.

High reverse gate bias step stress from $-10$ to $-42$ V was performed on AlGaN/GaN HEMTs and resulted in a large increase in gate leakage current, with a sharp one order of magnitude increase in current at the critical voltage. The critical voltage of 0.14-μm gate length devices at room temperature was observed to be $-30$ V. Over 20 devices were step-stressed at temperatures ranging from 25 to 150°C, exhibiting an activation energy of about 42 meV for the critical voltage. As the stress temperature of the devices increased, the critical voltage was found to decrease linearly. This is due to the breakdown of the interfacial oxide layer from electric field-induced bond breakage resulting in the consumption of the oxide interface between the Ni/Au Schottky contact and the GaN cap. Disassociation of oxygen from Ga under high electric fields results in diffusion of oxygen to form $NiO_x$, likely enhanced by thermal and strain effects. Gate leakage current at $V_{CRI}$ was similar ($\sim 10^{-7}$ A) regardless of stress temperature, indicating similar leakage paths for all stress temperatures. Future work will further investigate the interfacial oxide breakdown by performing both step stress and steady state TDDB voltage testing on AlGaN/GaN HEMTs with unintentional and intentional gate oxides.

Submicron gate length AlGaN/GaN HEMTs were stressed under Class AB RF conditions with drain bias of 10, 20, and 25 V. AlGaN/GaN HEMTs show excellent reliability up to a drain bias of 20 V; however, the Schottky contact properties show substantial degradation in terms of threshold voltage shift, gate leakage current, and

Schottky barrier height increase. This indicates that the gate contact is the cause for device failure under long-term life tests. Additionally, RF stressing at a drain bias of 25 V is above the threshold necessary for the onset of degradation and resulted in rapid, permanent degradation in $P_{OUT}$ and $I_{DS}$.

The degradation of AlGaN/GaN HEMTs stressed under high electric field conditions with various device structures and designs was studied. The presence of an interfacial oxide layer was investigated through both TEM and HAADF-STEM and laser-assisted atom probe tomography. This oxide layer is seen to be reactive with the Ni from the Ni/Au gate metal stack after high reverse gate bias stress. In some instances, oxygen was observed diffusing into the underlying AlGaN layer and further through threading dislocations that terminated below the Schottky contact. Both DC and RF high electric field stress indicated a decrease in Schottky barrier height and large increases in gate leakage current. However, a small decrease in saturate drain current was observed in all instances. Devices with Pt/Ti/Au gate metal scheme showed superior stability up to gate voltages of −100 V. These studies indicate that the reliability of AlGaN/GaN HEMTs can be significantly improved for high electric field applications through the use of much more stable gate metals, such as Pt.

**Acknowledgments** The work performed at UF is supported by an AFOSR MURI monitored by James Hwang and Gregg Jessen.

## References

1. G. Meneghesso, G. Verzellesi, F. Danesin, F. Rampazzo, F. Zanon, A. Tazzoli, M. Meneghini, E. Zanoni, Reliability of GaN high-electron-mobility transistors: state of the art and perspectives. IEEE Trans. Device Mater. Reliab. **8**, 332–343 (2008)
2. S. Singhal, J.C. Roberts, P. Rajagopal, T. Li, A.W. Hanson, R. Therrien, J.W. Johnson, I.C. Kizilyalli, K.J. Linthicum, GaN-ON-Si failure mechanisms and reliability improvements, in *Reliability Physics Symposium Proceedings, 2006. 44th Annual., IEEE International*, *2006* Boston, MA, pp. 95–98 (2006)
3. A. Sozza, C. Dua, E. Morvan, B. Grimber, S.L. Delage, A 3000 hours DC life test on AlGaN/GaN HEMT for RF and microwave applications. Microelectron. Reliab. **45**, 1617–1621 (2005)
4. E. Zanoni, G. Meneghesso, M. Meneghini, A. Tazzoli, N. Ronchi, A. Stocco, F. Zanon, A. Chini, G. Verzellesi, A. Cetronio, C. Lanzieri, and M. Peroni, Long-term stability of Gallium Nitride high electron mobility transistors: a reliability physics approach, in *Microwave Integrated Circuits Conference, 2009. EuMIC 2009 Strasbourg, France. European*, pp. 212–217 (2009)
5. M. Asif Khan, High electron mobility transistor based on a GaN-AlxGa1-xN heterojunction. Appl. Phys. Lett. **63**, 1214 (1993)
6. C. Rivera, The role of electric field-induced strain in the degradation mechanism of AlGaN/GaN high-electron-mobility transistors. Appl. Phys. Lett. **94**, 053501 (2009)
7. M. Dammann, A. Leuther, F. Benkhelifa, T. Feltgen, W. Jantz, Reliability and degradation mechanism of AlGaAs/InGaAs and InAlAs/InGaAs HEMTs. Physica Status Solidi A **195**, 81–86 (2003)
8. M. Piazza, C. Dua, M. Oualli, E. Morvan, D. Carisetti, F. Wyczisk, Degradation of TiAlNiAu as ohmic contact metal for GaN HEMTs. Microelectron. Reliab. **49**, 1222–1225 (2009)

9. U. Chowdhury, J.L. Jimenez, C. Lee, E. Beam, P. Saunier, T. Balistreri, S.Y. Park, T. Lee, J. Wang, M.J. Kim, J. Joh, J.A. del Alamo, TEM observation of crack- and pit-shaped defects in electrically degraded GaNHEMTs. IEEE Electron Device Lett. **29**, 1098–1100 (2008)
10. J.A. del Alamo, J. Joh, GaN HEMT reliability. Microelectron. Reliab. **49**, 1200–1206 (2009)
11. J. Joh, J.A. del Alamo, Mechanisms for electrical degradation of GaN high-electron mobility transistors, in *Electron Devices Meeting, 2006. IEDM '06. International, 2006* San Francisco, CA, pp. 1–4 (2006)
12. M. Faqir, G. Verzellesi, A. Chini, F. Fantini, F. Danesin, G. Meneghesso, E. Zanoni, C. Dua, Mechanisms of RF current collapse in AlGaN-GaN high electron mobility transistors. IEEE Trans. Device Mater. Reliab. **8**, 240–247 (2008)
13. A. Sarua, Piezoelectric strain in AlGaN/GaN heterostructure field-effect transistors under bias. Appl. Phys. Lett. **88**, 103502 (2006)
14. G. Meneghesso, Hot carrier effects in AlGaAs/InGaAs high electron mobility transistors: failure mechanisms induced by hot carrier testing. J. Appl. Phys. **82**, 5547 (1997)
15. M.J. Uren, D.G. Hayes, R.S. Balmer, D.J. Wallis, K.P. Hilton, J.O. Maclean, T. Martin, C. Roff, P. McGovern, J. Benedikt, P.J. Tasker, Control of Short-Channel Effects in GaN/AlGaN HFETs, in *European Microwave Integrated Circuits Conference, 2006. The 1st, 2006*. Rome, Italy, pp. 65–68 (2006)
16. M. Buchta, K. Beilenhoff, H. Blanck, J. Thorpe, R. Behtash, S. Heckmann, H. Jung, Z. Ouarch, M. Camiade, GaN technologies and developments: status and trends, in *Microwave and Millimeter Wave Technology (ICMMT), 2010 International Conference on, 2010* Chicago, ILL., pp. 488–491 (2010)
17. M.A. Mastro, J.R. LaRoche, N.D. Bassim, C.R. Eddy, Simulation on the effect of non-uniform strain from the passivation layer on AlGaN/GaN HEMT. Microelectron. J. **36**, 705–711 (2005)
18. D.K. Schroder, *Semiconductor Material and Device Characterization*, 2006th edn. (IEEE Press/Wiley, Piscataway/Hoboken, 2006)
19. O. Oda, *Compound Semiconductor Bulk Materials and Characterizations* (World Scientific, Singapore/Hackensack, 2007)
20. L. Brillson, Nanoscale luminescence spectroscopy of defects at buried interfaces and ultrathin films. J. Vac. Sci. Technol. B **19**, 1762 (2001)
21. M. Micovic, A. Kurdoghlian, P. Hashimoto, M. Hu, M. Antcliffe, P. J. Willadsen, W. S. Wong, R. Bowen, I. Milosavljevic, A. Schmitz, M. Wetzel, and D. H. Chow, GaN HFET for W-band power applications, in *Electron Devices Meeting, 2006. IEDM '06. International, 2006* Washington, DC, pp. 1–3 (2006)
22. S. Arulkumaran, High-temperature effects of AlGaN/GaN high-electron-mobility transistors on sapphire and semi-insulating SiC substrates. Appl. Phys. Lett. **80**, 2186 (2002)
23. S.A. Vitusevich, S.V. Danylyuk, N. Klein, M.V. Petrychuk, A.Y. Avksentyev, V.N. Sokolov, V.A. Kochelap, A.E. Belyaev, V. Tilak, J. Smart, A. Vertiatchikh, L.F. Eastman, Separation of hot-electron and self-heating effects in two-dimensional AlGaN/GaN-based conducting channels. Appl. Phys. Lett. **82**, 748–750 (2003)
24. S.A. Vitusevich, A.M. Kurakin, N. Klein, M.V. Petrychuk, A.V. Naumov, A.E. Belyaev, AlGaN/GaN high electron mobility transistor structures: self-heating effect and performance degradation. IEEE Trans. Device Mater. Reliab. **8**, 543–548 (2008)
25. R. Gaska, A. Osinsky, J.W. Yang, M.S. Shur, Self-heating in high-power AlGaN-GaN HFETs. IEEE Electron Device Lett. **19**, 89–91 (1998)
26. Y.C. Chou, D. Leung, I. Smorchkova, M. Wojtowicz, R. Grundbacher, L. Callejo, Q. Kan, R. Lai, P.H. Liu, D. Eng, A. Oki, Degradation of AlGaN/GaN HEMTs under elevated temperature lifetesting. Microelectron. Reliab. **44**, 1033–1038 (2004)
27. M. Kuball, Measurement of temperature distribution in multifinger AlGaN/GaN heterostructure field-effect transistors using micro-Raman spectroscopy. Appl. Phys. Lett. **82**, 124 (2003)

28. T. Batten, A. Manoi, M.J. Uren, T. Martin, M. Kuball, Temperature analysis of AlGaN/GaN based devices using photoluminescence spectroscopy: challenges and comparison to Raman thermography. J. Appl. Phys. **107**, 074502–074502-5 (2010)
29. T. Batten, J.W. Pomeroy, M.J. Uren, T. Martin, M. Kuball, Simultaneous measurement of temperature and thermal stress in AlGaN/GaN high electron mobility transistors using Raman scattering spectroscopy. J. Appl. Phys. **106**, 094509–094509-4 (2009)
30. J. Kim, J.A. Freitas Jr., J. Mittereder, R. Fitch, B.S. Kang, S.J. Pearton, F. Ren, Effective temperature measurements of AlGaN/GaN-based HEMT under various load lines using micro-Raman technique. Solid-State Electron. **50**, 408–411 (2006)
31. R. Aubry, J.C. Jacquet, C. Dua, H. Gerard, B. Dessertenne, M.A. di Forte-Poisson, Y. Cordier, S.L. Delage, Thermal characterisation of AlGaN/GaN HEMTs using micro-Raman scattering spectroscopy and pulsed I-V measurements, in *Silicon Carbide and Related Materials 2003* Nice, France, Pts 1 and 2, vol. 457–460, pp. 1625–1628 (2004)
32. C. Chih-Yang, E.A. Douglas, K. Jinhyung, L. Liu, L. Chien-Fong, C. Byung-Hwan, D.J. Cheney, B.P. Gila, F. Ren, G.D. Via, D.A. Cullen, Z. Lin, D.J. Smith, J. Soohwan, S.J. Pearton, Electric-field-driven degradation in off-state step-stressed AlGaN/GaN high-electron mobility transistors. IEEE Trans. Device Mater. Reliab. **11**, 187–193 (2011)
33. R. Coffie, Y. Chen, I.P. Smorchkova, B. Heying, V. Gambin, W. Sutton, Y.C. Chou, W.B. Luo, M. Wojtowicz, A. Oki, Temperature and voltage dependent RF degradation study in Algan/gan HEMTs, in *Reliability Physics Symposium, 2007*. Toronto, Canada. *Proceedings. 45th Annual. IEEE International*, pp. 568–569 (2007)
34. S. Demirtas, J.A. del Alamo, Critical voltage for electrical reliability of GaN high electron mobility transistors on Si substrate, in *Reliability of Compound Semiconductors Digest (ROCS), 2009*. Boston, MA, pp. 53–56 (2009)
35. S. Demirtas, J. Joh, J.A. del Alamo, High voltage degradation of GaN high electron mobility transistors on silicon substrate. Microelectron. Reliab. **50**, 758–762 (2010)
36. J. Joh, J.A. del Alamo, J. Jimenez, A simple current collapse measurement technique for GaN high-electron mobility transistors. IEEE Electron Device Lett. **29**, 665–667 (2008)
37. J.W. Joh, J.A. del Alamo, Critical voltage for electrical degradation of GaN high-electron mobility transistors. IEEE Electron Device Lett. **29**, 287–289 (2008)
38. J. Joh, F. Gao, T. Palacios, J.A. del Alamo, A model for the critical voltage for electrical degradation of GaN high electron mobility transistors. Microelectron. Reliab. **50**, 767–773 (2010)
39. C.-F. Lo, L. Liu, T.-S. Kang, R. Davies, B.P. Gila, S.J. Pearton, I.I. Kravchenko, O. Laboutin, Y. Cao, W.J. Johnson, F. Ren, Improvement of off-state stress critical voltage by using Pt-Gated AlGaN/GaN high electron mobility transistors. Electrochem. Solid-State Lett. **14**, H264–H267 (2011)
40. E. Zanoni, F. Danesin, M. Meneghini, A. Cetronio, C. Lanzieri, M. Peroni, G. Meneghesso, Localized damage in AlGaN/GaN HEMTs induced by reverse-bias testing. IEEE Electron Device Lett. **30**, 427–429 (2009)
41. P. Ivo, A. Glowacki, R. Pazirandeh, E. Bahat-Treidel, R. Lossy, J. Wurfl, C. Boit, G. Trankle, Influence of GaN cap on robustness of AlGaN/GaN HEMTs, in *Reliability Physics Symposium, 2009*. Portland, OR. *I.E. International*, pp. 71–75 (2009)
42. T. Beechem, A. Christensen, S. Graham, D. Green, Micro-Raman thermometry in the presence of complex stresses in GaN devices. J. Appl. Phys. **103**, 124501–124501-8 (2008)
43. J. Jungwoo, X. Ling, J.A. del Alamo, Gate current degradation mechanisms of GaN high electron mobility transistors, in *Electron Devices Meeting, 2007. IEDM 2007. Washington, DC. IEEE International*, pp. 385–388 (2007)
44. J. Joh, J.A. del Alamo, K. Langworthy, S. Xie, T. Zheleva, *Role of stress voltage on structural degradation of GaN high-electron-mobility transistors*. Microelectron. Reliab. Corrected Proof (In Press)
45. J.D. Guo, F.M. Pan, M.S. Feng, R.J. Guo, P.F. Chou, C.Y. Chang, Schottky contact and the thermal stability of Ni on n-type GaN. J. Appl. Phys. **80**, 1623–1627 (1996)

46. N. Miura, T. Nanjo, M. Suita, T. Oishi, Y. Abe, T. Ozeki, H. Ishikawa, T. Egawa, T. Jimbo, Thermal annealing effects on Ni/Au based Schottky contacts on n-GaN and AlGaN/GaN with insertion of high work function metal. Solid-State Electron. **48**, 689–695 (2004)
47. H. Jung, R. Behtash, J.R. Thorpe, K. Riepe, F. Bourgeois, H. Blanck, A. Chuvilin, U. Kaiser, Reliability behavior of GaN HEMTs related to Au diffusion at the Schottky interface. Physica Status Solidi (C) **6**, S976–S979 (2009)
48. W. Gotz, N.M. Johnson, C. Chen, H. Liu, C. Kuo, W. Imler, Activation energies of Si donors in GaN. Appl. Phys. Lett. **68**, 3144–3146 (1996)
49. T. Mattila, R.M. Nieminen, Ab initio study of oxygen point defects in GaAs, GaN, and AlN. Phys. Rev. B **54**, 16676 (1996)
50. J.C. Zolper, R.G. Wilson, S.J. Pearton, R.A. Stall, Ca and O ion implantation doping of GaN. Appl. Phys. Lett. **68**, 1945–1947 (1996)
51. M.R. Holzworth, N.G. Rudawski, S.J. Pearton, K.S. Jones, L. Lu, T.S. Kang, F. Ren, J.W. Johnson, Characterization of the gate oxide of an AlGaN/GaN high electron mobility transistor. Appl. Phys. Lett. **98**, 122103–3 (2011)
52. Lu Liu, T. Kang, D. Cullen, L. Zhou, J. Kim, C. Chang, E. Douglas, S. Jang, D.J. Smith, S. J. Pearton, J.W. Johnson, F. Ren, Effect of source field plate on the characteristics of off-state, step-stressed AlGaN/GaN high electron mobility transistors. J. Vac. Sci. Technol. **B 29**, 032204 (2011)
53. A. Chini, V. Di Lecce, M. Esposto, G. Meneghesso, E. Zanoni, RF degradation of GaN HEMTs and its correlation with DC stress and I-DLTS measurements, in *Microwave Integrated Circuits Conference, 2009. EuMIC 2009*. Lille, France. *European*, pp. 132–135 (2009)
54. E.A. Douglas, S.J. Pearton, B. Poling, G.D. Via, L. Liu, F. Ren, Electrochem. Solid-State Lett. **14**, H464 (2011)

# Chapter 14
# GaAs Device Reliability: High Electron Mobility Transistors and Heterojunction Bipolar Transistors

**F. Ren, E.A. Douglas, and Stephen J. Pearton**

**Abstract** The two main GaAs-based electronic device technologies are high electron mobility transistors (HEMTs) and heterojunction bipolar transistors (HBTs). Both technologies are commercialized for use in low-noise amplifiers, radar, and fiber optic data transmission systems. In this chapter, we will summarize the degradation mechanisms that limit the lifetime of these devices. A variety of contact and surface degradation mechanisms have been reported but differ in the two device technologies – for HEMTs, the layers are thin and relatively lightly doped compared to HBT structures, and there is a metal Schottky gate that is directly on the semiconductor. By contrast, the HBT relies on pn junctions for current modulation and has only ohmic contacts. This leads to different degradation mechanisms for the two types of devices.

## 14.1 Introduction

The basic unit of failure in Si device technology is the FIT (failure unit), defined as 1 failure/$10^9$ device hours. For 100 devices on test, a failure rate of 1,000 FIT would mean there would be only 1 failure in 1 year. Given that a relatively small number of devices will actually show failure, it is critical to both enhance the failure rate through accelerated testing (the five common stresses used are temperature, voltage, current, humidity, and temperature cycling) and treat the resulting reliability data correctly. Under accelerated aging conditions, different failure mechanisms may be accelerated by different amounts for the same applied stress.

F. Ren
Department of Chemical Engineering, University of Florida, Gainesville, FL 32606, USA

E.A. Douglas • S.J. Pearton (✉)
Department of Materials Science and Engineering, University of Florida, Gainesville, FL 32606, USA
e-mail: spear@mse.ufl.edu

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_14,
© Springer Science+Business Media New York 2013

A quick look at the global reliability data for compound devices shows no obvious improvement in 20 years [1]. Is this an indictment of our industry or the fact we have always been moving to the "next" device technology? In the past 40 years, the focus of GaAs-based device technologies shifted from metal–semiconductor field-effect transistors (MESFETs) to various types of high electron mobility transistors (HEMTs) and then to heterojunction bipolar transistors (HBTs). Much of this was driven by the heavy emphasis on defense funding for the necessary research, with few large commercial applications until the cell phone market exploded. By contrast, in the Si industry, the basic device has always been the metal-oxide-semiconductor field-effect transistor (MOSFET). In the initial work on GaAs MESFETs, the main reliability issues were due to the reaction of gate metals with the underlying semiconductor, typically referred to as gate sinking, and also the need for effective surface passivation to prevent creation of surface states that lead to unwanted depletion effects and carrier trapping. With the shift in emphasis to epitaxial layered structures such as HEMTs and HBTs in the 1990s, issues such as dopant diffusion (especially the base dopant in HBTs) and hydrogen passivation and contamination (in both HEMTs and HBTs) became important. As the requirements for high-power operation and integration level increased, field-driven failure mechanism related to trap creation, surface breakdown, and contact failure were all identified as being important, and in addition, the failure rate of passive components began to be an issue. With the introduction of GaN-based transistors as well as the continued maturation of the GaAs technologies, defect reduction and a focus on improved capacitors and interconnects were areas of focus. The continual introduction of new materials in compound semiconductor technologies was a drawback compared to Si, where problems such as mobile ion migration, electromigration, time-dependent dielectric breakdown, corrosion, electrostatic discharge, and soft errors could be systematically solved and protocols and models developed for their understanding and mitigation. This allowed the Si industry to focus on designing and building-in reliability at the wafer level and to introduce defect reduction efforts and strict process controls. The recent introduction of new materials in Si technologies, such as Cu for interconnects and high- and low-K materials for gates and interconnects, probably means some new areas of reliability concern, but the huge experience base should minimize unexpected problems.

In Si MOS devices, time-dependent failure mechanisms include surface charge accumulation or injection, dielectric breakdown, electromigration, contact degradation, and corrosion due to contamination. In compound semiconductors, there are added issues of local regions of non-stoichiometry that affect field distributions and increase recombination, oxidation of AlGaAs or AlGaN, high densities of dislocations, and other extended defects in some structures and high surface state densities. In GaAs devices, it is important to pay close attention to voltage and current acceleration stress mechanisms. Many studies in Si indicate that the reaction rate of the failure mechanism is proportional to a power of the applied voltage as well as temperature, that is,

$$R(T,V) = R_O(T)V^{\gamma(T)} \tag{14.1}$$

where the coefficient $R_O(T)$ is an Arrhenius function of $T$ and the power dependence varies between 1 and 4.5. This determines how much acceleration upon increasing the bias voltage used during stressing. If dielectric breakdown is the dominant failure mode, then at a given field, a fraction of the devices will fail in a short time, with no additional failures until an increased field is applied.

## 14.2 Typical Accelerated Stress Protocols

While GaAs reliability studies have typically focused on temperature enhancement of failure to get adequate statistics (the so-called 3 temperature test is common), it is clearly necessary to use voltage and current enhancement. In this case, devices are tested at different temperatures, and an activation energy ($E_a$) is extracted from the relationship

$$\mathrm{Ln}(t_1/t_2) = E_a/k(1/T_1 - 1/T_2) \tag{14.2}$$

where $t_{1,2}$ are time to failure at temps $T_{1,2}$ and $k$ is Boltzmann's constant.

Some typical approaches to stressing at fixed temperature are illustrated in Fig. 14.1, which shows the so-called step stress approach, in which the parameter of interest such as voltage, current, or rf power is stepped in some fashion, either monotonically or with a recovery period in between. The latter types of stressing are useful for detecting the presence of traps that empty during the recovery period.

When one simply uses temperature as the accelerant, serious errors in estimating device lifetime are possible. Figure 14.2 highlights the problem of using temperature acceleration as the only reliability test [2]. If there is more than one degradation mechanism present in the device, testing at high temperature and then extrapolating back to the normal device operating temperature may completely miss the real mechanism that limits the device lifetime at that particular temperature. This seems to be generally understood, and there is increasing use of voltage, current, and rf stressing at temperatures closer to the actual device operating temperature. It should be kept in mind that actual field returns may be dominated by other factors, including ESD, capacitor defects, and assembly/packaging issues, rather than device degradation.

In compound semiconductor device technology, it is being increasingly realized that one must look for not only the dominant factor causing degradation but also whether there is recovery during the off-state (trap generation), the role of the ambient (in case of oxidation or hydrogen effects), and the electric field distributions (including inspection of nonuniformity in gate dimensions). A complete description of the device degradation might also include visual inspection, dc, rf, base noise spectra ($f$, $V$), and check for high leakage. It is common to define failure as a change of only 10–20% in some key performance parameter, such as, for example, a drop of this magnitude in HBT gain.

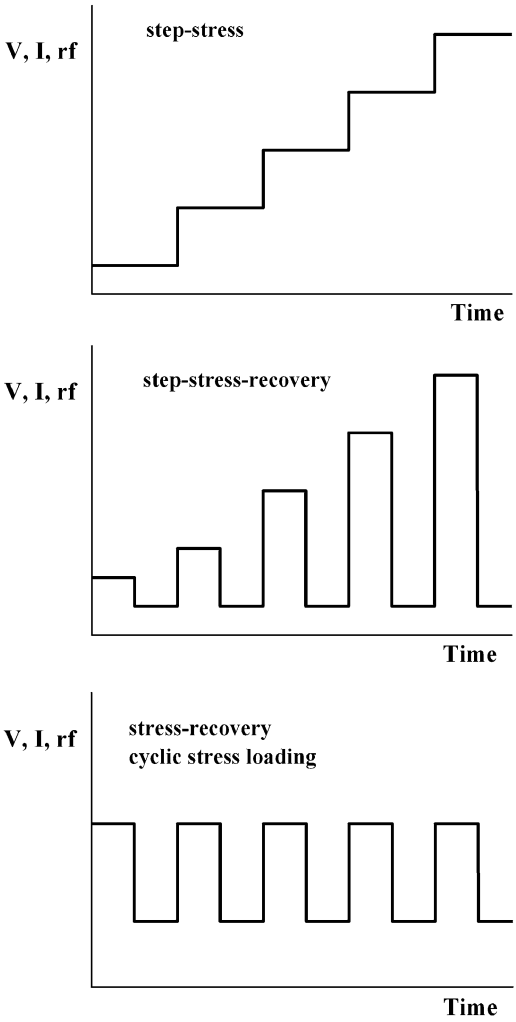



**Fig. 14.1** Examples of typical time dependence of stressing protocols for III-V electronic devices. The parameter that is varied with time can be voltage, current, rf power, among others

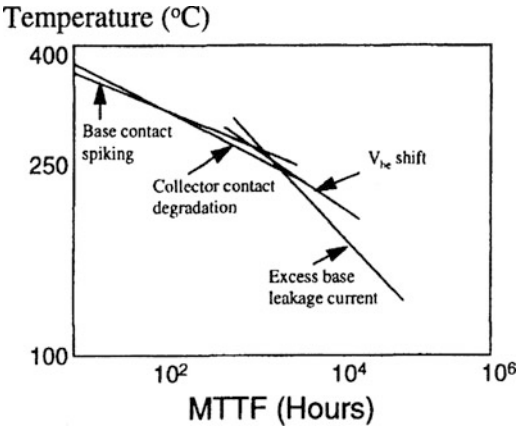



**Fig. 14.2** Mean time to failure of GaAs-based HBTs, showing how different degradation mechanisms dominate at different temperatures (After Henderson [2])

## 14.3 GaAs HEMTs

The most commonly reported degradation mechanisms for both GaAs- and InP-based HEMTs include contact problems (such as sinking gates in which the gate metal begins to react with the underlying semiconductor), creation of surface states (which can be manifested as what is commonly called gate lag), mechanical stress (the absorption of H into Ti-based Schottky gates can lead to compressive stress due to piezo effects in the semiconductor) [3], hot carrier-induced impact ionization at the gate edge or avalanche breakdown in the semiconductor, fluorine contamination, and corrosion (mainly related to Al oxidation) [3–29]. Metamorphic HEMT (MHEMT) technology has been developed using metamorphic buffer layers to grow InAlAs/InGaAs on larger diameter GaAs substrates to overcome the limitations of InP substrates – smaller wafer size, higher cost, and brittle nature. The commercial applications of MHEMTs are predominantly in low-noise mm-wave amplifiers for radio communications, automotive collision avoidance radar, and high-bit-rate fiber systems [18]. The choice of whether they can be used in place of InP-based HEMTs depends upon their DC/RF performance and the chip cost requirements. A number of studies have shown that MHEMTs can exhibit similar reliability to InP HEMTs, with over $10^6$ h mean-time-to-failure at 125°C [21, 22].

However, the InAlAs/InGaAs MHEMTs require a burn-in step to improve the device stability. Current device designs tend to suffer from device degradation, and a costly burn-in process is typically performed to make the device more stable and eliminate the early degradation when the devices are placed in service. The transistors are generally biased at certain gate and drain voltages for 24–60 h before sending the devices to customers. During the burn-in process, the drain current decreases and ohmic contact resistance increases with time and level off around 36 h. Minimizing the burn-in time or eliminating the burn-in step is highly desirable to reduce the device fabrication cost. In order to effectively identify the failure mechanisms, both a high-temperature storage test and DC stress were used in this study. Besides MHEMT devices themselves, transmission line method (TLM) patterns were also used to isolate the gate effect on the device degradation.

The MHEMTs were obtained from a commercial vendor. The Ti/Pt-based Schottky gate length was 150 nm, with 1.2 μm spacing between both gate/drain and gate/source. The devices had a two-finger design with gate width 75 μm. The TLM patterns also present on the device chip employed 45 × 70 μm pads with gaps of 3, 6, 9, 12, and 15 μm.

Typical dc characteristics of the MHEMTs prior to stressing showed that the maximum drain–source current density was 0.27 A/mm, with a gate current in the hundreds of nA range. The unity current gain, $f_t$, was 94 GHz, while the maximum frequency of oscillation, $f_{max}$, was 124 GHz.

The devices were stressed in one of two ways. Some of the MHEMTs were biased at a drain–source voltage of 3 V for 36 h at 165°C. Other devices were given

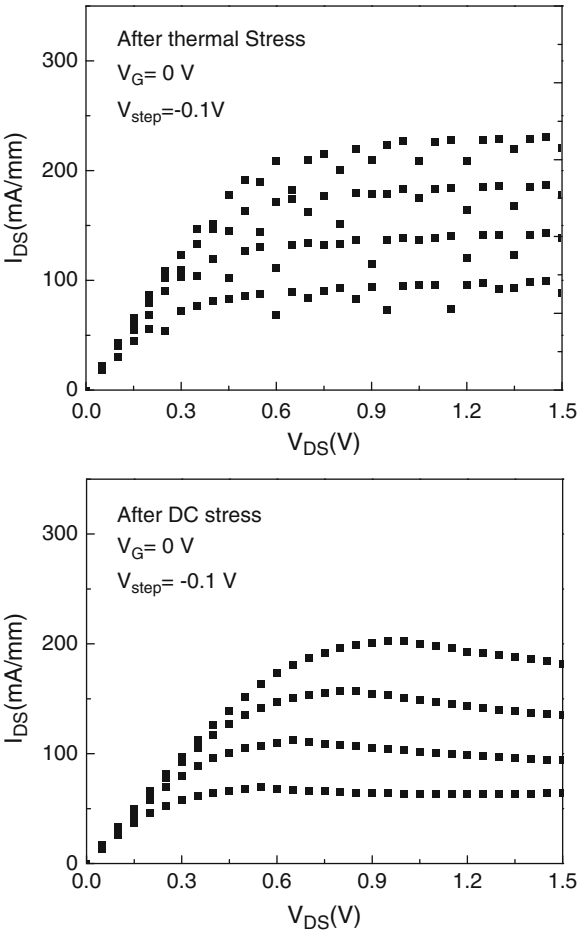



**Fig. 14.3** The drain current ($I_{DS}$) of the InAlAs/InGaAs MHEMT after heating stress at 250°C for 36 h as function of drain voltage ($V_{DS}$) (*top*) and the drain current ($I_{DS}$) of the InAlAs/InGaAs MHEMT after DC stress at 2.7 V for 36 h (*bottom*) (Reprinted with permission from [30]. Copyright 2010 American Vacuum Society)

a thermal storage test in an oven at 250°C for 36 h. The dc characteristics were measured before and after both kinds of stressing using an Agilent 4145 parameter analyzer. Some of the devices were also examined by cross-sectional transmission electron microscopy (TEM) to look for reactions of the contacts with the underlying semiconductor. Energy-dispersive X-ray spectroscopy (EDS) elemental analysis was performed to obtain the elemental profiles near the reacted contacts.

Figure 14.3 shows the drain current ($I_{DS}$) of the InAlAs/InGaAs MHEMT after both heating stress at 250°C for 36 h as function of drain voltage ($V_{DS}$) (top) and the drain current ($I_{DS}$) of the InAlAs/InGaAs MHEMT after DC stress at 3 V for 36 h (bottom). Both forms of stress lead to a reduction in drain current density of ~12.5% compared to the unstressed devices. Thus, after the burn-in process, the MHEMTs suffer from higher parasitic resistance.

To further investigate the origin of the degradation in drain–source current, the sheet resistance and specific contact resistance of the devices were obtained from TLM data as a function of thermal storage time and a function of constant bias voltage stress time. Using the TLMs to examine the increase of the parasitic resistance isolated the effect of the gate sinking on the degradation of drain–source current from the effect of ohmic metal contact degradation. As shown in Fig. 14.4, the total resistance of TLMs increases significantly with time in the first 12 h of thermal storage at 250°C, while the specific contact resistance increases much more than sheet resistance and shows that the contact between ohmic metal and semiconductor dominated the degradation during the thermal storage. Similar data for the constant current stressed devices is shown in Fig. 14.5. In this case, the sheet resistance increased around 18%, while specific contact resistance was reduced by 40%. The device resistance increase was dominated by changes in the sheet resistance instead of contact resistance.

TEM cross sections of a degraded thermal storage HEMT and a constant current stressed HEMT are illustrated in Fig. 14.6 (top) and Fig. 14.6 (bottom), respectively. Both samples showed metal spikes with the ohmic metal diffusing into epitaxial layer, which were formed during the high-temperature ohmic annealing. For the constant current stressed sample, the density of the spikes was higher around the edge of the source ohmic contact pad and drain ohmic contact pads, as illustrated in the Fig. 14.8 (top). Interestingly enough, the region of the high-density spikes in the TEM picture matched the estimated transfer length of the TLM measurement, as shown in top plot in Fig. 14.6. Thus, the formation of the high-density spikes could result from the current-induced electromigration.

The drain current density of the commercial power MHEMTS fabricated with the structure described here is around 0.5–1 A/mm. In the fabrication, the final metal often has a thickness of 4–6 μm, while the ohmic metal contact has a thickness of less than 0.3 μm. In these devices, the ohmic metal is alloyed with the semiconductor, and the resistance of the alloyed metal is also larger than the unalloyed metal stack. The 4–6-μm-thick final metal does not exhibit a problem with a current density of 0.5–1 A/mm. However, as shown by the TEM image of Fig. 14.7, the metal thickness of the region at the edge of the ohmic contact pad is too thin to sustain the current density to avoid electromigration. For example, the ohmic metal thickness for the device shown in Fig. 14.8 is roughly 250 nm. During the burn-in process, 20 mA was used to stress a device with gate width of 75 μm (20 mA/75 μm = 266 mA/mm). Therefore, the current density is given as 20 mA/(75 μm × 0.25 μm) = $1 \times 10^5$ A/cm$^2$, which is the current density through the ohmic pad. Such a high density of current flowing across the thin ohmic metal and then across the metal semiconductor interface into the semiconductor causes the ohmic metal diffusion during the burn-in process. This caused the electromigration-induced voids and the formation of additional metal spikes at the edge of the ohmic metal contact pads, as shown in Fig. 14.7, the source contact pad (left), and drain contact pad (right) after performing the constant current stressing.

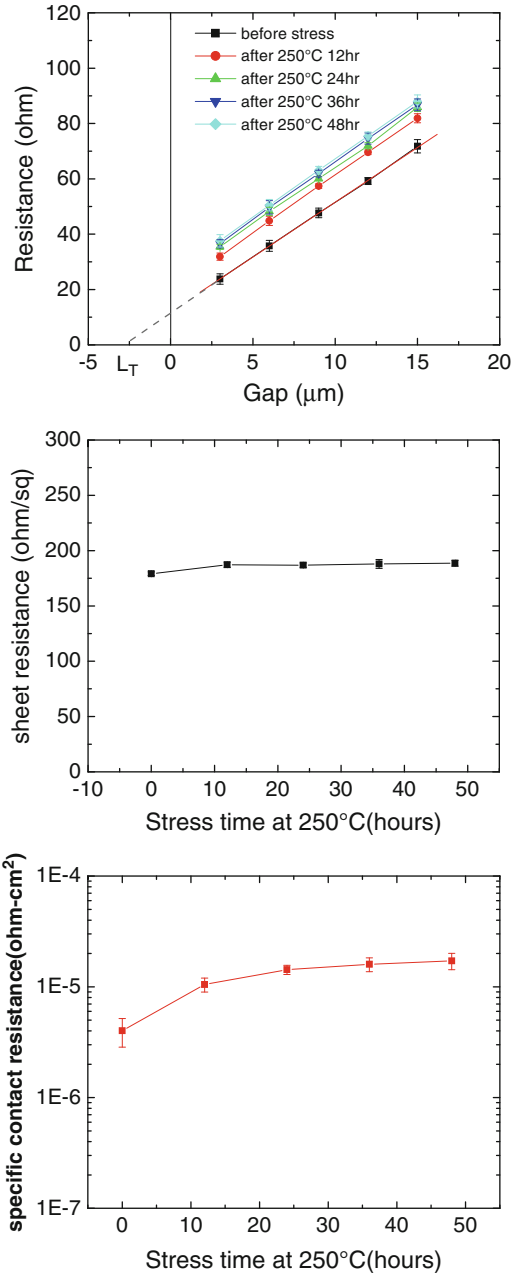



**Fig. 14.4** Data from a TLM pattern stored at 250°C for 48 h (*top*) resistance as a function of gap distance, (*middle*) sheet resistance, and (*bottom*) specific contact resistivity (Reprinted with permission from [30]. Copyright 2010 American Vacuum Society)

The gate ideality factor of the unstressed HEMT extracted from the gate current–voltage (I-V) characteristics was ~1.6, indicative of both recombination and thermionic emission electron transport mechanisms being present (Fig. 14.8). The gate characteristics of the thermal and DC-stressed HEMTs showed significant

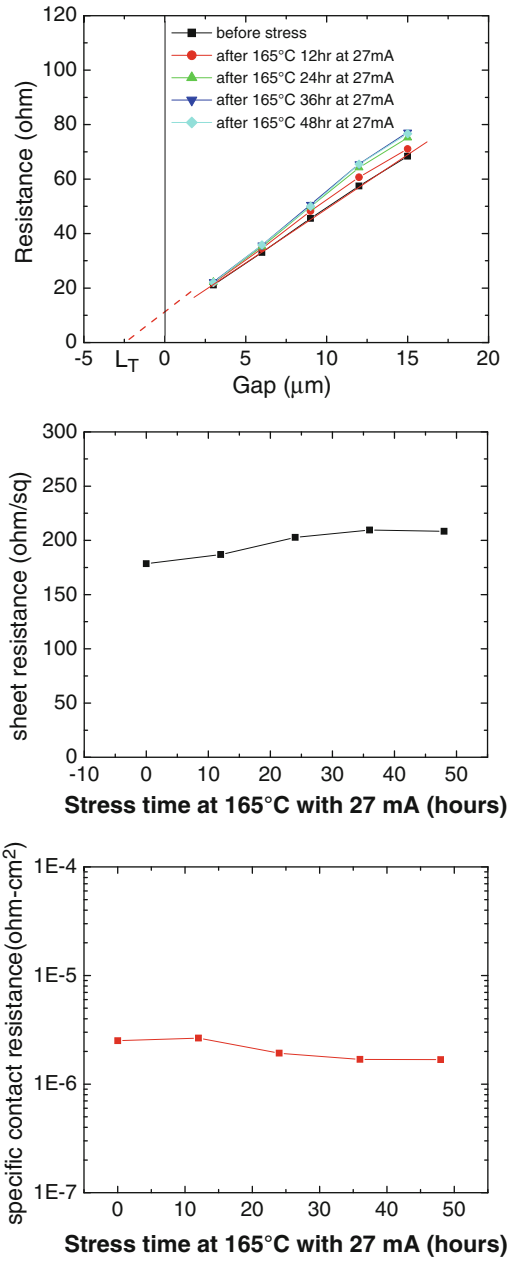



**Fig. 14.5** Data from a TLM pattern on a MHEMT stressed with a DC stress of 27 mA at 165°C – (*top*) resistance versus gap distance, (*middle*) sheet resistance, and (*bottom*) specific contact resistance (Reprinted with permission from [30]. Copyright 2010 American Vacuum Society)

degradation, and gate current increased several orders in both forward and reverse bias conditions. Low-magnification cross-section view TEM images of a Pt/Ti/Pt/Au mushroom gate after 165°C, $V_{DS}$ 3 V, $J_{DS}$ 300 mA/mm for 36 h showed that the bottom Pt of the Pt/Ti/Pt/Au mushroom gate diffused into the InAlAs gate contact

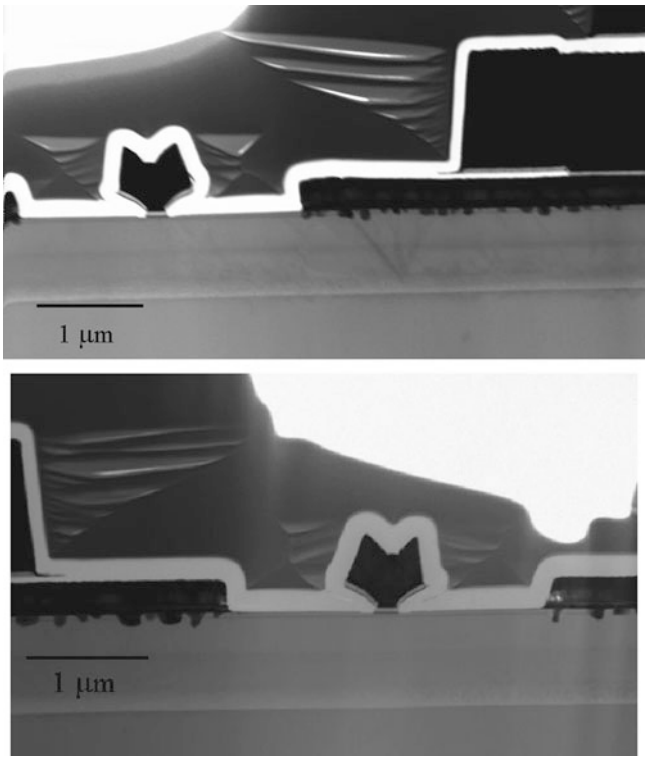



**Fig. 14.6** Low-magnification cross-section view TEM image of InAlAs/InGaAs MHEMT after stored at 250°C for 48 h. (*Top*) low-magnification cross-section view TEM image of InAlAs/InGaAs MHEMT after DC stress for 36 h (*bottom*) (Reprinted with permission from [30]. Copyright 2010 American Vacuum Society)

layer. Energy-dispersive X-ray spectroscopy (EDS) elemental analysis was used to analyze the Pt diffusion depth of the gate region and showed that around 5–10 nm of Pt diffused into the InAlAs layer.

In order to address the needs for elimination of burn-in process and better reliability, we studied the mechanisms of InAlAs/InGaAs MHEMT degradation. We investigated the degradation mechanism of MHEMTs under DC stress at drain voltage of 2.7 V and current density of 167 mA/mm and thermal storage. After DC stress, the Pt element of mushroom gate diffused into the InAlAs layer, which was confirmed with TEM images and EDS analysis. The high-voltage bias led to drain current degradation and gate sinking. The gate current increasing level of thermal stress (oven storage at 250°C for 36 hrs) was much larger than that of DC stress. The increases of the TLM resistance during DC stress and thermal storage resulted from different mechanisms. The Rc deterioration was dominant in the thermal storage test, and the Rs increase was observed in the DC stress. Metal spike formation and ohmic metal diffusion are observed during the thermal storage [30].

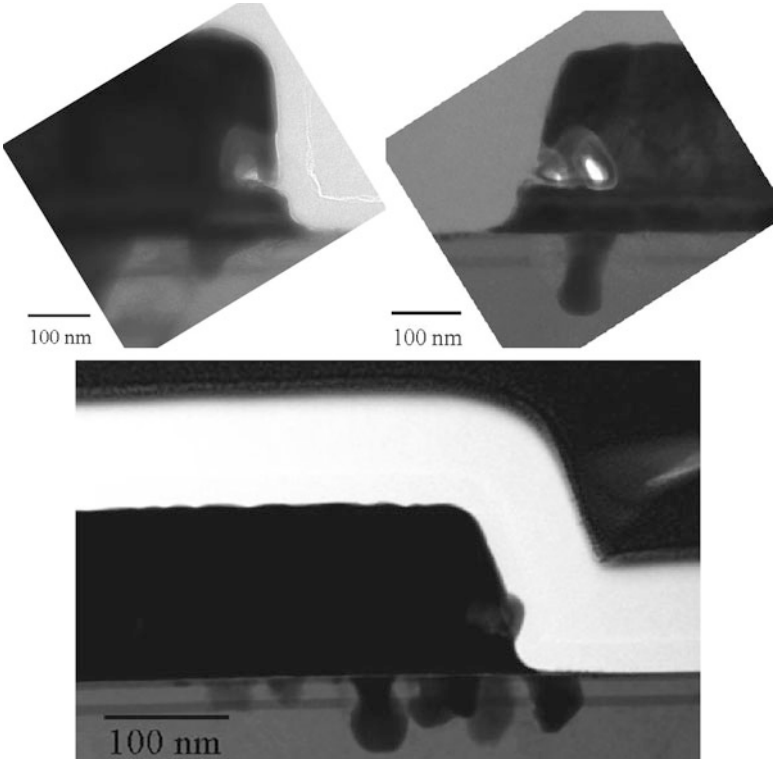



**Fig. 14.7** Higher-magnification TEM images of ohmic contact at the edge of source and drain contact (*top*, *left*, and *right*). Higher-magnification TEM images of ohmic contact at the edge of the contact (*bottom*)

## 14.4 GaAs HBTs

The path to commercialization of GaAs-based HBTs saw a number of alterations to the choice of base dopants and emitter layer material [31]. For high-current operation, it is necessary to use C as the base dopant instead of Be or Zn because of its higher doping capability (almost $10^{21}$ $cm^{-3}$) and superior stability. At high doping levels, Be is not soluble, and migration of charged Be interstitials leads to shifts in threshold voltage during device operation. In addition, AlGaAs can be replaced by InGaP, which provides lower surface recombination velocity, reduced susceptibility to oxidation, larger $\Delta E_V$, less recombination in the emitter, and improved emitter–base ledge passivation [32–39].

Some reported mechanisms for HBT device degradation under temperature or current density acceleration include midgap trap formation, hydrogen de-passivation from dopants, dislocation propagation, contact degradation, spiking, and base dopant diffusion. It was reported that the time to failure could be expressed as [2, 40]

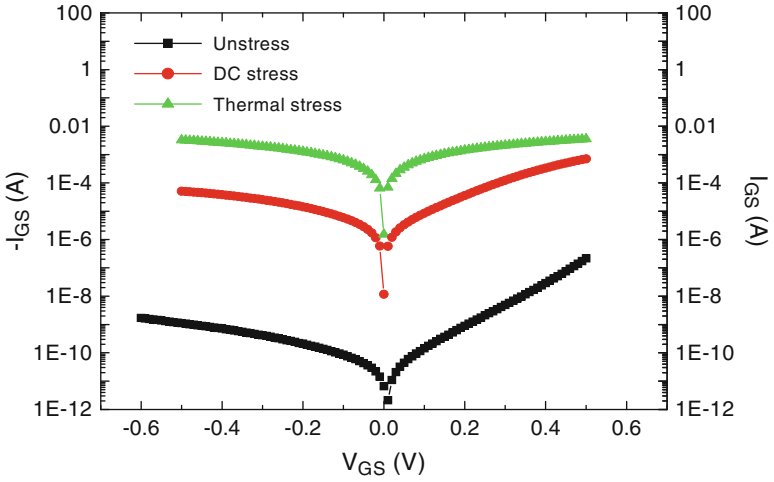



**Fig. 14.8** Gate I-V characteristics from MHEMTs before and after thermal or electrical stress

$$\mathrm{TTF} = CJ^{-\alpha}e(-\mathrm{Ea}/k\mathrm{Tj}) \tag{14.3}$$

where $C$ is const., $J$ is current density, $\alpha$ is current exponent, Ea is activation energy, and Tj is the junction temperature. Ea typically exhibits values around ~1 eV [2, 37, 40]. As we discussed earlier, the initial use of high Be base doping levels led to instability problems. At $p > 4 \times 10^{19}$ cm$^{-3}$, recombination-enhanced diffusion of $Be^+$ interstitials from base into emitter can occur, leading to a positive shift in $V_{BE}$ and decrease in gain. Switching to C as the dopant, significantly improved reliability, and extremely high base doping levels could be achieved, with $p = 3{-}10 \times 10^{21}$ cm$^{-3}$. However, even in HBTs with C-doped base layers, there are still issues with metallization stability, surface passivation, and avalanche breakdown.

One important issue arose with the use of implant isolation for maintaining device planarity. A schematic of a typical device is shown in Fig. 14.9. To achieve very high frequency operation, the total layer thicknesses were beyond what could be isolated with conventional energy implantation of oxygen ions, and thus a combined implant process that involved oxygen ions and protons to reach the region near the substrate interface was developed. We have correlated this gain degradation with the presence of hydrogen in the base layer, where it forms electrically inactive C–H complexes. During device operation, there is a gradual reactivation of the C acceptors due to minority-carrier-enhanced debonding of the hydrogen, and this produces a time-dependent decrease in current gain. The hydrogen in the base layer may be introduced either during the epitaxial growth from the trimethylgallium or arsine source gases, or during subsequent device fabrication. In the latter case, there are two particular opportunities for hydrogen introduction. The first is the implant isolation definition of the active region, normally achieved

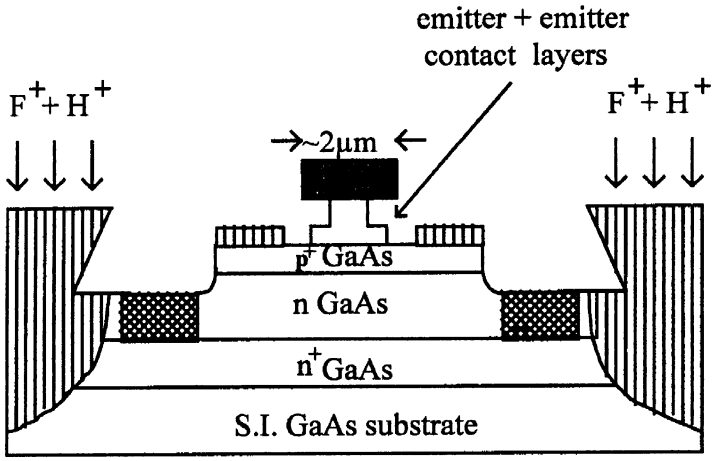



**Fig. 14.9** Schematic of implant isolated AlGaAs/GaAs HBT structure

by a multiple energy $F$ and H implant scheme, followed by annealing to maximize the resistance of the isolation region. We have also found that a second opportunity is the plasma-enhanced chemical vapor deposition of SiNx onto the exposed base and emitter mesa sidewall, which is followed by an etch back to leave the nitride sidewall spacer in place.

The typical signature of the effect of hydrogen is a decrease in HBT gain within a few minutes of biasing the device. The forward bias injection of electrons into the base leads to reactivation of carbon acceptors through the reaction

$$(\mathrm{C-H})^0 + \mathrm{e}^- \rightarrow \mathrm{C}^- + \mathrm{H}^+ + \mathrm{e}^- \tag{14.4}$$

After this injection-enhanced reactivation of base dopants, the released atomic hydrogen may form molecules (which are electrically inactive), but the effective hole concentration, $p$, increases, so gain decreases. The rate of this process is dependent on injected current density and time.

The rapid initial gain degradation during stress can be linked to a shift of the collector current characteristic to higher $V_{\mathrm{BE}}$. By comparing the evolution of the collector current shift, $\Delta V_{\mathrm{C0}}$, to that of the base current $I_{\mathrm{B}}$ during stress, it is apparent that they share the same dependence in time and on $T_{\mathrm{chuck}}$ (Fig. 14.10) [41]. Here, $\Delta V_{\mathrm{C0}}$ was obtained by periodically returning the devices to room temperature and measuring the Gummel characteristics. The shift $\Delta V_{\mathrm{C0}}$ could be caused by a decreased effective junction area of over 35% or by an increase in the base–emitter barrier height. A measurement of the base–emitter capacitance, $C_{\mathrm{BE}}$, before and after stress, showed an increase, rather than a decrease, in value of about 5%. Therefore, it is necessary to consider a change of the base–emitter barrier height toward reduction of the injected current.

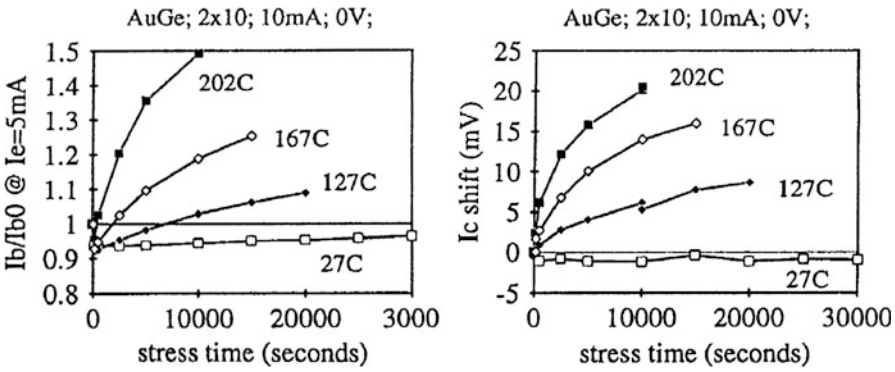



**Fig. 14.10** Comparison of the increase in base current (at fixed Ie) and the shift of the collector versus Vbe characteristic under various stress conditions

A gradual decrease in the base sheet resistance during stress was observed, which also occurs on the same timescale as the gain degradation [41]. The sheet resistance could be measured using a modified transmission line measurement, where test devices with different emitter width were used. In this special split-base geometry, the base metallizations of each side of the emitter were left disconnected, and the resistance between them is

$$R_{\text{split}} = W_E L_E{}^{-1}\text{Rb, square} + 2R_{\text{con}} \tag{14.5}$$

where $R_{\text{con}}$ is the base contact resistance of each half device. The base sheet resistance Rb,square is obtained from the slope of $R_{\text{split}}$ versus $W_E$ for a set of devices. Here, $R_{\text{split}}$ was measured several times during stress for a set of adjacent devices at a constant current density. A relatively low Je of 25 kA/cm$^2$ was used to keep the total current at values tolerable by the metallization for the wider devices, and a chuck temperature of 167°C was used to accelerate the degradation. Figure 14.11 shows the gradual decrease of the $R_{\text{split}}$ versus $W_E$ slope during stress. Van der Pauw structures with dimension of 50 × 50 μm$^2$ and processed alongside with the HBT devices were used to check the base layer sheet resistance independently. The device base sheet resistance has an initial value of 1,500Ω per square and decreased toward the van der Pauw value of 250Ω per square as electrical stress is applied (Fig. 14.11) [41]. All of this data is consistent with diffusion of hydrogen from the implant isolation region into the base during the anneal to maximize the resistance of this region. The hydrogen passivates some of the carbon acceptors and increases the resulting device gain. As the device is biased and electrons are injected into the base, the carbon acceptors are reactivated and the gain decreases.

An effective solution to this problem is to replace the proton implantation with $He^+$ ions. The modified implant isolation schedule is shown in Fig. 14.12, along with the resulting sheet resistance of the implanted region as a function of

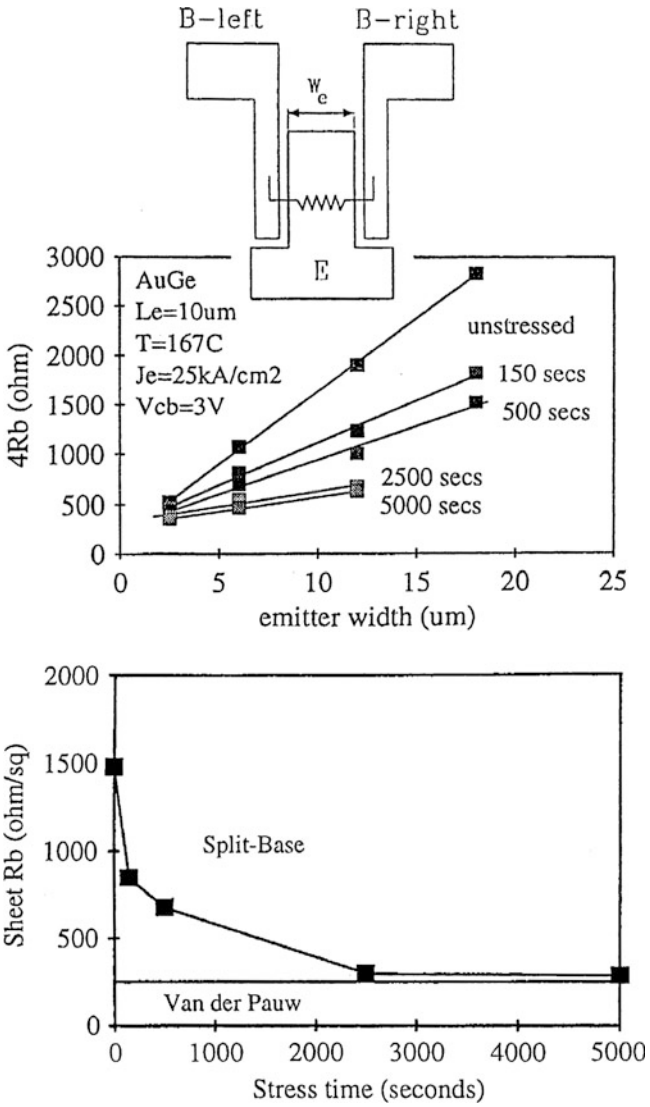



**Fig. 14.11** Resistance between base terminals of split-base devices versus emitter width, at various intervals during stress. The base sheet resistance is obtained from the slope and decreases during stress (*top*) and base sheet resistance versus duration of stress, decreasing to the value observed in van der Pauw test structures (*bottom*)

post-implant annealing temperature. Devices fabricated with this scheme did not exhibit the gain degradation [42]. This shows that the main source of the hydrogen was in fact the implant isolation region. When the hydrogen was replaced by He, there was no passivation of the carbon acceptors in the base, and therefore no time-dependent gain exhibited by the HBTs [43–45].

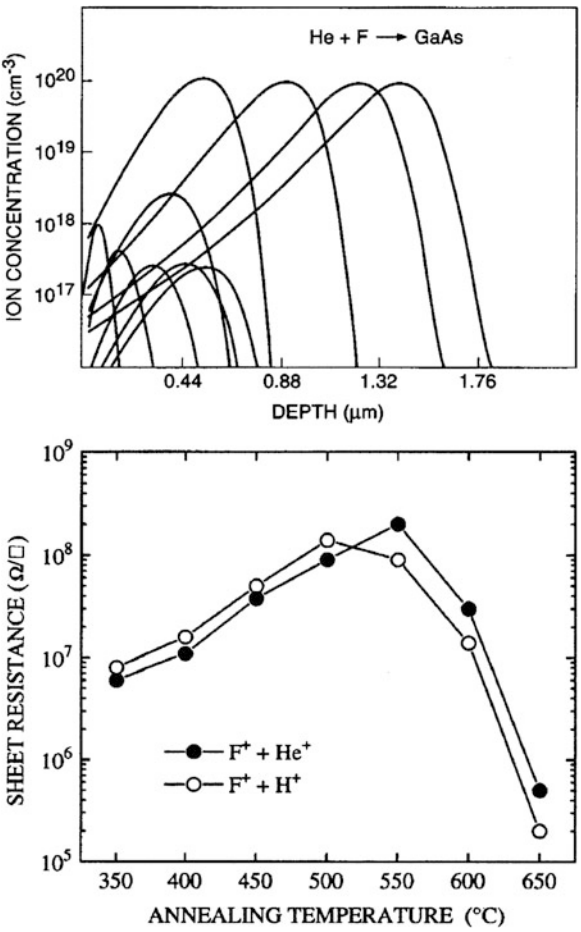



**Fig. 14.12** Implant isolation scheme for isolating AlGaAs/GaAs HBTs with total layer thickness of over 1.5 μm. Conventional energy $F^+$ ions cannot reach this distance, and lighter ions at a heavier dose are used to create enough traps to make the GaAs semi-insulating. The resulting resistivity of the implanted HBT structure as a function of annealing temperature is shown at the bottom of the figure

## 14.5 Summary

The main degradation mechanisms in AlGaAs/GaAs HEMTs and HBTs have been reviewed. Elevated temperatures, high electric fields, and high-current densities are all accelerators of degradation. The reaction of the gate metal with the underlying semiconductor, ohmic contact degradation, and creation of surface traps are all issues for HEMTs, while HBTs suffer mainly from contact reaction and junction leakage mechanisms. Careful design of the device geometry is important to avoid high-field regions and large current densities through contacts.

**Acknowledgment** This work was supported by an AFOSR MURI, monitored by Gregg Jessen.

## Appendix– Glossary of Terms in Reliability Studies

| | |
|---|---|
| **THB** | Temperature-humidity-bias. A standard set of conditions defined under Electronics Industry Alliance/JES D22-A101-B. 85°C/85% rel. humidity |
| **HAST** | Highly accelerated stress test. A standard set of conditions defined under EIA JES D22-A110B. 110/130°C/85% rel. humidity |
| **Burn-in test** | An initial stress at high current or temperature often found to weed out defective parts and stabilize devices for further test |
| **TLM** | Transmission line pattern used for extracting contact resistance – provides larger area for material analysis. Lateral fields present, vertical fields absent, may be used to identify degradation mechanism in limited cases |
| **SEC** | Standard evaluation circuitry |
| **MTTF** | Median or mean-time-to-failure (time required for 50% population to fail) |
| **ESD/EOS** | Electrostatic discharge/electrical over-stress. ESD occurs when there is a difference in electrostatic potential between two charged objects; an exchange of electrostatic energy occurs causing their potentials to become balanced. EOS is the term for stress that exceeds the rated V/I and occurs when this exceeds the values guaranteed by the IC |
| **HBM** | Human body model. Modeling of a situation where energy from an external electrical power or electrostatic energy source becomes charged within the human body to become discharged as electrostatic energy |
| **MM** | Machine model. Modeling of a situation where electrostatic energy is charged within the machine through a grounding problem or other negative influence |
| **DIR** | Design-in reliability |
| **PLR** | Package level reliability |
| **SPC** | Statistical process control using statistical techniques to measure and analyze the variation in processes. Most often used for manufacturing processes, the intent of SPC is to monitor product quality and maintain processes to fixed targets |
| **EFR** | Early failure rate |
| **HTOL** | High temperature operating life |
| **HTRB** | High temperature reverse bias |
| **HFGC** | High forward gate current |
| **Corrosion** | Metal degradation due to chemical or electrolytic reactions in presence of moisture, contaminants, or bias |

**Electromigration** Movement of metal atoms of a metal line in the direction of the current flow through that line. Metal atoms are removed from the starting end of the metal line and accumulate at the other end, forming voids at the entrance and hillocks at the exit of the line. Electromigration can result in open circuits (due to the voids) or line-to-line short circuits (due to the hillocks). Accelerated by temperature and current density and is modeled with $t_f = CJ\text{-ne}^{(\text{Ea/kT})}$

**Kink** Abrupt increase in drain current of HEMTs at a certain gate voltage, usually followed by a rapid increase in output conductance. A low-frequency phenomenon related to trapping effects

**Gate sinking** Reaction of HEMT gate metal with semiconductor at elevated temperatures

**COTS** Commercial-off-the-shelf

**TDDB** Time-dependent dielectric breakdown charge injection mechanism, leading to destruction of dielectric layers over time. During the buildup stage, charges invariably get trapped in various parts of the oxide as current flows in the oxide. The trapped charges increase in number with time, forming high electric fields (electric field = voltage/oxide thickness) and high-current regions along the way. This process of electric field buildup continues until the runaway stage is reached. During the runaway stage, the sum of the electric field built up by charge injection and the electric fields applied to the device exceeds the dielectric breakdown threshold in some of the weakest points of the dielectric

**Gate lag** A delayed response of the channel current to modulation of the gate potential. Results from traps on semiconductor surface near the gate

**Hot carrier effects** Phenomenon involving the injection of highly energetic carriers into the gate oxide layer and the substrate, resulting in volume charge buildup that can shift transistor threshold voltages. This mechanism is accelerated by low temperatures. Field accelerates electrons in channel (often observed in HEMTs). When Ee > Eg, impact ionization occurs, creating e–h pairs. Holes flow to the device source, overcome ΔEv, collected by gate. This leads to a negative gate component which may dominate at high VDS. The shape of the IG-VGS curves may identify impact ionization. Manifested by VT shift, VB ↑, gm, and ft ↓, power slump

**BIR** Built-in reliability. A phrase used in the Si industry (1900–1995 timeframe) where there was an emphasis on process control, in-line screening, and wafer level reliability (WLR) testing

## References


1. W.J. Roesch, Microelectron. Reliab. **46**, 1218 (2006)
2. T.A. Henderson, in *Reliability of Compound Semiconductors Workshop*, 2006 (unpublished)
3. J.A. del Alamo, A.A. Villanueva, IEDM Tech. Dig. 1019–1022 (2004)
4. G. Meneghesso, E. Zanoni, Microelectron. Reliab. **42**, 685 (2002)
5. Y.K. Kukai, S. Sugitani, T. Enoki, Y. Yamane, IEEE Trans. Device Mater. Reliab. **8**, 289 (2008)
6. G. Meneghesso, C. Canali, P. Cova, E. De Bortoli, E. Zanoni, IEEE Electron Device Lett. **17**, 232 (1996)
7. Y.C. Chou, R. Grundbacher, D. Leung, R. Lai, P.H. Liu, Q. Kan, M. Beidenbender, M. Wojtowicz, D. Eng, A. Oki, IEEE Electron Device Lett. **25**, 64 (2004)
8. W. Roesch, in *GaAs REL Workshop*, Monterey, 1999, pp. 119–126
9. A.A. Villanueva, J.A. del Alamo, F. Hisaka, K. Hayashi, M. Somerville, IEEE Trans. Device Mater. Reliab. **8**, 283 (2008)
10. M. Borgarino, R. Menozzi, Y. Baeyens, P. Cova, F. Fantini, IEEE Trans. Electron Devices **45**, 366 (1998)
11. R.E. Leoni, J.C.M. Hwang, IEEE Trans. Electron Devices **46**, 1608 (1999)
12. K.J. Choi, J.L. Lee, J. Electron. Mater. **30**, 885 (2001)
13. T. Hisaka, Y. Nogami, H. Sasaki, A.A. Villanueva, A. Hasuike, N. Yoshida, J.A. del Alamo, in *Proceedings of the Gallium Arsenide Integrated Circuits Symposium*, Philadelphia, 2003, pp. 67–70
14. A. Villanueva, J.A. del Alamo, T. Hisaka, T. Ishida, IEDM Tech. Dig. 393–396 (2007)
15. S. Mertens, J. del Alamo, IEEE Trans. Electron Devices **49**, 1849 (2002)
16. P. Ersland, H. Jen, X. Yang, in *GaAs Rel. Workshop*, San Diego, 2003, pp. 3–6
17. T. Hisaka, in *Reliability on Compound Semiconductors (ROCS) Workshop*, Monterey, 2004, pp. 81–88.
18. I. Thayne, K. Elgaid, D. Moran, X. Cao, E. Boyd, H. McLelland, M. Holland, S. Thoms, C. Stanley, Thin Solid Films **515**, 4373 (2007)
19. S. Bollaert, Y. Cordier, M. Zaknoune, H. Happy, V. Hoel, S. Lepilliet, D. Théron, A. Cappy, Solid-State Electron. **44**, 1021 (2000)
20. M. Dammann, M. Chertouk, W. Jantz, K. Köhler, G. Weimann, Microelectron. Reliab. **40**, 1709 (2000)
21. P.F. Marsh, C.S. Whelan, W.E. Hoke, R.E. Leoni, T.E. Kazior, Microelectron. Reliab. **42**, 997 (2002)
22. M. Dammann, A. Leuther, R. Quay, M. Meng, H. Konstanzer, W. Jantz, M. Mikulla, Microelectron. Reliab. **44**, 939 (2004)
23. M. Dammann, F. Benkhelifa, M. Meng, W. Jantz, Microelectron. Reliab. **42**, 1569 (2002)
24. T. Suemitsu, Thin Solid Films **515**, 4378 (2007)
25. C.S. Whelan, P.F. Marsh, W.E. Hoke, R.A. McTaggart, P.S. Lyman, P.J. Lemonias, S.M. Lardizabal, R.E. Leoni, S.J. Lichwala, T.E. Kazior, IEEE J. Solid State Circuits **35**, 1307 (2000)
26. C. Karnfelt, R. Kozhuharov, H. Zirath, I. Angelov, IEEE Trans. Microw. Theory Tech. **54**, 2887 (2006)
27. Y. Campos-Roca, C. Schworer, A. Leuther, M. Seelmann-Eggebert, H. Massler, IEEE Microw.Wirel. Compon. Lett. **15**, 466 (2005)
28. M. Dammann, A. Leuther, A. Tessmann, H. Massler, M. Mikulla, G. Weimann, Electron. Lett. **41**, 699 (2005)
29. A. Leuther, A. Tessmann, I. Kallfass, R. Losch, M. Seelmann-Eggebert, N. Wadefalk, F. Schafer, J.D. Gallego Puyol, M. Schlechtweg, M. Mikulla, O. Ambacher, in *2009 I.E. Conference on Indium Phosphide & Related Materials*, Newport Beach, 2009, pp. 188–191
30. K.H. Chen, C.Y. Chang, L.C. Leu, C.F. Lo, B.H. Chu, S.J. Pearton, F. Ren, J. Vac. Sci. Technol. B **28**, 365 (2010)

31. A.A. Immorlica, Reliability of III-V electronic devices, in *III-V Compound Semiconductors*, ed. by T. Li, M. Mastro, A. Dadgar (CRC Press, Boca Raton, 2011)
32. M.G. Adlerstein, J.M. Gering, IEEE Trans. Electron. Device. **47**, 434 (2000)
33. X. Liu, J.S. Yuan, J.J. Liou, Microelectron. Reliab. **48**, 1212 (2008)
34. B.P. Yan, Y.F. Yang, C.C. Hsu, H.B. Lo, E.S. Yang, Microelectron. Reliab. **41**, 1959 (2001)
35. O. Ueda, Microelectron. Reliab. **39**, 1839 (1999)
36. D. Pavlidis, Microelectron. Reliab. **39**, 1801 (1999)
37. T. Henderson, Microelectron. Reliab. **39**, 1033 (1999)
38. J. William, Roesch Microelectron. Rel. **51**, 188 (2011)
39. S. Ghosh, F. Marc, C. Maneux, B. Grandchamp, G.A. Koné, T. Zimmer, Microelectron. Reliab. **50**, 1587 (2010)
40. Y. Chen, Q. Wang, S. Kayali, Microelectron. Reliab. **45**, 1869 (2005)
41. M.R. Frei, T.-Y. Chiu, C.R. Abernathy, F. Ren, T.R. Fullowan, J. Lothian, S.J. Pearton, B. Tseng, R.K. Montgomery, P.R. Smith, Solid-State Electron. **46**, 1301 (2002)
42. S.J. Pearton, C.R. Abernathy, J.W. Lee, F. Ren, C.S. Wu, J. Vac. Sci. Technol. B **13**, 15 (1995)
43. F. Ren, S.J. Pearton, W.S. Hobson, T.R. Fullowan, J. Lothian, A.W. Yanof, Appl. Phys. Lett. **56**, 860 (1990)
44. S.J. Pearton, F. Ren, J. Lothian, T. Fullowan, A. Katz, P. Wisk, C. Abernathy, R. Kopf, R. Elliman, M. Ridgeway, C. Jagadish, J. Williams, J. Appl. Phys. **71**, 4949 (1992)
45. S.J. Pearton, F. Ren, S. Chu, C.R. Abernathy, W.S. Hobson, R. Elliman, J. Appl. Phys. **74**, 6580 (1993)

# Chapter 15
# Novel Dielectrics for GaN Device Passivation and Improved Reliability

**F. Ren, Stephen J. Pearton, B.P. Gila, C.R. Abernathy, and R.C. Fitch**

**Abstract** Proper surface cleaning processes and the type of passivation film ($SiN_X$, $Sc_2O_3$, MgO) used to reduce the current collapse phenomena in the devices are very critical to reduce the inter-device isolation leakage currents in mesa-isolated AlGaN/GaN high electron mobility transistors and to improve reliability. Three different passivation layers ($SiN_X$, MgO, and $Sc_2O_3$) were examined for their effectiveness in mitigating surface state-induced current collapse in AlGaN/GaN high electron mobility transistors (HEMTs). The plasma-enhanced chemical vapor deposited $SiN_X$ produced ~80–85% recovery of the drain-source current, independent of whether $SiH_4/NH_3$ or $SiD_4/ND_3$ plasma chemistries were employed. Both the $Sc_2O_3$ and MgO produced essentially complete recovery of the current in GaN-cap HEMT structures and ~80–95% recovery in AlGaN-cap structures. The $Sc_2O_3$ had superior long-term stability with no change in HEMT behavior over 5 months aging. The use of MOSFETs could allow the use of complementary devices, thus producing less power consumption and simpler circuit design. The same novel oxides employed for alleviating many of the problems encountered in current Schottky-based devices were successfully used as the gate dielectric for MOS-diode and MOSFETs. Both MgO and $Sc_2O_3$ were shown to provide low interface state densities (in the $10^{11}$ $eV^{-1}$ $cm^{-2}$ range) on n- and p-GaN, making them useful for gate dielectrics for metal-oxide-semiconductor (MOS) devices and

F. Ren
Department of Chemical Engineering, University of Florida,
Gainesville, FL 32611, USA

S.J. Pearton (✉) • B.P. Gila • C.R. Abernathy
Department of Materials Science and Engineering, University of Florida, Gainesville, FL 32611, USA
e-mail: spear@mse.ufl.edu

R.C. Fitch
Sensors Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base,
OH 45433-7322, USA

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_15,
© Springer Science+Business Media New York 2013

also as surface passivation layers to mitigate current collapse in GaN/AlGaN high electron mobility transistors (HEMTs). Clear evidence of inversion was demonstrated in gate-controlled MOS p-GaN diodes using both types of oxide. Charge-pumping measurements on diodes undergoing a high temperature implant activation anneal showed a total surface state density of ~$3 \times 10^{12}$ $cm^{-2}$. On HEMT structures, both oxides provided effective passivation of surface states and these devices showed improved output power. The MgO/GaN structures were also found to be quite radiation resistant, making them attractive for satellite and terrestrial communication systems requiring a high tolerance to high energy (40 MeV) protons.

## 15.1 Introduction

AlGaN/GaN high electron mobility transistors (HEMTs) appear well suited to high speed and high temperature applications including high frequency wireless base stations and broadband links, commercial and military radar, and satellite communications [1–34]. These devices are capable of producing very high power densities (>12 W/mm), along with high breakdown voltage and low noise figures. The highest performance devices are typically grown on SiC substrates because of their high thermal conductivity. These devices can achieve excellent power performance well above that obtained with competing technologies such as GaAs and SiC metal-semiconductor field-effect transistors (MESFETs). However, a commonly observed problem in these devices is the so-called current collapse in which the application of a high drain-source voltage leads to a decrease of the drain current and increase in the knee voltage[4, 6, 12–17, 20]. This phenomenon can also be observed by a current dispersion between dc and pulsed test conditions or a degraded RF output power. The mechanisms include the presence of surface states on the cap layer or trapping centers in the resistive buffer underlying the active channel. The carriers in the 2-dimensional electron gas can be lost either to the surface or buffer traps [4, 6, 35]. The buffer traps are a function of the epitaxial growth condition. The effect of the surface traps may be mitigated by the use of appropriate passivation layers, most often $SiN_X$ deposited by plasma-enhanced chemical vapor deposition (PECVD) [18, 21, 36]. The use of $SiN_X$ passivation typically restores 70–80% of the dc current [35]. Other dielectric passivation layers such as MgO or $Sc_2O_3$ deposited by plasma-assisted molecular beam epitaxy (MBE) can provide even higher levels of current recovery under optimum conditions [19, 35, 36]. Besides depositing a proper surface passivation layer, surface cleaning processes prior to the deposition of the passivation dielectric appear to be very critical to rescue surface traps.

The use of metal-oxide-semiconductor (MOS) or metal-insulator-semiconductor (MIS) gates for HEMTs produces a number of advantages over the more conventional Schottky metal gates, including lower leakage current and greater voltage swing [2, 22, 26, 28–34, 37–43]. The dielectric materials used for HEMT passivation have also been employed for gate dielectric for the MOSFET and MOSHEMTs including $SiO_2$ [3, 22–27, 38], Gd ($Ga_2O_3$) [22, 23, 27, 34], AlN [37], $SiN_X$ [14, 32, 33], MgO [24, 39], and $Sc_2O_3$ [39].

In this chapter, we will describe progress on surface cleaning prior to the deposition of the passivation layer and different dielectrics for HEMT passivation and gate dielectrics [39, 41–43]. These are key requirements for reliable device operation. Table 15.1 shows a comparison of the properties of these oxides with those of the other commonly used passivation materials on GaN. These novel oxides do not contain hydrogen and may have advantages over $SiN_X$ in that respect because atomic hydrogen diffuses rapidly and could enter the GaN or gate metal over extended periods of device operation.

## 15.2 Effects of Surface Treatments Prior to Dielectric Passivation

The outstanding microwave power performance demonstrated by AlGaN/GaN HEMTs makes them excellent candidates for phased array radar, wireless base stations, or satellite communications systems. Figure 15.1 shows a comparison of typical device performance from different technologies. However, it is generally observed that the RF power from GaN transistors is considerably lower than expected from the dc characteristics. This RF power decrease is manifested by a drain current collapse caused by the presence of surface states between the gate and drain which depletes the channel. This phenomenon can also be observed by a current dispersion between dc and pulsed test conditions or a degraded RF output power. The effect of the surface traps may be mitigated by the use of appropriate passivation layers, most often $SiN_X$ deposited by plasma-enhanced chemical vapor deposition (PECVD). However, the magnitude of the isolation current is a strong function of the surface treatment employed.

The effects of surface treatments included ashing in an $O_2$ barrel etcher, UV-ozone exposure for 25 min at 25°C or dipping for 10 s in $NH_4OH:H_2O$ (1:1), followed by either PECVD of $SiN_X$ or MBE deposition of MgO or $Sc_2O_3$ surface passivation films on the inter-device isolation currents were investigated. The unpassivated devices show nA isolation current, as illustrated in Fig. 15.2.

Note that $O_2$ plasma ashing, ozone treatment, or $NH_4OH:H_2O$ dips by themselves do not significantly alter the isolation current, but $SiN_X$ deposition with any surface cleaning step always increases the current by at least an order of magnitude. The activation of this current remains at ~1.5 eV, indicating that it is the density rather than the nature of the surface states that is altered by the nitride deposition. The increased trap density could originate from energetic ion bombardment or preferential loss of $N_2$ from the GaN surface during the plasma deposition. While the ashing, ozone exposure, or chemical rinsing themselves do not increase isolation leakage, they do not prevent current collapse in the HEMTs. It is always necessary to deposit a passivation layer to mitigate the current collapse. Figure 15.2 also shows the isolation current at 40 V bias between neighboring HEMTs as a function of $SiN_X$ pre-deposition treatment. Over a limited temperature range $T$ (25–225°C), this current (I) is thermally activated with the form

**Table 15.1** Material properties for SiC, GaN, and various dielectrics

| | SiC | GaN | $SiO_2$ | AlN | $Gd_2O_3$ | $Sc_2O_3$ | CaO | MgO | $Mg_{.5}Ca_{.5}O$ | $Mg_{.75}Ca_{.25}O$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Structure | W (4 H) | W | A | W or A | B | B | N | N | N | N |
| Lattice constant | 3.07 | 3.186 | – | 3.113 | 10.813 | 9.845 | 4.799 | 4.2112 | 4.505 | 4.356 |
| Atomic spacing in the (111) plane | | | – | – | 3.828 | 3.4807 | 3.393 | 2.978 | 3.186 | 3.08 |
| Mismatch to GaN (%) | – | – | – | 2.3 | 20.1 | 9.2 | 6.6 | −6.5 | 0 | – |
| Mismatch to SiC (%) | | | | 1.1 | 24.3 | 13 | 10.2 | −3.3 | | 0.3 |
| $T_{MP}$ (K) | | 2,800 | 1,900 | 3,500 | 2,668 | 2,678 | 2,860 | 3,073 | | |
| Bandgap (eV) | 3.26 | 3.4 | 9 | 6.2 | 5.3 | 6.3 | 7 | 8 | | |
| Electron affinity (eV) | | 3.4 | 0.9 | 0–2.9 | 0.63 | | | 0.7 | | |
| Work function (eV) | 4.5 | | | 0.9–1.2 | 2.1–3.3 | 4 | | 3.1–4.4 | | |
| Dielectric constant | 9.7–10.3 | 9.5 | 3.9 | 8.5 | 11.4 | 14 | 11.8 | 9.8 | | |

Data comes from references [44–56]
*W* wurtzite, *A* amorphous, *B* bixbyite, *N* NaCl

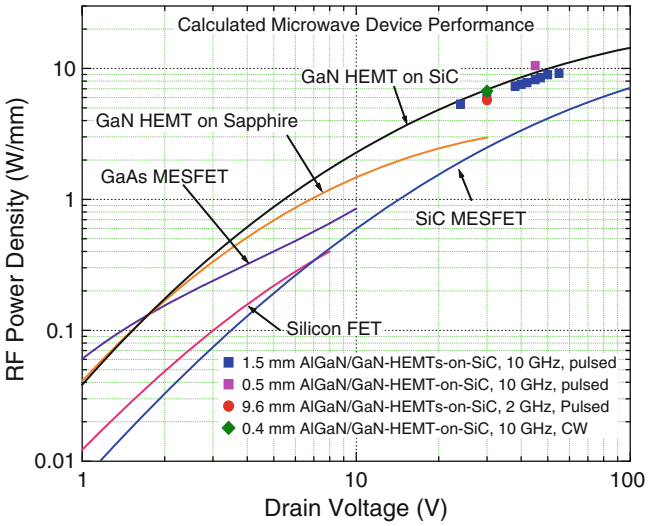



**Fig. 15.1** Power output as a function of drain bias for transistors made from different materials

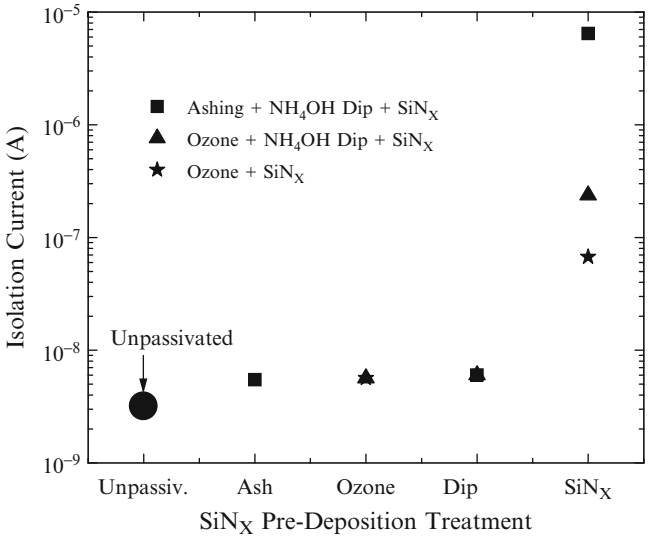



**Fig. 15.2** Isolation current measured at 40 V for mesa-isolated HEMTs, as a function of different $SiN_X$ pre-deposition treatments (Reprinted with permission from [57]. Copyright 2003 American Institute of Physics)

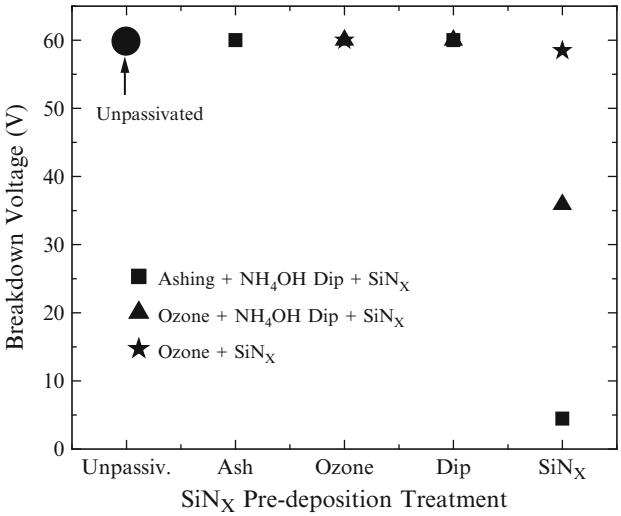



**Fig. 15.3** Breakdown voltage between neighboring devices for mesa-isolated HEMTs, as a function of different $SiN_X$ pre-deposition treatments (Reprinted with permission from [57]. Copyright 2003 American Institute of Physics)

$$I = I_0 \exp\left(\frac{-E_a}{KT}\right) \tag{15.1}$$

where $I_0$ is a constant, $K$ is Boltzmann's constant, and $E_a$ is ~1.5 eV. We assume this activation energy represents an approximate ionization energy of the bulk or surface states controlling the inter-device leakage current. This value is in fair agreement with the values for surface-related deep traps measured on AlGaN/GaN HEMTs by deep level transient spectroscopy, namely ~1.4 eV [43, 58].

The corresponding reverse breakdown voltages ($V_B$) for the mesa-isolated HEMTs (defined as a reverse current density of $\geq$1 μA/cm$^2$) is shown in Fig. 15.3. There is a significant reduction in $V_B$ only in the case of $SiN_X$ deposition preceded by either ozone exposure or ashing compared with cleaning in $NH_4OH$:$H_2O$. A plausible explanation is that the $NH_4OH$:$H_2O$ treatment reduces the native oxide thickness to the point that the surface is more vulnerable to plasma damage during the $SiN_X$ plasma deposition step. In the other process sequences, the presence of a thicker native oxide may help shield the surface from ion bombardment or fluxes of hydrogen radicals.

There was also a strong time dependence to the measured isolation leakage current in some cases. As an example, Fig. 15.4 shows the I-V characteristics for the unpassivated case (before UV) where the slope from 10 to 40 V is $8.6 \times 10^{-11}$ A/V corresponding to a resistance of $1.7 \times 10^{10}$ Ω between the HEMTs. Directly after UV-ozone exposure, the leakage current was in the μA range with a corresponding resistance of ~$8 \times 10^7$ Ω. When remeasured after standing in room ambient for 3 days, the current was reduced to the nA range (Fig. 15.3) and the linear

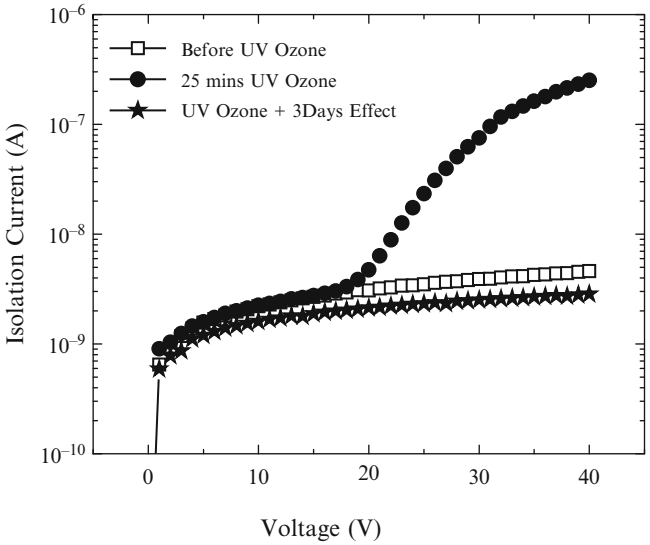



**Fig. 15.4** I-V characteristics for isolation current between neighboring devices both before and after UV-ozone treatment (Reprinted with permission from [57]. Copyright 2003 American Institute of Physics)

portion of the I-V characteristic yielded a slope of $\sim 5 \times 10^{-11}$ A/V or an equivalent resistance of $2 \times 10^{10}$ Ω. Note in this figure that only the case of $SiN_X$ deposition, regardless of pretreatment, was effective in reducing current collapse in the HEMTs.

Figure 15.5 shows the isolation current increase at 40 V as a function of the type of passivation employed. In each case, the surface was treated in UV ozone prior to the film deposition. The lowest isolation currents were obtained with $Sc_2O_3$, and these were roughly a factor of 6 lower than with the PECVD $SiN_X$. This is consistent with the lower surface state densities measured for the $Sc_2O_3$/GaN and MgO/GaN interfaces and reported previously [36, 59]. The isolation leakage currents in mesa-isolated AlGaN/GaN HEMTs are found to be dominated by surface states whose density is a strong function of the cleaning and passivation films employed. The isolation current is always higher after $SiN_X$ deposition, while MBE-deposited $Sc_2O_3$ produces the lowest values for any of the films that reduce the current collapse problem in the HEMTs.

## 15.3 Comparison of Dielectrics for AlGaN/GaN HEMT Surface Passivation to Reduce Current Collapse

Gate-lag measurements on the HEMTs have been widely used as a metric for establishing the effectiveness of the device passivation [35]. In this method, the drain current ($I_{DS}$) response to a pulsed gate-source voltage ($V_G$) is measured.

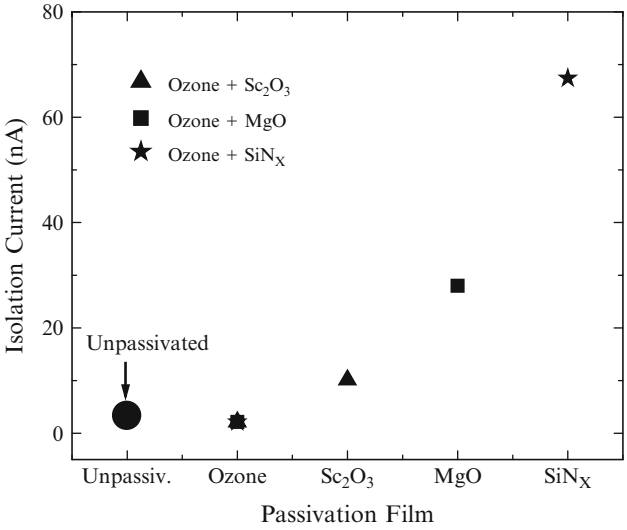



**Fig. 15.5** Isolation current measured at 40 V for mesa-isolated HEMTs, as a function of the film used for surface passivation (Reprinted with permission from [57]. Copyright 2003 American Institute of Physics)

Figure 15.6 shows the normalized $I_{DS}$ as a function of gate voltage ($V_G$) for both dc and gate pulsed measurements on unpassivated HEMTs. In the data, $V_G$ was pulsed from –5 to 0 V at 0.1 MHz and 10% duty cycle. The large differences between dc and pulsed drain currents of the unpassivated devices was due to the presence of surface traps that depleted the channel in the access regions between the gate and drain contacts.

Conventional plasma-enhanced chemical vapor deposited $SiN_X$ is widely used to passivate the AlGaN/GaN HEMTs, and the precursors for PECVD $SiN_X$ film were $SiH_4$ and $NH_3$. The deposition temperature of the PECVD system is kept $\leq$300°C, and the Si–H and N–H bonds of $SiH_4$ and $NH_3$ cannot be completely broken [59, 60]. Therefore, typical PECVD $SiN_x$ does not have a fixed stoichiometry but rather contains varying amount of Si, N, and H. The hydrogen content can be as high as 39% (atomic) and is strongly dependent on the deposition conditions. The $SiN_x$ film deposited without load-lock system often also contains around 5% oxygen when low deposition temperature is used (<250°C). The ratio of Si–H to N–H also depends on the deposition conditions [61]. The rearrangement of hydrogen bonds and the out-diffusion of free hydrogen in the film can create the traps and cause the instability in devices [62–65]. This trap formation can occur at subsequent high temperature process steps or during the device operation. Under the high electric field conditions, the hot electrons can inject into the dielectric film and will release the bonded H therefore creating electron traps.

SiNx film can also be deposited by using deuterated precursors: $SiD_4$ and $ND_3$. Si–D and N–D bond energies are larger than those of Si–H and N–H bonds, and the diffusion of deuterium in Si is slower than that of hydrogen. Deuterium has also

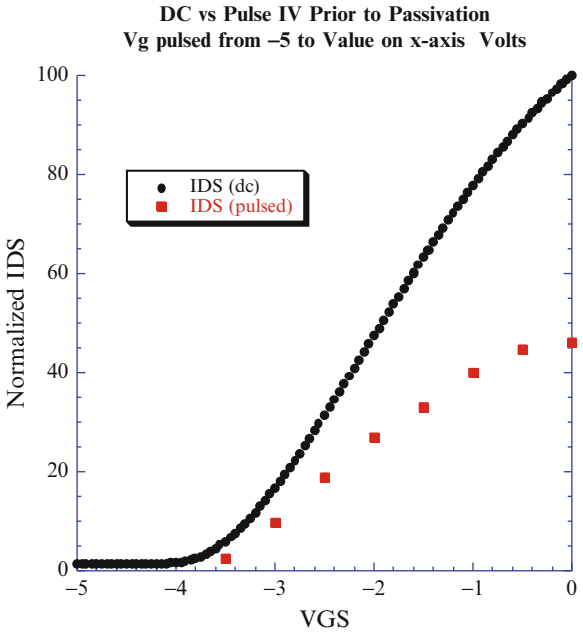



**Fig. 15.6** Gate-lag measurements on unpassivated AlGaN/GaN HEMTs. $V_G$ is switched from –5 to 0 V

been used in Si-MOSFETs to passivate the interface states, which shows better thermal stability as compared to hydrogen due to the higher bonding energy of the Si–D bond. Thus, SiNx deposited with deuterated precursors should have better stability as compared to the SiNx deposited with silane and ammonia.

$Gd_2O_3$, $Sc_2O_3$, and MgO have been used as gate oxide for the metal-oxide-semiconductor (MOS) technology, and oxide/GaN interface state densities were ~$2 \times 10^{11}$–$1 \times 10^{12}$ $cm^{-2}$ [40–43]. $Sc_2O_3$ and MgO have smaller lattice mismatch to GaN and larger bandgap than $Gd_2O_3$. These oxides do not contain hydrogen and may have advantages over $SiN_X$ in that respect because atomic hydrogen diffuses rapidly and could enter the GaN or gate metal over extended periods of device operation. Thus, the novel MgO as well as $Sc_2O_3$ layers could be employed to passivate AlGaN/GaN HEMTs.

Schematics of HEMT structures used in the studies of device passivation are shown in Fig. 15.7. Both $Al_2O_3$ and SiC substrates were used. The doping in this structure is basically due to piezo-induced carriers. The second type of HEMT employed an undoped $Al_{0.3}$ $Ga_{0.7}$ N-cap layer on top of a doped $Al_{0.3}$ $Ga_{0.7}$ N donor layer. The doping in this structure is due to the intentional Si-doping of the donor layer. The ohmic (Ti/Al/Pt/Au) and gate (Ni/Au) contacts were deposited by e-beam evaporation and patterned by lift-off. The 100 Å thick MgO, $Sc_2O_3$, or $SiN_X$ layers were deposited on completed devices (0.25–1.2 μm gate length, 100 μm gate width) using either molecular beam epitaxy (MBE) for the oxides or PECVD for the $SiN_X$, with deposition temperatures of 100°C in the oxide case and 250°C for the $SiN_X$ case. The $SiN_X$ films were deposited with either $SiH_4$ + $NH_3$ or $SiD_4$ + $ND_3$ to examine the effect of deuterated precursors.

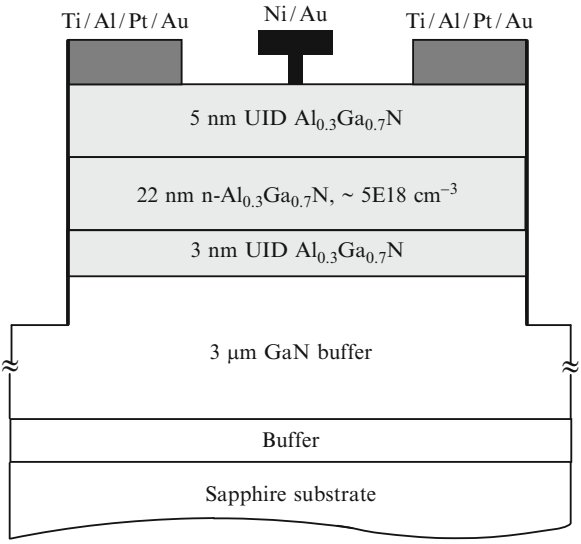



**Fig. 15.7** Schematic cross sections of AlGaN/GaN HEMT structures

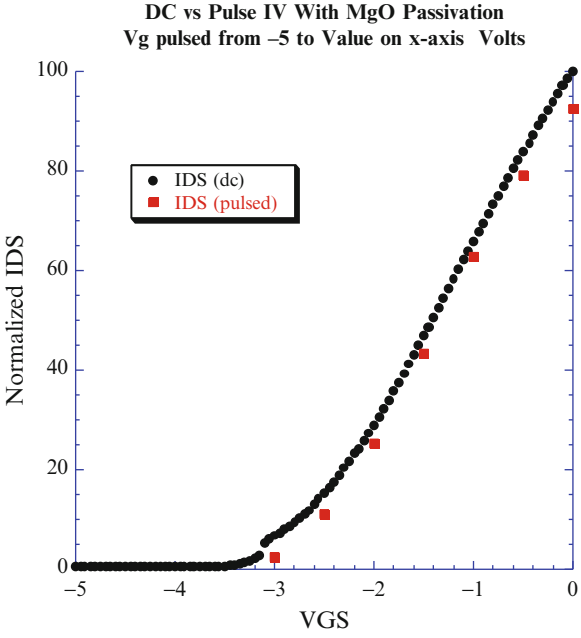



**Fig. 15.8** Gate-lag measurements on MgO-passivated AlGaN/GaN HEMTs. $V_G$ is switched from –5 to 0 V

Figure 15.8 shows the normalized $I_{DS}$ of MgO-passivated HEMTs as a function of gate-source voltage, $V_G$, switched from –5 to 0 V with $V_{DS}$ held constant at a low value (3 V) to avoid the complications of device heating. The MgO-passivated HEMTs showed an increase in drain-source current of 20% in the dc mode which is consistent with passivation of surface state. This is clear evidence for the assumption that surface

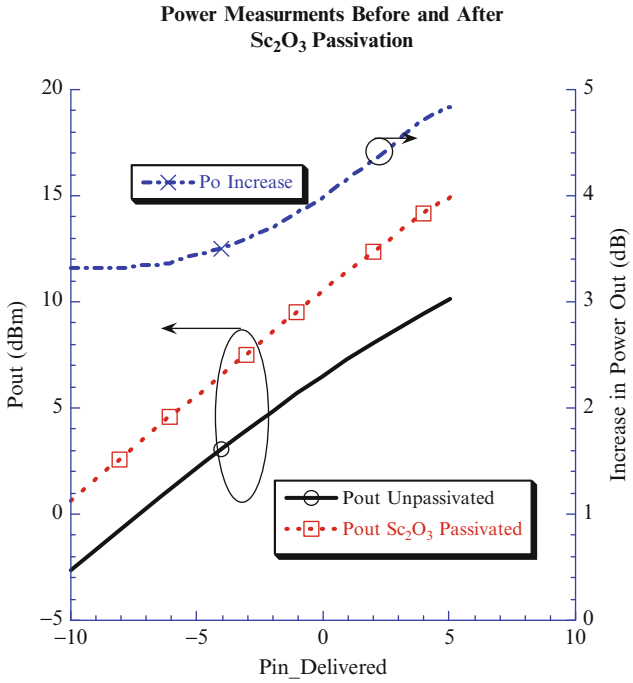



**Fig. 15.9** Output power characteristics before and after $Sc_2O_3$ passivation of 0.5 × 100 μm$^2$ devices measured at 4 GHz and a bias point of $V_D = 10$ V, $V_G = -2$ V

states are the cause of the gate-lag phenomena and also that MgO passivation mitigates this problem. Similar results were also obtained with $Sc_2O_3$ passivation.

Typical load-pull data for HEMTs before and after $Sc_2O_3$ passivation are shown in Fig. 15.9 for a measurement frequency of 4 GHz. In all cases, the drain voltage, $V_D$, was held at 10 V, while the gate voltage $V_G$ was –2 V. The wafer measurements employed mechanical tuners for matching, and there was no harmonic termination under class A operation. The devices were matched for the power testing prior to passivation and were tested under these same conditions after oxide deposition. Note the increased power output. By contrast, the difference in output power before and after $SiN_X$ passivation was <2 dB in all cases. This is consistent with the higher percentage recovery of $I_{DS}$ during the gate-lag measurements with $Sc_2O_3$ passivation. The $Sc_2O_3$ passivation has shown excellent aging characteristics when measured under dc test conditions, with no change in HEMT performance over a period of >5 months, as shown in Fig. 15.10 (top). Figure 15.10 (bottom) shows gate-lag measurement for MgO-passivated HEMTs immediately after MgO deposition and after 5 months aging without bias on the devices under room conditions (temperature and humidity). The $I_{DS}$ increases ~20% upon passivation, as mentioned earlier, and there is almost complete mitigation of the degradation in $I_{DS}$ immediately after MgO deposition. However, after 5 months aging, there is a clear difference between the dc and pulsed data, indicating that the MgO passivation has lost some of its effectiveness.

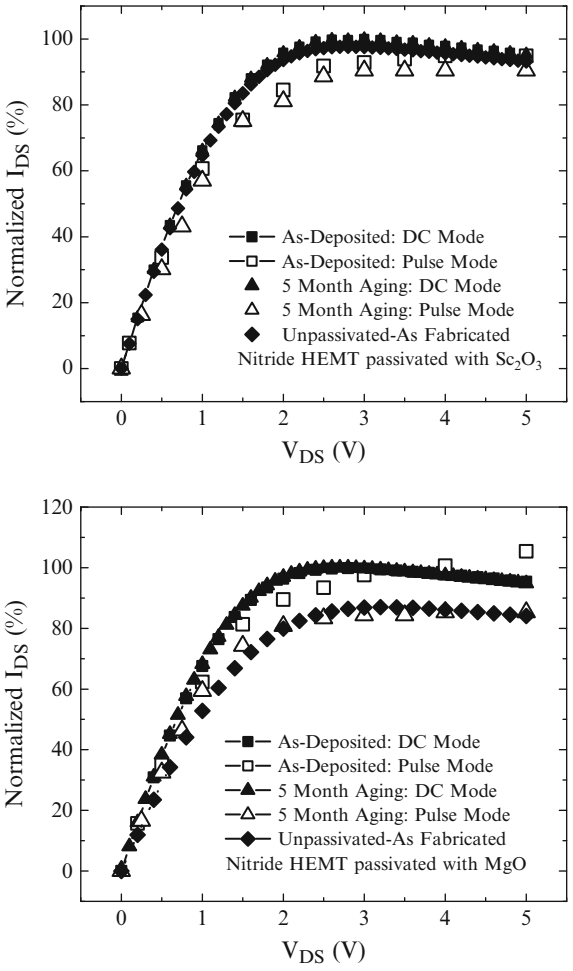



**Fig. 15.10** *Top*: Gate-lag measurements before and after $Sc_2O_3$ passivation and following 5 months aging of 1.2 × 100 μm$^2$, GaN-cap HEMTs. At top $V_G$ was switched from –5 to 0 V, while at bottom it was switched from –5 V to the value shown on the *X*-axis. *Bottom*: Gate-lag measurements before and after MgO passivation and following 5 months aging of 1.2 × 100 μm$^2$, GaN-cap HEMTs. At top $V_G$ was switched from –5 to 0 V, while at bottom it was switched from –5 V to the value shown on the *X*-axis

We have noticed in separate experiments that the MgO/GaN interface deteriorates over time if left uncapped, that is, we see increases in interface state trap density in MOS diodes on which the metal is deposited a long period after the MgO was deposited on the GaN. However, if the MgO is immediately covered by the gate metal, these changes are not observed. In a real manufacturing process, the MgO-passivated HEMTs would be covered with a dielectric such as $SiN_X$ to provide mechanical protection, and therefore MgO should not be disqualified as a candidate to provide effective passivation of GaN/AlGaN HEMTs. Obviously, further work is needed to establish the reliability of $Sc_2O_3$ and MgO passivation on these devices, but the preliminary data with $Sc_2O_3$ looks very promising.

The gate-lag data is shown in Fig. 15.11 (top) for the HEMTs before and after $SiN_X$ passivation. Note that both the hydrogenated and deuterated precursor dielectric produce a recovery of 70–75% in the drain-source current. This range

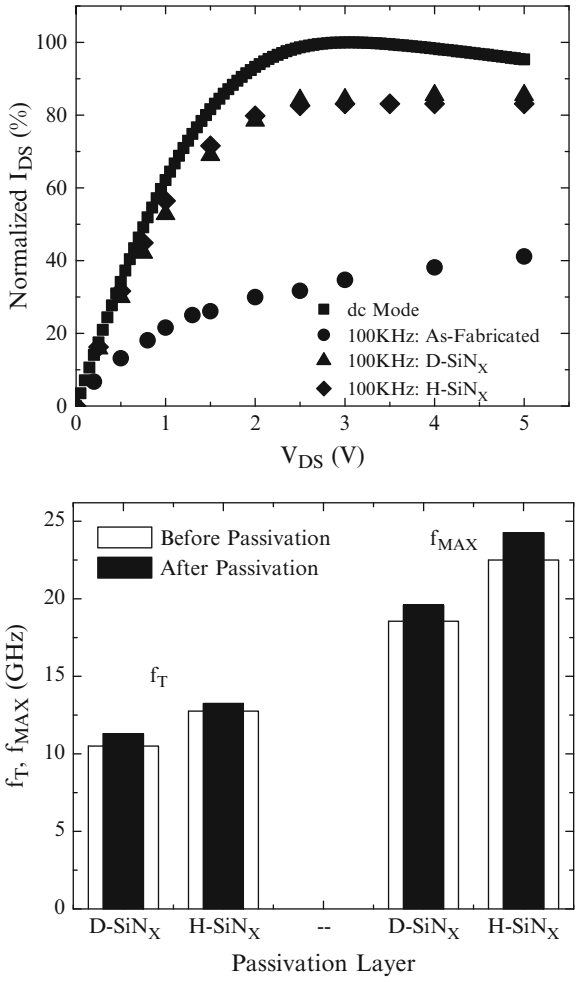



**Fig. 15.11** *Top*: Normalized $I_{DS}$ versus $V_{DS}$ for 1.2 × 100 μm$^2$, AlGaN/GaN HEMTs before and after $SiN_X$ passivation using either hydrogenated or deuterated precursors. *Bottom*: Small signal RF performance AlGaN/GaN HEMTs before and after $SiN_X$ passivation using either hydrogenated or deuterated precursors

represents results obtained from five different devices with each type of dielectric. The unity current gain frequency ($f_T$) and maximum frequency of oscillation ($f_{MAX}$) data before and after passivation are shown at the bottom of Fig. 15.11. Note that there is actually a slight increase in both parameters. There were no systematic differences between the results for hydrogenated and deuterated precursor $SiN_X$, even though the latter typically produces slightly denser films when deposited under the same PECVD conditions.

There was also an increase in forward and reverse gate currents after passivation with $SiN_X$. These results are also consistent with a decrease in surface depletion and decrease in surface trap density and indicate that there is no major degradation to the dc performance of the HEMTs upon addition of the $SiN_X$ layer. We believe the increase in gate leakage is not from the passivation itself but may originate from degradation of the gate metalization during the oxide desorption step at 350°C.

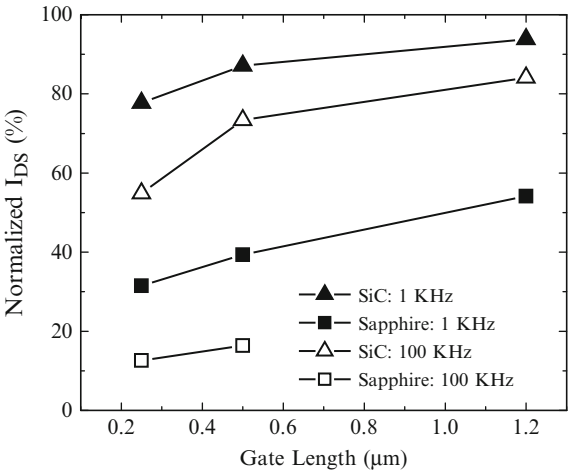



**Fig. 15.12** Normalized $I_{DS}$ as a function of both gate length and substrate type for GaN-cap HEMTs at pulse frequency of 1 and 100 KHz

The effects of both substrate type and HEMT gate length on the change in drain-source current are shown in Fig. 15.12. The shorter the gate length, the more pronounced the degradation in current because of larger surface area and higher electric field in the channel between source, S, and drain, D (the S-D distance was fixed in all cases). Figure 15.13 (top) shows typical gate-lag data for $0.5 \times 100\ \mu m^2$ AlGaN-cap HEMT grown on sapphire substrates. The decrease in drain-source current becomes more pronounced at high measurement frequencies. The degradation in current was less significant when these same structures were grown on SiC substrates, as shown in Fig. 15.13 (bottom). The defect density will be lower in the latter case due to the closer lattice match between GaN and SiC, and this appears to affect the resultant surface state density. This suggests that at least some of the surface traps are related to dislocations threading to the surface.

## 15.4 Growth and Properties of Novel Oxides for HEMT Passivation and Gate Oxide of MOSHEMTs and MOSFETs

It is clear that substantial room for improvement remains in the choice of dielectrics for the use with wide bandgap semiconductors. Based on recent work, it is believed there are some attractive materials which offer promise for this application. All of the proposed dielectrics are expected to be thermally stable based on their melting points and upon thermal stability studies we have performed on similar materials. Each of the candidate materials has a band gap which is at least 2 eV larger than the semiconductor, and all except AlN have dielectric constants comparable to or larger than the underlying semiconductor.

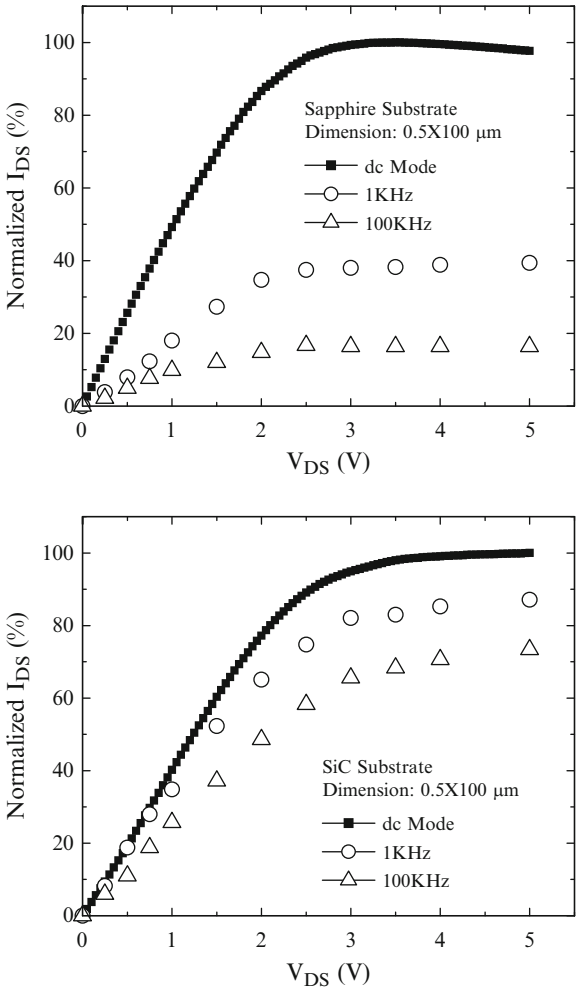



**Fig. 15.13** Gate-lag measurements on unpassivated 0.5 μm gate length, GaN-cap HEMTs grown on either sapphire (*bottom*) or SiC (*top*) substrates

### *15.4.1 Bixbyite Oxides*

In the development of GaN MOSFET power transistors, there have been investigations of the growth of $Gd_2O_3$ on GaN using elemental Gd and an ECR generated oxygen plasma. It is known that growth initiation is a critical step in the formation of epitaxial films. Four different procedures for growth initiation were tested on the (1 × 3) GaN surface produced by heating in vacuum to 700°C. First, a low temperature 300°C exposure to the oxygen plasma was studied. As shown in Fig. 15.14, the surface changes from a sharp (1 × 3) to a hazy (1 × 2). If the (1 × 3) GaN surface was exposed to the oxygen plasma at 650°C, the surface also immediately changed to the (1 × 2) surface, then after an additional 5-min

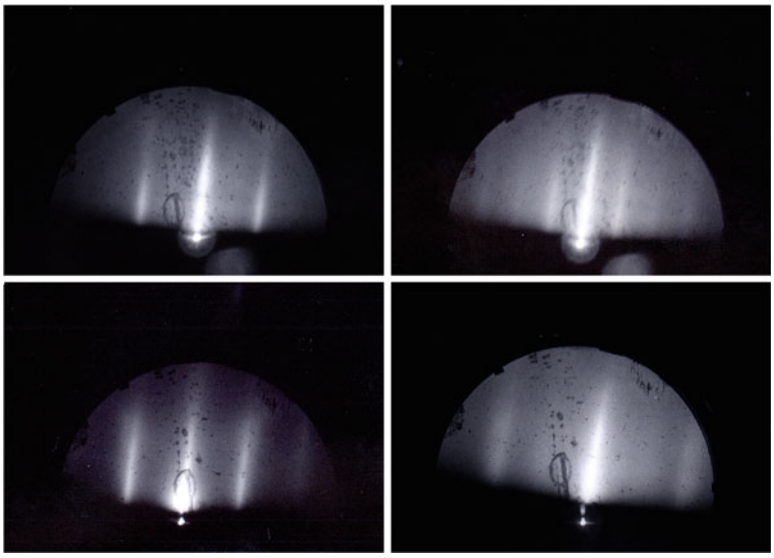


**Fig. 15.14** RHEED images of (*top*) a (1 × 3) GaN surface at 700°C and (*bottom*) the same surface at 300°C under an O plasma showing a (1 × 2)

exposure to the oxygen plasma, the surface became further reconstructed to a (4 × 4) as shown in Fig. 15.15. This (4 × 4) surface remained after the oxygen plasma was turned off. When the (1 × 3) surface was exposed to a Gd atomic beam, the RHEED pattern, seen in Fig. 15.16, is also a (1 × 2); however, the spacing of the main diffracted lines has changed indicating that the spacing between the atoms has changed and that a different crystal structure has formed.

Growth on the low temperature oxygen plasma-exposed surface initially indicated a good $Gd_2O_3$ crystal structure as indicated by RHEED, but after 10 min of growth, a symmetric RHEED pattern emerged, as seen in Fig. 15.17, indicating that the $Gd_2O_3$ crystal quality had been degraded. Reducing the growth rate produced the same symmetric pattern. Growth was also performed on the (1 × 2) GaN surface produced from the short oxygen exposure at 650°C. Here, the initial $Gd_2O_3$ RHEED pattern remained for the duration of the growth; however, extra spots were seen in the RHEED images (see Fig. 15.18) indicating that the layer was partly polycrystalline.

Further, growth experiments with reduced gadolinium effusion cell temperatures and a reduced substrate temperature of 605°C gave final RHEED patterns of the $Gd_2O_3$ crystal surface without the extra spots, indicating a uniform single crystal film. Samples grown on the high temperature oxygen plasma-treated surface (4 × 4) and the gadolinium exposed surface showed sharp $Gd_2O_3$ crystal RHEED patterns and smooth surfaces on the order of 0.5 nm RMS roughness.

High resolution TEM proved that the $Gd_2O_3$ grew on the GaN in a planar fashion as seen in Fig. 15.19 (right). There are dislocations visible in the image indicating

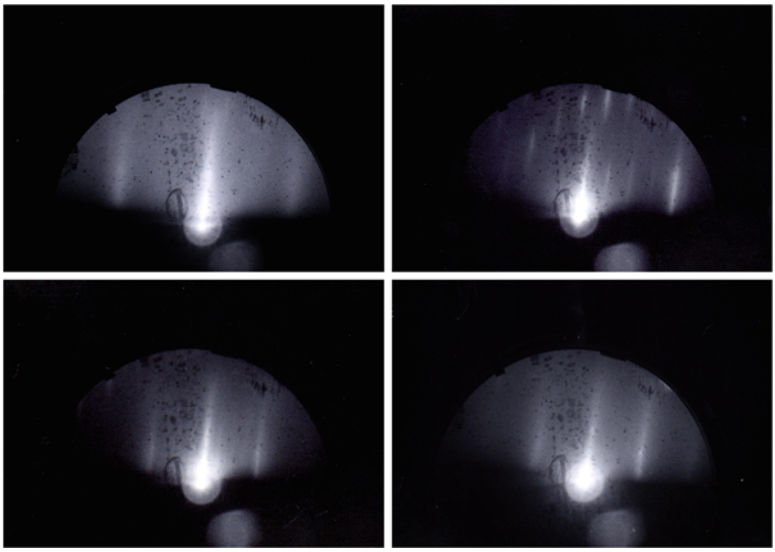


**Fig. 15.15** RHEED images of (*top*) a (1 × 3) GaN surface after receiving an O plasma exposure for 5 min at 700°C showing a faint (4 × 4) and (*bottom*) a (1 × 1)$Gd_2O_3$ surface grown at 610°C on the (4 × 4) oxygen treated surface

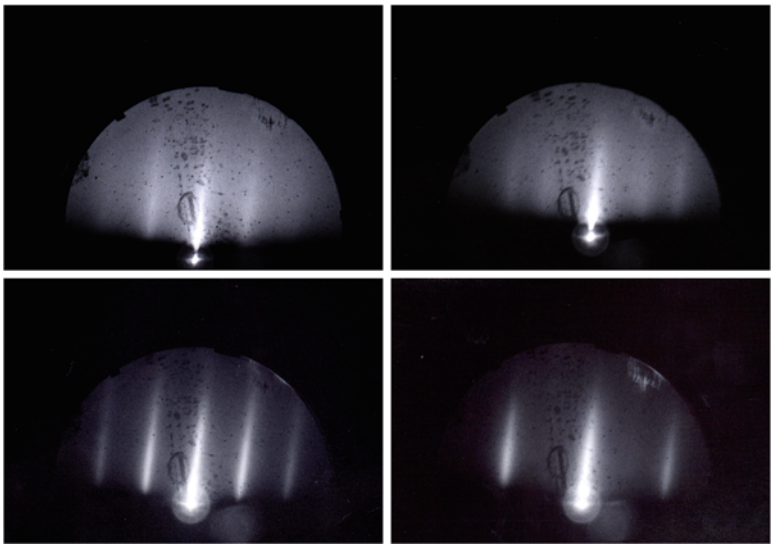


**Fig. 15.16** RHEED image of (*top*) the GaN surface after exposure to Gd for 1 min at 650°C and (*bottom*) a (1 × 1) $Gd_2O_3$ surface from a layer grown at 610°C

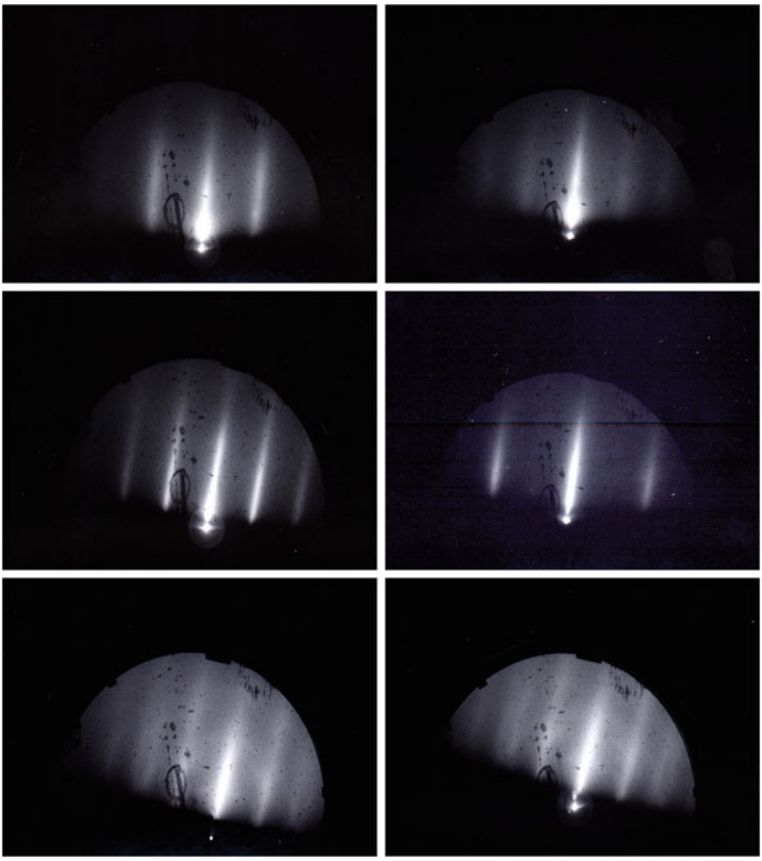


**Fig. 15.17** RHEED images of (*top*) a (1 × 2) 300°C oxygen treated GaN surface, (*middle*) a (1 × 1) $Gd_2O_3$ surface after 10 min of growth at 650°C, and (*bottom*) a symmetric pattern after 90 min of growth. The <11–20> is at left and the <1–100> is at right

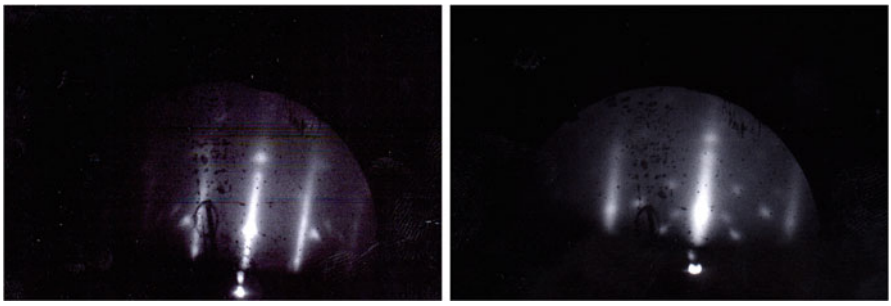


**Fig. 15.18** RHEED image of $Gd_2O_3$ (1 × 1) grown at 650°C at 0.9 nm/min indicating extra spots due to defects

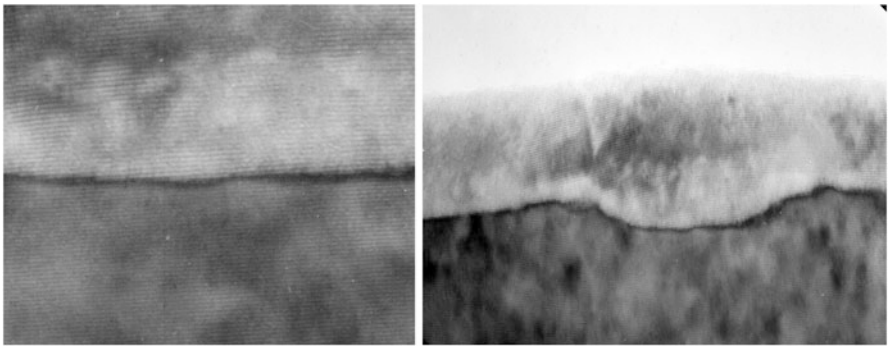


**Fig. 15.19** TEM images of $Gd_2O_3$ grown on GaN at 650 C using elemental sources. The GaN is on bottom and the $Gd_2O_3$ is on top. Notice the difference in lattice spacing. The bar length is 5.0 nm

that the film is highly defective. However, large areas of dislocation-free material were seen in XTEM. From this image, it can be seen that the $Gd_2O_3$/GaN interface was clean. The dark layer between the $Gd_2O_3$ and the GaN is a thickness effect from the ion milling process used to fabricate the TEM sample. On the left side of the image, the $Gd_2O_3$ lattice can be seen in contact with the GaN lattice. A lower resolution TEM image, as illustrated in Fig. 15.19 (left), indicated that the $Gd_2O_3$ planarized the GaN quite well. In fact, the $Gd_2O_3$ filled in the void in the GaN surface with the same registry as the entire film. Thus, there was no polycrystalline morphology in the void. There were, however, some pockets of crystal lattice that were tilted and rotated as seen in the image. This can be attributed to the dislocations arising from the single crystal material in the void meeting single crystal material on the surface. X-ray diffraction of this sample showed only one peak for the $Gd_2O_3$ that coincided with the (111) peak from published JCPDS cards. This peak, shown in Fig. 15.20, had a shoulder to the left side and had a full width at half maximum (FWHM) of 883 arcsec. The intensity of the X-ray system used was too low to resolve this shoulder in triple axis mode so the cause of this shoulder is unknown. X-ray diffraction of the sample that showed the extra RHEED spots (Fig. 15.17) is shown in Fig. 15.20. Here, a second peak was clearly seen which corresponds to the (321) plane of $Gd_2O_3$. Also, the FWHM of this (111) peak was over 1,600 arcsec. X-ray diffraction was also performed on the gadolinium initiated sample and the 5-min high temperature oxygen plasma initiated sample, seen in Fig. 15.21. The difference in the two samples was quite large. The gadolinium initiated growth showed a FWHM of 1,023 arcsec. for the (111) peak. The high temperature oxygen plasma initiated growth showed a very broad peak that was several thousand arcseconds wide. As a result of this study, it was determined that Gd initiation is the best method for growth initiation and that single crystal epitaxial $Gd_2O_3$ can be deposited on GaN. The optimized process was used to fabricate the devices mentioned in the previous section.

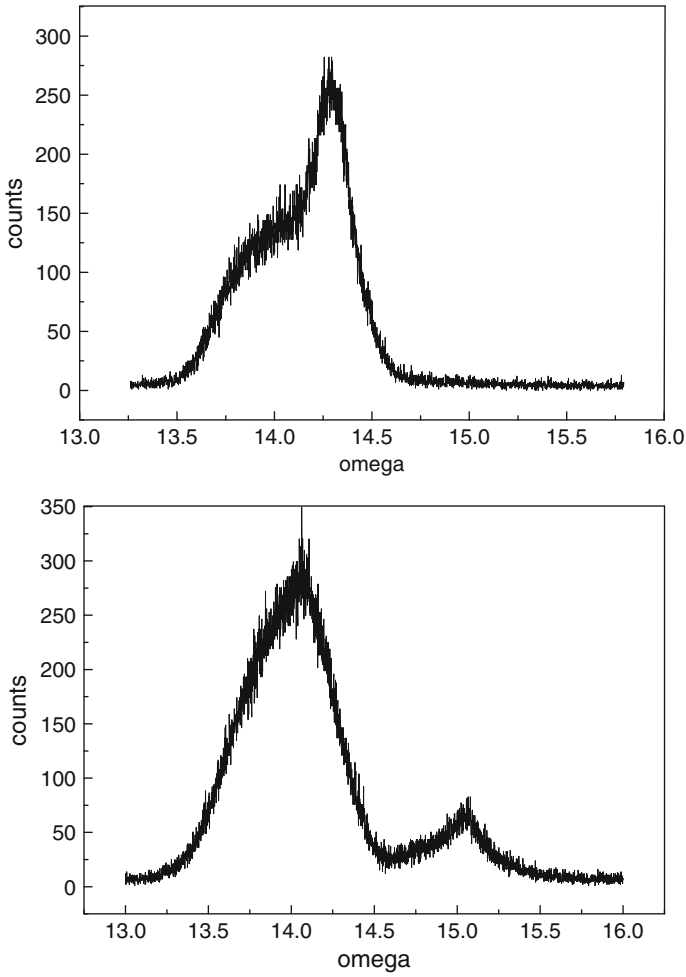



**Fig. 15.20** X-ray diffraction plots of $Gd_2O_3$ grown on GaN at 650°C. The FWHM of the peak at left is 883 arcsec and at right is 1,623 arcsec. The RHEED pattern from the film at right showed spots indicative of polycrystalline structure

### *15.4.2 $Sc_2O_3$*

$Sc_2O_3$ exists in the same Bixbyite structure as $Gd_2O_3$ but has a much smaller lattice constant, see Table 15.1, which should make it less defective when deposited on GaN. Like $Gd_2O_3$, it has a high melting point suggesting good thermal stability. The material is often used as an optical coating for high power photonic devices and has a band gap of 6.3 eV. Based upon the reported work function of 4 eV, it is likely that the electron affinity is ~0.85 eV. If this estimate is correct, then, unlike $Gd_2O_3$, the band offsets for $Sc_2O_3$ should allow fabrication of devices on both n- and p-type

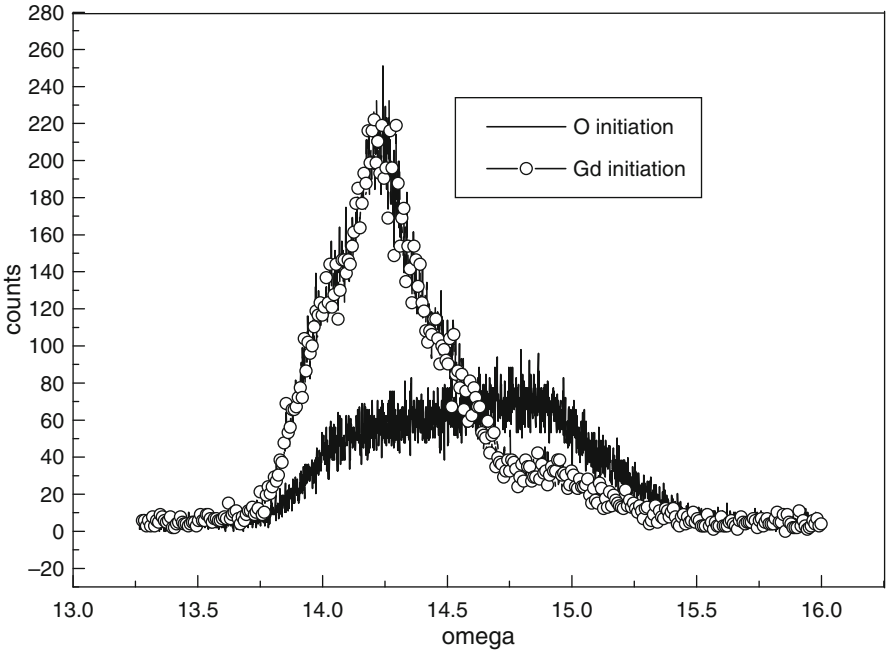



**Fig. 15.21** XRD plot of samples grown on a Gd exposed GaN surface and a 5 min high temperature O plasma-exposed GaN surface. FWHM of the Gd initiated sample is 1,058 arcsec

material. The dielectric constant is also favorable as it is slightly larger than that of $Gd_2O_3$, 14 vs. 11.

In light of the success we have had with $Gd_2O_3$, we believe that $Sc_2O_3$ has a high probability of showing improvement over existing dielectrics. By limiting the $Sc_2O_3$ thickness to ~10 nm, we should be able to minimize the defect concentration. However, a large mismatch still exists in this system, and we expect that an amorphous overlayer will still be required. We have found that reducing the $Gd_2O_3$ deposition temperature to 100°C is sufficient to produce amorphous material on GaN. However, when this material is deposited on crystalline $Gd_2O_3$, a poly-crystalline layer is produced. It is possible that using an e-beam source of $Gd_2O_3$ will help to suppress the surface mobility and allow for the deposition of amorphous material in spite of the template effect of the crystalline $Gd_2O_3$.

### *15.4.3 MgO/CaO*

MgO is a rock salt dielectric which has been explored as an intermediate buffer layer for growth of ferroelectric materials on semiconductors or as a potential gate dielectric for GaAs or Si. While MgO deposition by MBE has been successfully

demonstrated by a number of groups, the crystal quality of the films deposited on GaAs and Si has been poor due to the large lattice mismatch between the MgO and the semiconductor substrate. GaN has a smaller lattice constant than GaAs and is thus a much closer match to MgO. CaO is also a rock salt dielectric but with a substantially larger lattice spacing than MgO. A ternary oxide consisting of MgO and CaO could thus in principle be closely matched to the GaN resulting in a substantial reduction in defect density. Similarly, a lattice matched alloy should also be possible for SiC. An additional advantage of this system is the large band gap and, thus, large band offsets are expected relative to either an n- or p-type semiconductor. Further, the dielectric constants for both MgO (9.8) and CaO (11.8) are substantially higher than for $SiO_2$.

One problem with this material system is the large difference in ionic radius between Mg and Ca, which causes severe immiscibility. At the temperature regime needed for this work, the solubility of the compounds is minimal. However, the use of MBE as the deposition method often allows for the formation of metastable phases. In fact, it has been demonstrated that the MgCaO ternary can be grown with compositions covering the entire compositional range when deposited at a temperature of 300°C.

It is possible to use two approaches to form lattice matched MgCaO on GaN and SiC. One employs low growth temperatures to overcome the immiscibility while the other relies on alternating layers of MgO and CaO, often referred to as the digital alloy approach, to form a material with the properties of the average composition of the film. On GaN, the layers would be alternated sequentially, while on SiC, the ratio of MgO layers to CaO layers would be 3:1. An additional consideration with the SiC substrates is the possibility of $SiO_2$ formation at the SiC/oxide interface. By initiating growth with a metal layer, the formation of $SiO_2$ can be suppressed.

### *15.4.4 Amorphous AlN*

While crystalline AlN shows some promise as a gate dielectric, it still suffers from leakage and low breakdown due to dislocations and other defects. Presumably, these arise from the small but finite lattice mismatch which exists between AlN and SiC or GaN. An alternative approach is the use of amorphous AlN. It is expected that reducing the growth temperature should suppress the surface mobility sufficiently to prevent nucleation of a crystalline phase. Rogers et al. (unpublished) have in fact shown that by using a growth temperature of 200°C in the CVD deposition of AlN on Si it is possible to reduce the leakage and improve the breakdown relative to crystalline AlN grown on the same substrate material. The absence of XRD peaks in their material suggests that the deposited material is amorphous.

Amorphous AlN has also been deposited using a gaseous Al precursor and an RF nitrogen plasma in a metal-organic molecular beam epitaxy (MOMBE) system. As shown in Fig. 15.22, it was possible to deposit material at 325°C in which

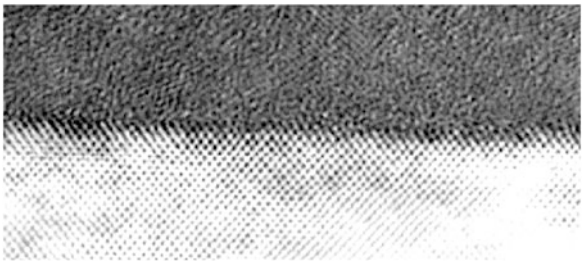


**Fig. 15.22** XTEM of AlN deposited on Si at 325°C showing crystalline pockets dispersed in an amorphous matrix

approximately 30% of the volume is amorphous. On SiC, this has produced MIS capacitors with a gate breakdown of 6.8 MV/cm and C-V analysis shows little or no evidence of hysteresis. Similar experiments on GaN were not as successful, with breakdowns of only 1 MV/cm attainable. This is most likely due to enhanced crystallinity from the template effect of the GaN substrate.

In order to improve the electrical behavior of the AlN/SiC and AlN/GaN structures, it will be necessary to further reduce the crystallinity of the material. This can most likely be accomplished by further reducing the growth temperature. In order to accomplish this, it will be necessary to switch from the gaseous Al precursor, which has a lower thermal decomposition temperature of ~325°C, to an elemental Al effusion oven which can in principle be used at much lower substrate temperatures. In addition to using the amorphous AlN as a sole dielectric, it can also be incorporated into stacked gate structures employing the crystalline oxides discussed previously.

## 15.5 Insulators for GaN MOS and MISFETs

GaN MOSFETs are far less developed than for Si or even SiC. Most work in this area has focused on the amorphous dielectrics, and in particular on $SiO_2$. Silicon oxide deposited by plasma-enhanced chemical vapor deposition (PECVD) has been reported to give interface state densities on the order of low $1 \times 10^{11}$ $eV^{-1}$ $cm^{-2}$. Silicon oxide deposited by electron-beam (EB) evaporation has shown an interface state density of $5.3 \times 10^{11}$ $eV^{-1}$ $cm^{-2}$. This dielectric has also been employed as the insulator in an insulated gate heterostructure FET (IG-HFET) by Khan et al. [41]. While these devices are promising, the low dielectric constant of $SiO_2$ and the absence of modulation at positive gate voltages (forward bias) above 3–4 V in $SiO_2$/GaN devices suggest that other alternatives should be considered.

Silicon nitride is another amorphous dielectric which has been explored. Deposition by PECVD on GaN resulted in an interface state density of $6.5 \times 10^{11}$ $eV^{-1}$ $cm^{-2}$. Electrical measurements showed the MISFET structure had a large flat band voltage shift (3.07 V) and a low breakdown voltage (1.5 MV/cm) of the dielectric. A unique dielectric structure of $SiO_2/Si_3N_4/SiO_2$ (ONO) was reported to have a breakdown field strength of 12.5 MV/cm for temperatures as

high as 300°C. The ONO structure was deposited by jet vapor deposition to a thickness of 10 nm/20 nm/10 nm. This multilayer structure allows for unique engineering of a dielectric and will most likely receive more attention in the future. However, the dielectric constant of the $SiO_2$ is still a disadvantage in this structure.

Gallium oxide is yet another material which has been investigated. Thermal oxidation of the GaN surface to form $Ga_2O_3$ has been performed in dry and wet atmospheres. Dry oxidation of GaN epilayers at temperatures below 900°C showed minimal oxidation. At temperatures above 900°C, a polycrystalline monoclinic $Ga_2O_3$ forms at a rate of 5.0 nm/h. This oxidation rate is too slow to be viable as a processing step. Wet oxidation of GaN also forms polycrystalline monoclinic $Ga_2O_3$ but at a rate of 50.0 nm/h at 900°C. From cross-sectional transmission electron microscopy (XTEM), the interface between the oxide and the GaN is found to be non-uniform. Scanning electron microscopy (SEM) shows that both films are rough and facetted. Electrical characterization of the oxide films shows the dry oxide dielectric has a breakdown field strength of 0.2 MV/cm and the wet oxide dielectric field strength of 0.05–0.1 MV/cm. As with GaAs, it does not appear that $Ga_2O_3$ will be a viable dielectric for GaN.

Due to the recent success of gallium gadolinium garnet (GGG) as a dielectric in GaAs MOSFETs, attention has turned toward this as a dielectric material for GaN. This material is amorphous when deposited by e-beam evaporation. The large dielectric constant of this material is particularly attractive and the bandgap, though significantly lower than that of $SiO_2$, is expected to be adequate for n-type material. Using e-beam evaporation in an MBE environment for deposition of the dielectric, a $Ga_2O_3(Gd_2O_3)$/GaN depletion mode MOSFET was fabricated [23]. While the device showed significantly less gate leakage at elevated temperature relative to a conventional MESFET fabricated on the same GaN layer, the device still could not be modulated at voltages above roughly 3 V. This again suggests problems with leakage at the dielectric/GaN interface.

In addition to the amorphous dielectrics, crystalline AlN has also been considered. Aluminum nitride deposited by MBE and MOCVD has been used to create MISFET devices and IG-HFET devices. The AlN gate in the MISFET structure grown by MOMBE at 400°C was polycrystalline and resulted in a dielectric breakdown field of 1.2 MV/cm [25]. The AlN IG-HFET structure was grown at 990°C, forming a single crystal film of 4.0 nm. This device operated in enhancement mode and had a pinch-off voltage of 0 V. Unlike the amorphous dielectrics, single crystal AlN and polycrystalline AlN films are expected to suffer from defects and grain boundaries that reduce the breakdown field sustainable in the material.

An attractive oxide alternative to both GGG and $SiO_2$ is $Gd_2O_3$, which also has a high dielectric constant, 11.4, and a band gap of 5.3 eV. This oxide exists in the $Mn_2O_3$ (Bixbyite) crystal structure which exhibits similar atomic symmetry in the (111) plane as the GaN (0001) basal plane. This similarity in symmetry offers the possibility to grow an epitaxial dielectric. Single crystal growth of $Gd_2O_3$ on GaAs has, for example, been demonstrated. However, in the case of GaAs, the expected orientation between the $Gd_2O_3$ and the substrate is (100) rather than (111). Further, the bond length mismatch between the $Gd_2O_3$ and GaAs (100) planes is

smaller than between the $Gd_2O_3$ (111) and GaN (0001), 4.2% vs. 20%. In spite of this mismatch, we have recently demonstrated that $Gd_2O_3$ can be deposited epitaxially on GaN using elemental Gd and an ECR oxygen plasma in GSMBE. The dislocations present in the film limit the breakdown field which can be sustained in the dielectric and necessitate the use of an $SiO_2$ overlayer in order to fabricate MOSFETs from this material. Transistors with 1-μm gate length fabricated using the $SiO_2/Gd_2O_3$ dielectric stack showed modulation up to a forward gate bias of 7 V [28]. This is the highest modulation yet reported for a GaN MOSFET. Thus, it appears that the amorphous dielectrics, in particular GGG, produce the best breakdown fields, while the $Gd_2O_3$ appears to produce the best interface with GaN. This is not surprising since in general one would expect the amorphous dielectrics to contain fewer defects and thus withstand higher fields before breakdown. By contrast, the crystalline dielectrics appear to be more amenable to the formation of a low defect interface and are better suited to passivation of the surface dangling bonds but suffer from high concentrations of bulk defects which limit the breakdown field. Both GGG and $Gd_2O_3$ appear to be stable at temperatures up to 950–1,000°C suggesting that both the interface and the bulk microstructure are capable of withstanding high temperature processing. It should be noted that both GGG and $Gd_2O_3$ are not well suited to fabrication of devices on either p-SiC or p-GaN since the valence band offset to these materials is expected to be either very small or in fact negative.

### *15.5.1 Gate-Controlled MOS Diodes*

Fabrication of gate-controlled MOS diodes on GaN is an excellent method for establishing the quality of the dielectrics. The use of MOS gates should enhance the performance of GaN/AlGaN transistors, which already shows excellent potential [3, 17, 19–21, 24, 28, 29, 31–34, 37–39, 66–91]. To fabricate gate-controlled diodes using the oxides as gate dielectrics, the starting sample was a 1-μm-thick p-GaN (hole concentration is ~ $2 \times 10^{17}$ $cm^{-3}$ at 25°C) layer grown on a 2-μm undoped GaN buffer grown on a $Al_2O_3$ substrate by metal-organic chemical vapor deposition. Both MgO and $Sc_2O_3$ were used as the gate dielectrics. The $Sc_2O_3$ was deposited by RF plasma-activated molecular beam epitaxy(MBE) at 650°C using elemental Sc evaporated from a standard effusion cell at 1,130°C and $O_2$ derived from an electron cyclotron plasma source operating at 200 W forward power (2.45 GHz) and $10^{-4}$ Torr. The $Sc_2O_3$ layers were ~400 Å thick. The smaller mismatch between the $Sc_2O_3$ and the GaN (0001) should lead to a lower defect density, and in fact, the FWHM in the X-ray diffraction scan is substantially reduced at 514 arcsec as compared to the 883 arcsec found for comparable $Gd_2O_3$ layers, as shown in Fig. 15.23.

The MgO was deposited by MBE at 100°C. The MgO precursors were elemental Mg and radio-frequency plasma-activated oxygen. The GaN samples were cleaned initially with a 3-min chemical etch in $HCl/H_2O$(1:1), $H_2O$ rinse, UV-ozone exposure for 25 min, rinse in buffered oxide etch solution (6:1, $NH_4F/HF$), and rinsed in $H_2O$. The samples were then loaded into the MBE system and heated at 650°C to

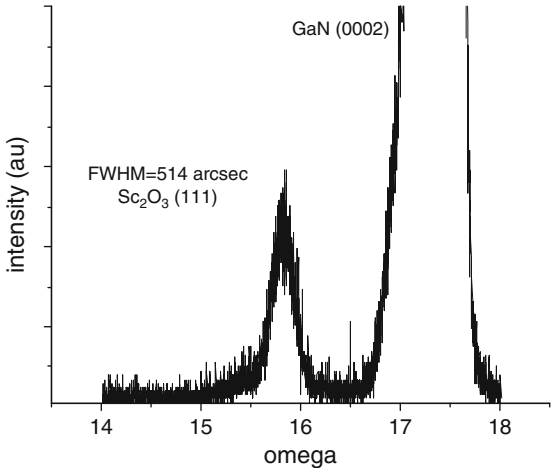



**Fig. 15.23** HXRD scans of $Sc_2O_3$ on h-GaN

ensure oxide removal. A standard effusion cell operating at 380°C was used for evaporation of the Mg, while the $O_2$ source was operated at 300 W forward power (13.56 MHz) and $2.5 \times 10^{-5}$ Torr. The MgO layers were ~800 Å thick. In separate measurements, we obtained interface state densities of 2–3 $\times$ $10^{11}$cm$^2$eV$^{-1}$ from the AC conductance and Terman methods [39–43].

$N^+$ regions were created by selective implantation of $Si^+$ at multiple energies and doses (70 keV, $2 \times 10^{13}$ cm$^2$, 195 keV, $6 \times 10^{13}$ cm$^{-2}$ and 380 keV, $1.8 \times 10^{14}$ cm$^{-2}$). The junction depth was ~0.4 μm from ion range simulations. The samples were then annealed at 950°C under $N_2$ to activate the Si-implanted regions. Windows were etched into the oxide and e-beam deposited p-Ohmic (Ni/Au), n-Ohmic (Ti/Al/Pt/Au), and gate metal (Pt/Au) were patterned by lithography. The separation of the $n^+$ regions was ~60 μm. A schematic of the completed gate-controlled diode is shown in Fig. 15.24, while Fig. 15.25 shows a plan view photograph of the completed device.

Figure 15.26 shows C-V characteristics of the MgO/GaN MOS-controlled diodes at 25°C in the dark as a function of the measurement frequency. In each case, −20 V was applied at the gated contact to provide a source of minority carriers. The frequency dispersion observed in inversion is due to the resistance of the inversion channel. At 5 KHz measurement frequency, we observe only deep depletion since the characteristics are dominated by majority carriers. As the frequency is decreased, a clear inversion behavior is observed due to charge flow into and from the $n^+$ regions external to the gate. In diodes without the $n^+$ regions to act as an external supply of minority carriers, we could not observe inversion, even up to measurement temperatures of 30°C. Similar results were obtained for $Sc_2O_3$/GaN diodes, as shown in Fig. 15.27. The shift in the "hooks" in the capacitance-voltage curves can be used to establish the total surface state density [41, 42].

Charge-pumping experiments can also be used to measure the surface state density. This density, $N_{SS}$, can be related to the pulse frequency $\upsilon$ and charge-pumping current $I_{CP}$ through the relation

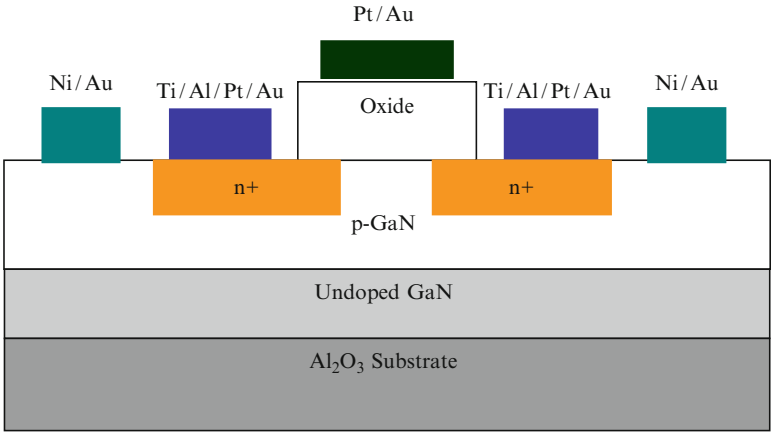



**Fig. 15.24** Schematic of $Sc_2O_3$/p-GaN gate-controlled diodes (Reprinted with permission from [41]. Copyright 2002 American Institute of Physics)

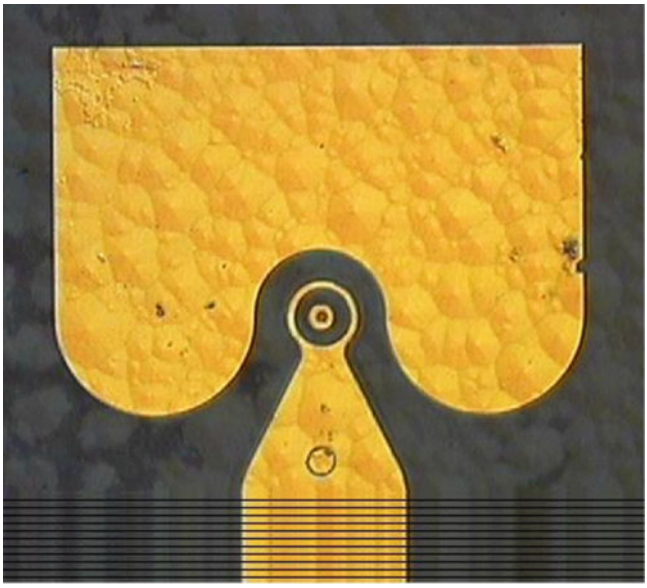


**Fig. 15.25** Photograph of completed gate-controlled diode (Reprinted with permission from [41]. Copyright 2002 American Institute of Physics)

$$N_{SS} = \frac{I_{CP}}{\upsilon A_C e}$$

where $A_C$ is the diode channel area and $e$ is the electronic charge. Figure 15.28 shows typical data measured at $\upsilon = 50$ kHz. From this data, we obtain $N_{SS} = 3 \times 10^{12}$ $cm^{-2}$ at the $Sc_2O_3$/p-GaN interface. Note that no effort has been made in our

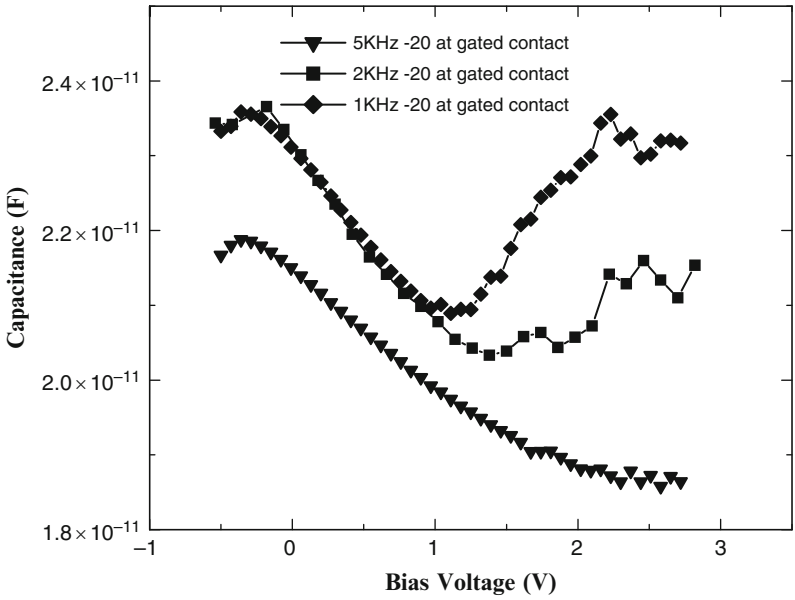



**Fig. 15.26** C-V characteristics of MgO/GaN MOS gate-controlled diodes at 25°C as a function of measurement frequency(+15 V bias in gated contact in each case) (Reprinted with permission from [41]. Copyright 2002 American Institute of Physics)

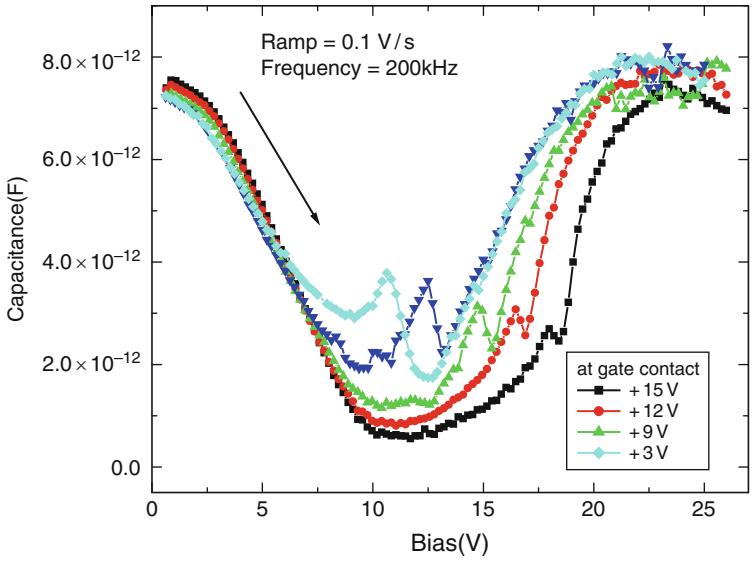



**Fig. 15.27** C-V characteristics of $Sc_2O_3$GaN MOS gate-controlled diodes at 25°C as a function of bias on the gate contact

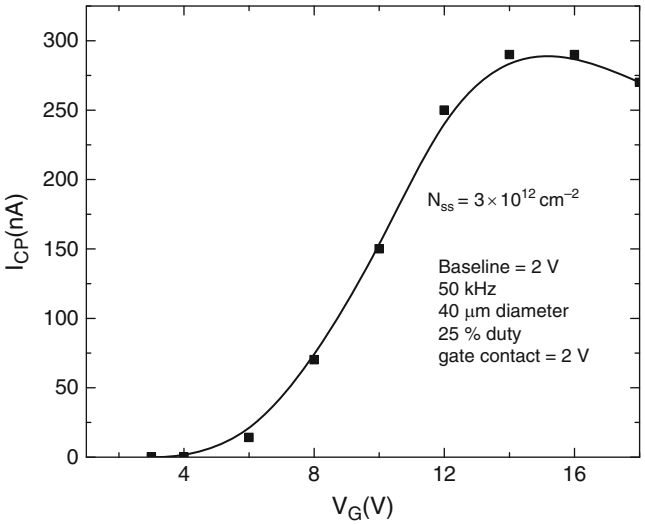



**Fig. 15.28** Charge-pumping current as a function of pulsed gate voltage

structures to passivate surface states through forming gas anneals or plasma hydrogenation. This is the first measurement of surface state density in GaN-based MOS structures that more fully approximates a transistor, that is, the structures show inversion and have gone through an extensive device processing sequence.

### *15.5.2 n-Channel Depletion Mode GaN MOSFET Using Stacked Gate Dielectric of $SiO_2/Gd_2O_3$*

The nitride layer structure consisted of 2-μm unintentionally doped GaN and a 700-Å Si-doped GaN channel grown by MOCVD on a c-$Al_2O_3$ substrate. Hall measurement indicated a channel doping level in the low-$10^{18}$ cm$^{-3}$. The wafer was prepared for oxide growth by a 3-min dilute HCl dip, 25-min UV-ozone exposure, and 5-min dilute HF dip. The $Gd_2O_3$ was grown in a Riber 2300 gas-source molecular beam epitaxy (GSMBE) system equipped with a Wavemat MPDR 610 electron cyclotron resonance (ECR) plasma source. After loading, the wafer was heated to 700°C to desorb the native oxides, then cooled to the growth temperature of 650°C. Growth proceeded at $1 \times 10^{-4}$ Torr oxygen pressure in the ECR plasma head at 200 W forward power, with a Gd cell temperature of 1,230°C. X-ray diffraction showed a single crystalline $Gd_2O_3$ structure for the 700 Å film.

A schematic of the processing sequence and final device structure of the planar n-channel depletion mode GaN MOSFET is given in Fig. 15.29 [84]. Device fabrication began with a wet etch of $Gd_2O_3$ to open source and drain ohmic contact

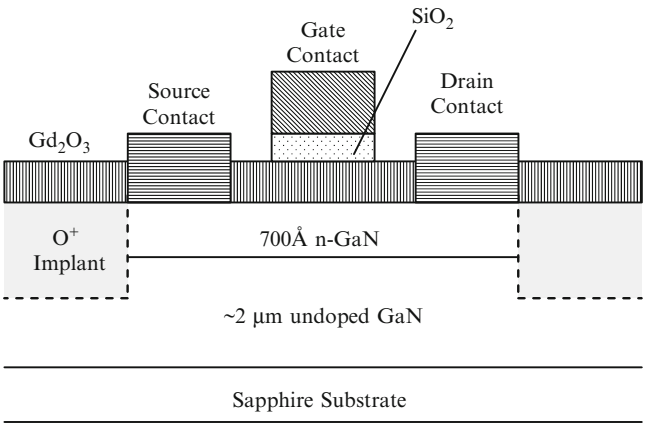



**Fig. 15.29** Schematic of stacked-gate GaN MOSFET

windows using HCl as the etchant and photoresist AZ-1818 as the mask. Ti(200 Å)/Al(600 Å)/Au(1000 Å) ohmic contacts were deposited by electron-beam evaporation. Oxygen implantation was used for device isolation. A similar $O^+$ implantation scheme has previously been shown to yield sheet resistance in GaN on the order of $10^{12}$ Ohm per square. To enhance the gate breakdown voltage, a 300 Å layer of $SiO_2$ was electron-beam evaporated onto the $Gd_2O_3$ surface in the photolithographically defined gate contact region. This was followed under vacuum by evaporation of Ti(100 Å)/Au(1000 Å) Schottky contacts to the $SiO_2$ surface.

When deposited under optimum conditions, $Gd_2O_3$ was grown epitaxially on GaN, as verified by reflection high energy electron diffraction (RHEED) and cross-sectional transmission electron microscopy (XTEM). The interface was quite smooth despite a somewhat rough nitride surface. However, there was still a relatively large difference in lattice spacing between $Gd_2O_3$ and GaN, which contributed to a high number of dislocations in the oxide film. These dislocations act as current leakage paths and caused the breakdown voltage of the as-grown $Gd_2O_3$ to be rather low. In order to improve the breakdown for device fabrication, amorphous $SiO_2$ was deposited on the $Gd_2O_3$ surface. The formation of a stacked gate dielectric of $SiO_2/Gd_2O_3$ allowed the interfacial properties of $Gd_2O_3$/GaN to be maintained, while reducing leakage by terminating the dislocations in the crystalline oxide. Figure 15.30 (top) illustrates the significant reduction in gate leakage after $SiO_2$ deposition. The breakdown field of the insulator improved from 0.3 to 0.8 MV/cm. Gate reverse leakage current density was $5 \times 10^{-6}$A/cm$^2$ at a gate-source bias ($V_{GS}$) of $-10$ V and remained below 10 nA at $V_{GS} = -70$ V for the $1 \times 200$ μm$^2$ device, clearly demonstrating the benefit of the insulated gate MOS structure.

Charge modulation from accumulation to depletion was observed. The absence of surface inversion is due to the wide band gap and correspondingly low room

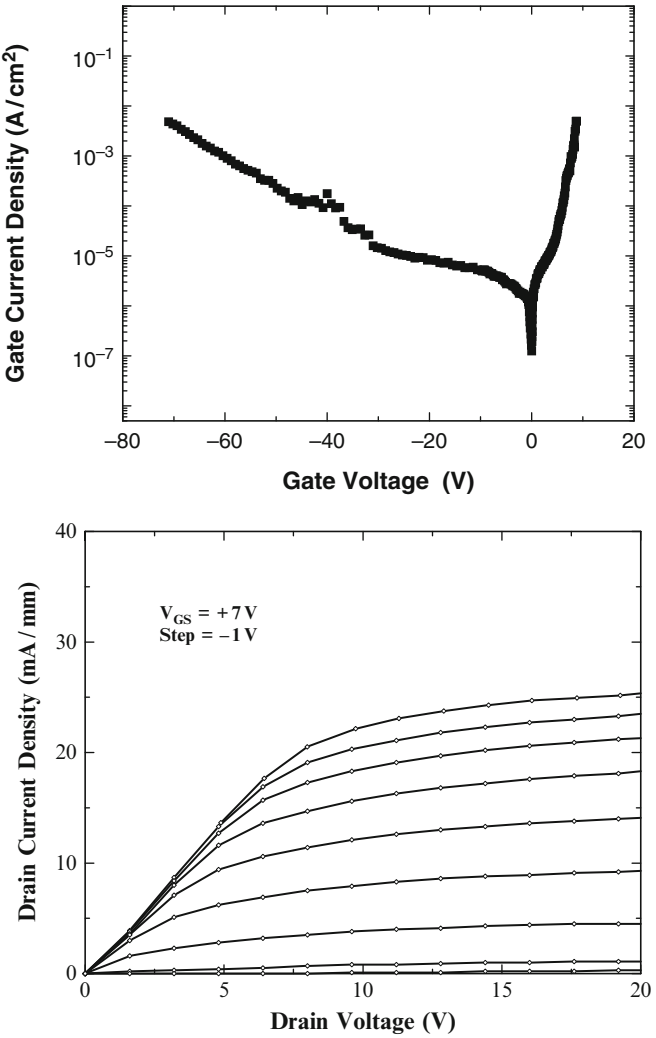



**Fig. 15.30** DC output characteristics of 1 × 200 μm$^2$ gate dimension GaN MOSFET

temperature minority carrier generation rate of GaN, consistent with previous reports on wide band gap MIS structures. Drain I-V characteristics of a 1 × 200 μm$^2$ gate dimension $SiO_2/Gd_2O_3$/GaN MOSFET were measured with an HP4145A and are shown in Fig. 15.30 (bottom). The drain-source breakdown exceeds 80 V. For these high breakdown devices, charge modulation could be demonstrated for gate voltages from +2 to −5 V. Other depletion mode devices of identical geometry could be modulated at an accumulation bias of +7 V, verifying the high quality of the $Gd_2O_3$/GaN interface. A maximum intrinsic transconductance of 61 mS/mm was measured at $V_{GS} = -0.5$ V and $V_{DS} = 20$ V.

### 15.5.3 MgO/p-GaN Enhancement Mode Metal-Oxide-Semiconductor Field-Effect Transistors (MOSFETs)

The sample structure consisted of ~2 μm thick Mg-doped p-GaN (p ~$10^{16}$ $cm^{-3}$) grown on c-plane sapphire substrate. In order to create the $n^+$ source and drain-contact regions for MOSFET fabrication, these samples were implanted with $^{27}Si^+$ ions at multiple energy conditions (10 keV, $1.6 \times 10^1$ $cm^{-2}$; 30 keV, $3.6 \times 10^{13}$ $cm^{-2}$; 70 keV, $7.2 \times 10^{13}$ $cm^{-2}$; 170 keV, $1.3 \times 10^{14}$ $cm^{-2}$; 340 keV, $2.2 \times 10^{14}$ $cm^{-2}$) at room temperature. This implant scheme was designed to produce a uniform Si profile to depth of ~0.3 μm [92]. The samples were encapsulated with 500 nm of $SiO_2$ deposited by plasma-enhanced chemical vapor deposition and annealed at 1,200°C under a flowing $N_2$ ambient. After HF removal of the $SiO_2$, the MgO was deposited by MBE at 100°C on the p-GaN layer as the gate dielectric. The MgO precursors were elemental Mg and radio-frequency plasma-activated oxygen. The GaN samples were cleaned initially with a 3-min chemical etch in $HCl/H_2O$ (1:1), $H_2O$ rinse, UV-ozone exposure for 25 min, rinse in buffered oxide etch solution(6:1, $NH_4F/HF$), and rinsed in $H_2O$. The samples were then loaded into the MBE system and heated at 650°C to ensure oxide removal. A standard effusion cell operating at 380°C was used for evaporation of the Mg, while the $O_2$ source was operated at 300 W forward power (13.56 MHz) and $2.5 \times 10^{-5}$ Torr. The deposited MgO layers were ~80-nm thick. Following this step, the MgO was removed in the source, the drain, and the p-epi contact areas, using $H_3PO_4$ etchant and an S-1818 resist mask. The e-beam deposited n-Ohmic (Ti/Al/Pt/Au), p-Ohmic(Ni/Au), and gate(Pt/Au) metals were followed by lift-off. A schematic of the final structure is shown in Fig. 15.31.

Figure 15.32 shows the typical gate I-V characteristics of the MgO/GaN MOSFET. For the oxide thickness of 80 nm, the corresponding oxide breakdown field was 1.75 MV/cm at a current density of 5 $mA/cm^2$. The I-V characteristic is shifted due to the presence of oxide charge. The gate contact exhibited good rectification, demonstrating the MgO produces excellent insulator characteristics and a much larger breakdown than if the Pt/Au was placed directly on the p-GaN

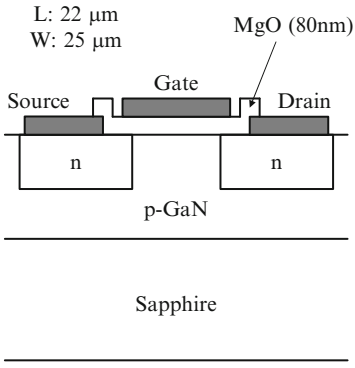



**Fig. 15.31** Schematic cross section of the MgO/p-GaN MOSFET (*top*) and a microscope image of the device layout (*bottom*)

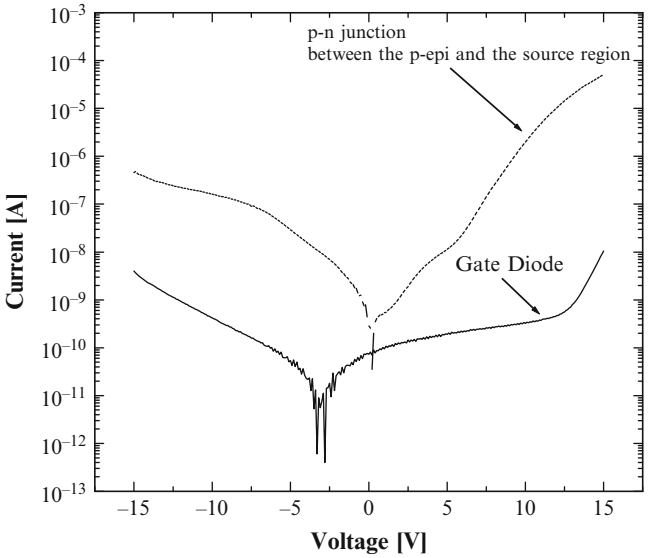



**Fig. 15.32** Transfer characteristics of MgO/p-GaN MOSFET at a drain-source voltage of 3 V

without any MgO. In that case, the forward turn-on voltage was <0.7 V and the reverse breakdown voltage was <1 V. In addition, the p-n junction between the p-GaN and the n-type source showed excellent rectifying characteristics (also shown in Fig. 15.32), demonstrating that the Si activation efficiency was sufficient to produce an abrupt junction .Room temperature capacitance-voltage characteristics measured at 1 MHz and with a sweep rate of 30 mV/s performed on companion diodes showed clear modulation from accumulation to depletion. Using the relation $C_{ox} = \varepsilon_o \varepsilon_{ox} A/T_{ox}$ (where $\varepsilon_o$ is the permittivity in a vacuum, $C_{ox}$ is the oxide capacitance, $A$ is the cross-sectional area of oxide, and $T_{ox}$ is the oxide thickness), the dielectric constant of the oxide, $\varepsilon_{ox}$, was calculated to be 9.9, a value which is in agreement with the tabulated value for MgO (9.8). A significant flat band voltage shift was observed, which indicated the existence of fixed charges in the oxide layer.

The drain-source I-V characteristics from the MOSFET showed the channel was formed at a gate voltage of > +6 V and the device could be modulated to + 12 V of gate voltage. There is no indication of negative resistance effects, and thus, self-heating does not appear to be a problem under the biasing conditions employed here. The drain current $I_{DS}$ can be expressed as [93]

$$I_{DS} = \mu_n \cdot C_i \frac{W}{L}(V_G - V_T)V_D$$

where $C_i$ is the oxide capacitance per unit area, $L$ and $W$ are the channel length and channel width, respectively, $\mu_n$ is the low field mobility, $V_G$ is the gate voltage,

$V_D$ is the drain voltage, and $V_T$ is the threshold voltage. The transconductance of the MOSFET (i.e., the change in drain current for a change in gate voltage) is maximized for high electron mobilities, thinner MgO dielectric layers, and larger values of $W/L$. MEDICI™ simulation reported previously indicate that the drain current should be in the mA range for the structure used here, and the lower currents obtained show that we need to improve carrier mobility significantly [94] and probably also the N-doping level obtained in the source/drain regions, which affects the series resistance. The maximum drain current was ~0.2 μA at 12 V gate voltage and the transconductance was 5.4 μS/mm . This is comparable to the first results for e-mode MOSFETs in GaAs, using $GdGa_2O_3$ gate oxides [95]. Clearly, much more work has to be done to increase the channel mobility. It is easier to activate n-type implants in GaN than p-type implanted dopants, but in principle, it should also be possible to use n-type epi and create p-type source and drain regions [96].

### *15.5.4 Radiation Damage Study on GaN MOS Diodes*

MgO/GaN metal-oxide-semiconductor diodes were irradiated with 40-MeV protons at a fluence of $5 \times 10^9$ $cm^{-2}$, simulating long-term (10 years) exposure in space-borne applications. The result of the proton irradiation was a decrease in device capacitance, consistent with the creation of deep electron traps that reduce the effective channel doping and also a decrease in breakdown field from ~$10^6$ V/cm in control devices to $0.76 \times 10^6$ V/cm in devices irradiated with the gate metal in place. The capacitance of the device irradiated without the contacts in place recovers to the same value as the contact diodes. The $D_{it}$ values are decreased by the $H_2$ anneal in both the unirradiated and irradiated devices.

Figure 15.33 shows the current density (J)-voltage (top) and capacitance-voltage (C-V) characteristics (bottom) from the MgO/GaN devices with 500 Å oxide thickness before and after proton irradiation. We also irradiated devices both before and after deposition of their gate contacts in order to investigate the effect of the presence of the metal over the MgO. While we could not measure any significant change in reverse breakdown voltage as a result of the proton irradiation, the $V_F$ values were decreased in both types of devices (i.e., those irradiated before or after deposition of the gate metal). The $D_{it}$ values extracted from the C-V characteristics using the Terman method (which is generally found to underestimate the trap density) were also essentially unchanged in the devices from which we could extract accurate data. Note that the forward breakdown field decreases from ~$10^6$ V/cm in the control samples to ~$0.76 \times 10^6$ V/cm for the devices irradiated with contacts in place, and for $0.46 \times 10^6$ V/cm for the structures irradiated prior to deposition of the contacts. This result indicates that the protons make displacement damage in the MgO which degrade its breakdown capabilities.

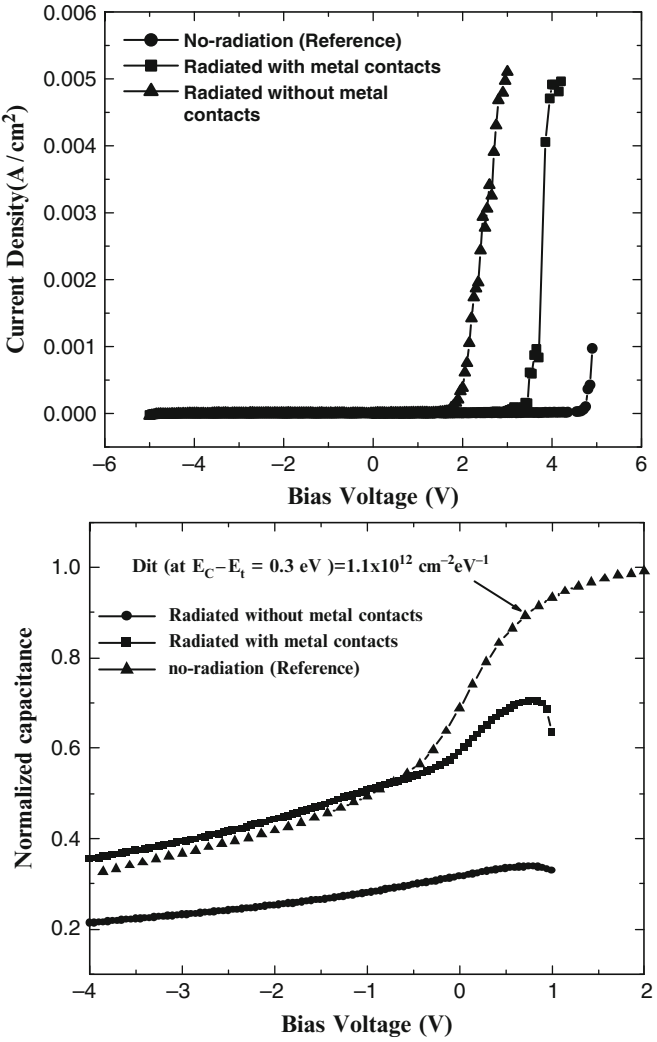



**Fig. 15.33** Current density-voltage (*top*) and capacitance-voltage characteristics (*bottom*) from MgO/GaN MOS devices before and after irradiation

## 15.6 Conclusions

The isolation leakage currents in mesa-isolated AlGaN/GaN HEMTs is found to be dominated by surface states whose density is a strong function of the cleaning and passivation films employed. UV-ozone treatment prior to the gate metal deposition effectively removes the surface contamination. The isolation current is always higher after $SiN_X$ deposition, while MBE-deposited $Sc_2O_3$ produces the lowest values for any of the films that reduce the current collapse problem in the HEMTs.

The use of $Sc_2O_3$ and MgO as a surface passivation and gate oxide on AlGaN/GaN HEMTs produces lower leakage currents, higher total drain current, increased reverse breakdown voltage, increased three-terminal breakdown voltage and larger output power, and power-added efficiency. The MOSHEMTs show experimental threshold voltages in good agreement with theoretical values, as expected from the lower surface state densities of the $Sc_2O_3$/GaN and MgO/GaN system. However, MgO was found to be sensitive to environmental and thermal conditions which impose limits on how the devices must be processed and packaged. The environmental degradation is believed to be due to interaction with water vapor in the air requiring capping of the MgO.

These passivation films play a role in determining the ultimate reliability of GaN devices by reducing the concentration of interface traps that lead to high leakage currents, self-heating, and device instabilities.

MgO/GaN MOS devices show a relatively high resistance to degradation from high energy proton exposure equivalent to approximately 10 years in low earth orbit. The proton irradiation causes a decrease in channel capacitance through the creation of deep electron traps, and a decrease of ~24% in forward breakdown field was observed for devices irradiated with the gate contact in place. The interface trap densities were essentially unchanged as a result of the proton irradiation. The MgO/GaN system appears to be promising for applications requiring radiation hard electronics.

**Acknowledgment** The work at UF was partially supported by an AFOSR MURI (G.Jessen).

## References

1. N.X. Nguyen, M. Micovic, W.-S. Wong, P. Hashimoto, L.-M. McCray, P. Janke, C. Nguyen, Electron. Lett. **36**, 468 (2000)
2. Y.-F. Wu, B.P. Keller, S. Keller, D. Kapolnek, P. Kozodoy, S.P. DenBaars, U.K. Mishra, Solid-State Electron. **41**, 1569 (1997)
3. M.A. Khan, X. Hu, G. Simin, A. Lunev, J. Yang, R. Gaska, M.S. Shur, IEEE Electron. Device. Lett. **21**, 63 (2000)
4. B.M. Green, K.K. Chu, E.M. Chumbes, J.A. Smart, J.R. Shealy, L.F. Eastman, IEEE Electron. Device. Lett. **21**, 268 (2000)
5. S.C. Binari, W. Kruppa, H.B. Dietrich, G. Kelner, A.E. Wickenden, J.A. Freitas Jr., Solid-State Electron. **41**, 1549 (1997)
6. K.K. Chu, J.A. Smart, J.R. Shealy, L.F. Eastman, *Proceedings of the State-of-The-Art Program on Compound Semiconductors (SOTAPOCS XX/X)* (Electrochemical Society, Pennington, 1998), pp. M.3–12
7. L.F. Eastman, V. Tilak, J. Smart, B.M. Green, E.M. Chumbes, R. Dimitrov, H. Kim, O.S. Ambacher, N. Weimann, T. Prunty, M. Murphy, W.J. Schaff, J.R. Shealy, IEEE Trans. Electron Devices **48**, 479 (2001)
8. Y.-F. Wu, D. Kapolnek, J.P. Ibbetson, P. Parikh, B.P. Keller, U.K. Mishra, IEEE Trans. Electron Devices **ED48**, 586 (2001)
9. J.W. Johnson, A.G. Baca, R.D. Briggs, R.J. Shul, C. Monier, F. Ren, S.J. Pearton, A.M. Dabiran, A.M. Wowchack, C.J. Polley, P.P. Chow, Solid-State Electron. **45**, 1979 (2001)

10. W. Lu, J. Yang, M.A. Khan, I. Adesida, IEEE Trans. Electron Devices. **ED48**, 581 (2001)
11. S.J. Pearton, J.C. Zolper, R.J. Shul, F. Ren, J. Appl. Phys. **86**, 1 (1999)
12. G. Simin, X. Hu, N. Ilinskaya, A. Kumar, A. Koudymov, J. Zhang, M. Asif Khan, R. Gaska, M. S. Shur, Electron. Lett. **36**, 2043 (2000)
13. E. Kohn, I. Daumiller, P. Schmid, N.X. Nguyen, C.N. Nguyen, Electron. Lett. **35**, 1022 (1999)
14. J.-S. Lee, A. Vescan, A. Wieszt, R. Dietrich, H. Leier, Y.-S. Kwon, Electron. Lett. **37**, 130 (2001)
15. X. Hu, A. Koudymov, G. Simin, J. Yang, M. Asif Khan, A. Tarakji, M.S. Shur, R. Gaska, Appl. Phys. Lett. **79**, 2832 (2001)
16. I. Daumiller, C. Kirchner, M. Kamp, K.J. Ebeling, E. Kohn, IEEE Electron. Device. Lett. **20**, 448 (1999)
17. A. Tarakji, G. Simin, N. Ilinskaya, X. Hu, A. Kumar, A. Koudymov, J. Yang, M. Asif Khan, M.S. Shur, R. Gaska, Appl. Phys. Lett. **78**, 2169 (2001)
18. E.M. Chumbes, J.A. Smart, T. Prunty, J.M. Shealy, IEEE Trans. Electron Devices **48**, 416 (2001)
19. B. Luo, J.W. Johnson, F. Ren, K.K. Allums, C.R. Abernathy, S.J. Pearton, R. Dwivedi, T.N. Fogarty, R. Wilkins, A.M. Dabiran, A.M. Wowchack, C.J. Polley, P.P. Chow, A.G. Baca, Appl. Phys. Lett. **79**, 2196 (2001)
20. S.J. Pearton, F. Ren, A.P. Zhang, K.P. Lee, Mat. Sci. Eng. Rep. **R30**, 55 (2000)
21. S.C. Binari, K. Ikossi, J.A. Roussos, W. Kruppa, D. Park, H.B. Dietrich, D.D. Koleske, A.E. Wickenden, R.L. Henry, IEEE Trans. Electron Devices **48**, 465 (2001)
22. G. Simin, A. Koudymov, A. Tarakji, X. Hu, J. Yang, M. Asif Khan, M.S. Shur, R. Gaska, Appl. Phys. Lett. **79**, 2651 (2001)
23. F. Ren, M. Hong, S.N.G. Chu, M.A. Marcus, M.J. Schurman, A. Baca, S.J. Pearton, C.R. Abernathy, Appl. Phys. Lett. **73**, 3893 (1998)
24. J.W. Johnson, B. Luo, F. Ren, B.P. Gila, V. Krishnamoorthy, C.R. Abernathy, S.J. Pearton, J.I. Chyi, T.E. Nee, C.M. Lee, C.C. Chuo, Appl. Phys. Lett. **77**, 3230 (2000)
25. B.P. Gila, J.W. Johnson, K.N. Lee, V. Krishnamoorthy, C.R. Abernathy, F. Ren, S.J. Pearton, ECS Proc. Vol. **2001–1**, 71 (2001)
26. M. Asif Khan, X. Hu, A. Tarakji, G. Simin, J. Yang, R. Gaska, M.S. Shur, Appl. Phys. Lett. **77**, 1339 (2001)
27. G. Simin, X. Hu, N. Ilinskaya, J. Zhang, A. Tarakji, A. Kumar, J. Yang, M. Asif Khan, R. Gaska, M.S. Shur, IEEE Electron Device. Lett. **22**, 53 (2001)
28. H. Morkoç, A. Di Carlo, R. Cingolani, Solid-State Electron. **46**, 157 (2002)
29. Y. Ohno, M. Kuzuhara, IEEE Trans. Electron Devices **ED48**, 517 (2001)
30. Y.-F. Wu, D. Kapolnek, J.P. Ibbetson, P. Parikh, B.P. Keller, U.K. Mishra, IEEE Trans. Electron Devices **48**, 586 (2001)
31. S. Arulkumaran, T. Egawa, H. Ishikawa, T. Jimbo, M. Umeno, Appl. Phys. Lett. **73**, 809 (1998)
32. I. Irokawa, Y. Nakano, Solid-State Electron. **46**, 1467 (2002)
33. T.S. Lay, M. Hong, J. Kwo, J.P. Mannaerts, W.H. Hung, D.J. Huang, Solid-State Electron. **45**, 1679 (2001)
34. T. Hashizume, E. Alekseev, D. Pavlidis, K.S. Boutros, J. Redwing, J. Appl. Phys. **88**, 1983 (2000)
35. B. Luo, J.W. Johnson, J. Kim, R.M. Mehandru, F. Ren, B.P. Gila, A.H. Onstine, C.R. Abernathy, S.J. Pearton, A.G. Baca, R.D. Briggs, R.J. Shul, C. Monier, J. Han, Appl. Phys. Lett. **80**, 1661 (2002)
36. B. Luo, R. Mehandru, J. Kim, F. Ren, B.P. Gila, A.H. Onstine, C.R. Abernathy, S.J. Pearton, R. Fitch, J. Gillespie, T. Jenkins, J. Sewell, D. Via, A. Crespo, Y. Irokawa, J. Electrochem. Soc. **149**, G613–619 (2002)
37. B.P. Gila, J. Johnson, R. Mehandra, B. Luo, A.H. Onstine, K.K. Allums, V. Krishamoorthy, S. Bates, C.R. Abernathy, F. Ren, S.J. Pearton, Phys. Status. Solidi **A188**, 239 (2001)
38. P.B. Klein, S.C. Binari, K. Ikossi, A.E. Wickenden, D.D. Koleske, R.L. Henry, Appl. Phys. Lett. **79**, 3527 (2001)
39. R. Mehandru, B.P. Gila, J. Kim, J.W. Johnson, K.P. Lee, B. Luo, A.H. Onstine, C.R. Abernathy, S.J. Pearton, F. Ren, Electrochem. Solid-State Lett. **5**, G51 (2002)

40. J. Kim, R. Mehandru, B. Luo, F. Ren, B.P. Gila, A.H. Onstine, C.R. Abernathy, S.J. Pearton, Y. Irokawa, Appl. Phys. Lett. **81**, 373 (2002)
41. J. Kim, R. Mehandru, B. Luo, F. Ren, B.P. Gila, A.H. Onstine, C.R. Abernathy, S.J. Pearton, Y. Irokawa, Appl. Phys. Lett. **80**, 4555 (2002)
42. J. Kim, B. Gila, R. Mehandru, J.W. Johnson, J.H. Shin, K.P. Lee, B. Luo, A. Onstine, C.R. Abernathy, S.J. Pearton, F. Ren, J. Electrochem. Soc. **149**, G482 (2002)
43. M. Asif Khan, Q. Chen, M.S. Shur, B.T. Dermott, J.A. Higgins, J. Burm, W.J. Schaff, L.F. Eastman, Solid-State Electron. **41**, 1555 (1997)
44. J.H. Edgar (ed.), *Properties of Group III Nitrides* (Inspec, the Institution of Electrical Engineers, London, 1994)
45. G.W. Kaye, T.H. Laby, *Tables of Physical and Chemical Constants and some Mathematical Functions*, 11th edn. (Longmans, New York, 1956)
46. C.T. Sah, *Fundamentals of Solid-State Electronics* (World Scientific, River Edge, 1991)
47. S.M. Sze, *Physics of Semiconductor Devices*, 2nd edn. (Wiley, New York, 1981)
48. W.M. Yim, E.J. Stofko, P.J. Zanzucchi, J.I. Pankove, M. Ettenburg, S.L. Gilbert, Epitaxially grown AlN and its optical band gap. J. Appl. Phys. **44**, 292 (1973)
49. W.L. Chin, T.L. Tansley, T. Osotchan, Electron mobilities in gallium, indium, and aluminum nitrides. J. Appl. Phys. **75**, 7365 (1994)
50. M.S. Shur, M.A. Khan, GaN/AlGaN heterostructure devices: photodetectors and field-effect transistors. Mater. Res. Bull. **22**(2), 44 (1997)
51. S.S. Derbeneva, S.S. Batsano, Dokl. Chem. **175**, 710 (1967)
52. S.S. Batsono, E.V. Dulepov, Sov. Phys. Solid State **7**(4), 995 (1965)
53. K.A. Gschneider, *Rare Earth Alloys* (Izd. Mir, Moscow, 1965)
54. V.N. Abramov, A.N. Ermoshkin, A.I. Kuznetsov, Sov. Phys. Solid State **25**, 981 (1983)
55. N. Daude, C. Jouanin, C. Gout, Phys. Rev. B **15**, 2399 (1977)
56. G.V. Samsonov, *The Oxide Handbook* (Plenum, New York, 1973)
57. N.A. Moser, J.K. Gillespie, G.D. Via, A. Crespo, M.J. Yannuzzi, G.H. Jessen, R.C. Fitch, B. Luo, F. Ren, B.P. Gila, Appl. Phys. Lett. **83**, 4178 (2003)
58. A.P. Zhang, L.B. Rowland, E.B. Kamensky, V. Tilak, J.C. Grande, J. Feetsov, A. Vatiakhikh, L.F. Eastman, J. Electron. Mater. **32**, 626 (2003)
59. W.A. Lanford, M.J. Rand, J. Appl. Phys. **49**, 2473 (1978)
60. H.J. Stein, V.A. Wells, R.E. Hampy, J. Electrochem. Soc. **126**, 1750 (1979)
61. R. Chow, W.A. Landoed, K.M. Wang, R. Rosler, J. Appl. Phys. **53**, 5630 (1982)
62. D.G. Esaev, V.M. Efimov, A.A. Shklyaew, Thin Solid Films **221**, 160 (1992)
63. G.V. Gadiyak, J. Appl. Phys. **82**, 5573 (1997)
64. S.T. Sah, Solid-State Electron. **33**, 147 (1990)
65. R.C. Sun, J.T. Clemens, J.T. Nelson, *Proceedings of the IEEE Reliability Physics* (IEEE, New York, 1980), p. 244
66. M.S. Shur, Solid-State Electron. **42**, 2131 (1998)
67. H. Morkoc, *Nitride Semiconductors and Devices* (Springer, Berlin, 1999)
68. B. Gaffey, L.J. Guido, X.W. Wang, T.P. Ma, IEEE Trans. Electron Devices **ED48**, 458 (2001)
69. N.-Q. Zhang, S. Keller, G. Parish, S. Heikman, S.P. DenBaars, U.K. Mishra, IEEE Electron Device. Lett. **21**, 421 (2000)
70. R. Therrien, G. Lucovsky, R.F. Davis, Phys. Status. Solidi. **A176**, 793 (1999)
71. P.B. Klein, J.A. Freitas Jr., S.C. Binari, A.E. Wickenden, Appl. Phys. Lett. **75**, 4016 (1999)
72. P.B. Klein, S.C. Binari, J.A. Freitas Jr., A.E. Wickenden, J. Appl. Phys. **88**, 2843 (2000)
73. P.B. Klein, S.C. Binari, K. Ikossi-Anastasiou, A.E. Wickenden, D.D. Koleske, P.L. Henry, D. S. Katzer, IEEE Electron. Lett. **37**, 661 (2001)
74. B. Luo, J.W. Johnson, B.P. Gila, A. Onstine, C.R. Abernathy, F. Ren, S.J. Pearton, A.G. Baca, A.M. Dabiran, A.M. Wowchack, P.P. Chow, Solid-State Electron. **46**, 467 (2002)
75. G. Simin, A. Tarakji, X. Hu, A. Koudymov, J. Yang, M.A. Khan, M.S. Shur, R. Gaska, Phys. Status. Solidi. A **188**, 219 (2001)

76. A. Tarakji, X. Hu, A. Koudymov, G. Simin, J. Yang, M.A. Khan, M.S. Shur, R. Gaska, Solid-State Electron. **46**, 1211 (2002)
77. A. Koudymov, X. Hu, K. Simin, G. Simin, M. Ali, J. Yang, M.A. Khan, IEEE Electron Device Lett. **23**, 449 (2002)
78. G. Simin, A. Koudymov, H. Fatima, J.P. Zhang, J. Yang, M.A. Khan, X. Hu, A. Tarakji, R. Gaska, M.S. Shur, IEEE Electron Device Lett. **23**, 458 (2002)
79. M. Asif Khan, X. Hu, A. Tarakji, G. Simin, J. Yang, R. Gaska, M.S. Shur, Appl. Phys. Lett. **77**, 1339 (2000)
80. X. Hu, A. Koudymov, G. Simin, J. Yang, M.A. Khan, A. Tarakji, M.S. Shur, R. Gaska, Appl. Phys. Lett. **79**, 2832 (2001)
81. F. Ren, M. Hong, S.N.G. Chu, M.A. Marcus, M.J. Schurman, A. Baca, S.J. Pearton, C.R. Abernathy, Appl. Phys. Lett. **73**, 3839 (1998)
82. H.C. Casey Jr., G.G. Fountain, R.G. Alley, B.P. Keller, S.P. DenBaars, Appl. Phys. Lett. **68**, 1850 (1996)
83. R. Therrien, G. Lucovsky, R.F. Davis, Phys. Stat. Solidi A. **176**, 793 (1999); Appl. Surf. Sci. **166**, 513 (2000).
84. J.W. Johnson, B.P. Gila, B. Luo, K.P. Lee, C.R. Abernathy, S.J. Pearton, J.I. Chyi, T.E. Nee, C. M. Lee, C.C. Chou, F. Ren, J. Electrochem. Soc. **146**, 4303 (2001)
85. P.A. Ivanov, M.E. Levinshtein, G. Simin, X. Hu, J. Yang, M.A. Khan, S.C. Rumyantsev, M.S. Shur, R. Gaska, Electron. Lett. **37**, 1479 (2001)
86. P. Chen, W. Wang, S.J. Chua, Y.D. Zheng, Appl. Phys. Lett. **79**, 3530 (2001)
87. D. Mistele, T. Rotter, R. Ferretti, F. Fedler, H. Klausiag, O.K. Semchinova, J. Stemmer, J. Aderhold, J. Graul, Mat. Res. Soc. Symp. Proc. **639**, G11.42.1 (2001)
88. B.P. Gila, J.W. Johnson, R. Mehandru, B. Luo, A.H. Onstine, K.K. Allums, V. Krishnamoorthy, S. Bates, C.R. Abernathy, F. Ren, S.J. Pearton, Phys. Status Solidi A **188**, 239 (2001)
89. M. Hong, K.A. Anselm, J. Kwo, H.M. Ng, J.N. Baillargeon, A.R. Kortan, J.P. Mannaerts, A.Y. Cho, C.M. Lee, J.I. Chyi, T.S. Lay, J. Vac. Sci. Technol., B. **18**, 1453 (2000)
90. Y. Nakano, T. Kachi, T. Jimbo, J. Vac. Sci. Technol. **B21**, 2220 (2003)
91. D.J. Fu, Y.H. Kwon, T.W. Kang, C.J. Park, K.H. Baek, H.Y. Cho, D.H. Shin, C.H. Lee, K.S. Chung, Appl. Phys. Lett. **80**, 446 (2002)
92. Y. Irokawa, J. Kim, F. Ren, K.H. Baik, B.P. Gila, C.R. Abernathy, S.J. Pearton, C. Pan, G.T. Chen, J.I. Chyi, Appl. Phys. Lett. **83**, 4987 (2003)
93. M. Shur, *Physics of Semiconductor Devices* (Prentice Hall, Englewood Cliffs, 1990)
94. H. Cho, K.P. Lee, B.P. Gila, C.R. Abernathy, S.J. Pearton, F. Ren, Solid-State Electron. **47**, 1597 (2003)
95. F. Ren, M. Hong, W.S. Hobson, J.M. Kuo, J.R. Lothian, J.P. Mannaerts, J. Kuo, S.N.G. Chu, Y.K. Chen, A.Y. Cho, *Tech. Dig. Int. Electron Devices Meeting*, 1996, p. 943
96. A. Vertiatchikh, L.F. Eastman, W.J. Schaff, I. Prunty, Electron. Lett. **38**, 388 (2002)

# Chapter 16
# Reliability Simulation

**M.E. Law, M. Griglione, E. Patrick, N. Rowsey, and D. Horton**

**Abstract** Reliability offers significant modeling challenges. Typically, parts cannot be tested until failure under normal operating conditions. Since the target is frequently a decade or longer of useful life, this is impractical. Consequently, accelerated testing is performed. This procedure only works when the physics is well understood, and the failure mechanism is not accelerated by factors not under the control of the testing. Consequently, modeling of the failure mechanism is crucial in making extrapolated predictions of lifetime. Technology Computer-Aided Design tools have advanced to the point where multiple physics can be included and the testing simulated fully. This chapter describes such an extended tool and provides examples of applying it to the understanding of two different failure mechanisms.

## 16.1 Introduction

Reliability offers significant modeling challenges. Typically, parts cannot be tested until failure under normal operating conditions. Since the target is frequently a decade or longer of useful life, this is impractical. Consequently, accelerated testing is performed. For example, testing to failure can be done under higher temperatures and lifetime activation can be extracted. The room temperature operation can then be extracted from the higher temperature and an Arrhenius relationship.

This technique is fine, if the controlling mechanism is simply temperature and the actual channel temperature is well known. GaN/AlGaN HEMT structures are frequently used as power amplifiers, and therefore, the channel temperature can be significantly higher than the ambient. This fact alone would throw off the

M.E. Law (✉) • M. Griglione • E. Patrick • N. Rowsey • D. Horton
Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA
e-mail: MLaw@eng.ufl.edu

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_16,
© Springer Science+Business Media New York 2013

extrapolation of lifetime. Further, failure is usually a combination of factors – like electric field (gate oxide degradation in Si MOS), current density (electromigration in back-end wiring), mechanical stressing (contact plug delamination), or complex interactions of these effects (gate sinking driven by field and mechanical stress). In these cases, the simple temperature extrapolation becomes irrelevant to the actual lifetime.

In these more complex cases, accurate modeling is vital to predicting the lifetime. This is where simulation can play a role. A reliability simulator should be able to simulate the conditions of the stress test, accurately reproduce results, and then use those physical insights to predict normal operation lifetime. This chapter describes an approach used to build such a simulator and has two operating examples.

## 16.2 FLOORS Overview

### *16.2.1 Finite Element Approach*

A finite element discretization method (FEM) conveniently fulfills the need for fully coupled, multiphysical modeling that is essential in reliability studies. Thus, FLOORS is based upon this approach. Electronic operation of semiconductor devices may depend on temperature and mechanical strain. Thus, the partial differential equations governing electronic operation, thermal transport, and elastic deformation will have various degrees of coupling within different simulations. Interaction between the three physical domains (electrical, thermal, and mechanical) is simplified by the use of the same basic functions in the discretization of each differential equation via the FEM.

The partial differential equations commonly used within the domains are given below.

$$\nabla^2 \Psi = -\frac{q}{\varepsilon}(p - n + N_{\mathrm{D}} - N_{\mathrm{A}}) \tag{16.1}$$

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \cdot J_n - \mathrm{r}_n + \mathrm{g}_n \quad \text{and} \quad \frac{\partial p}{\partial t} = \frac{1}{q}\nabla \cdot J_p - \mathrm{r}_p + \mathrm{g}_p \tag{16.2}$$

$$c\frac{\partial T}{\partial t} - \nabla \cdot K\nabla T = Q \tag{16.3}$$

$$\nabla \cdot \sigma = 0 \tag{16.4}$$

The first two equations (16.1) and (16.2), comprise the three-coupled partial differential equations commonly used for simulation of electronic operation in semiconductor devices. First is the Poisson equation, 16.1, where $\Psi$ is electrostatic

potential, $n$ is the density of electrons, $p$ is the hole density, $N_A$ and $N_D$ are acceptor and donor impurity densities, respectively, $q$ is charge, and $\varepsilon$ is permittivity. The electron and hole current continuity equations are given in 16.2, where $J_n$ and $J_p$ are the electron and hole current densities, respectively, and r and $g$ are recombination and generation terms. The thermal domain is characterized by the heat conduction equation given by 16.3, where $T$ is temperature, $c$ is the specific heat capacity, $K$ is thermal conductivity, and $Q$ is the heat generation term. The mechanical equilibrium equation, 16.4, governs elastic deformation, where stress $\sigma$ is related to strain $\varepsilon$ by the constitutive relation $\sigma = D(\varepsilon - \varepsilon_0)$. $D$, the stiffness matrix, is a function of Young's modulus and Poisson's ratio. Discretization of the equations within each domain has conventionally been accomplished with different methods.

Device simulation was made possible by the seminal work by Scharfetter and Gummel [1]. Finite difference or finite volume methods have been extensively used for discretization of the electrical partial differential equations for more than 40 years, e.g., (16.1) and (16.2) [2]. The Scharfetter–Gummel (SG) discretization method is applied to the continuity equations of Eq. 16.2 as a way to avoid instabilities otherwise found in a standard difference scheme [1]. Within this technique, the current density is formulated in terms of drift and diffusion components as shown in Eq. 16.5:

$$J_n = -q\mu n \nabla \Psi + qD\nabla n, \tag{16.5}$$

where $\mu$ is electron mobility and $D$ is electron diffusivity. A similar equation arises for hole current density. The SG method solves this equation one dimensionally assuming constant electric field. Thus, the current is defined along the grid edges with appropriate weighting factors. The SG method cannot rigorously be used with the FEM since it does not have an expression that continuously defines current in space. Some disadvantages of the SG method are that current is only defined on the edge of an element and thus evaluation of the total current over the element is not exact, plus current flow becomes dependent on the grid [2]. However, this method is very stable and reliable and remains the approach for many device simulators including PICES-II, PADRE, Sentaurus SDevice, and Silvaco Atlas.

Though using finite element methods to discretize the Poisson and continuity equations is less common, this approach is viable [3–5]. For formulation in a finite element-based discretization scheme, the continuity equations are written in terms of the electron and hole quasi-Fermi levels ($\phi_{n,p}$) as in Eq. 16.6:

$$J_n = -q\mu_n n \nabla \phi_n, \tag{16.6}$$

as compared to the drift-diffusion formula that is implemented with a finite volume SG scheme. Recent work has shown that simulation results obtained by the finite volume SG method or quasi-Fermi FEM are comparable and that the FEM approach is preferable for situations when current flow is not in the direction of the grid as in ionizing radiation in single event upsets [6]. Another advantage for

implementing with the FEM is that direct computation of a Joule heat generation term, $Q$, is possible. Based on a formulation for Joule heating (given in Eq. 16.7 [8])

$$Q = \frac{q|J_n|^2}{\mu_n n} \tag{16.7}$$

and the quasi-Fermi representation of current density (Eq. 16.6), the heat generation term can easily be computed with the quasi-Fermi solution variables. Thus, the Poisson and continuity equations of the electrical domain and the heat conduction equation of the thermal domain may easily be coupled to the mechanical equilibrium equations, which have long-standing roots in the FEM.

### *16.2.2 Alagator*

Alagator is the scripting language employed by FLOORS to describe the physical equations to solve. The equations are stored as character strings in a parameter database that is initialized on startup with appropriate defaults. Each solution variable/material combination can have a different differential equation. Interfaces between materials are treated as a different material, so interface transport can be handled. A matrix is then constructed of rows of solution variables (electrostatic potential, holes, elastic stress) and columns of materials (GaN, AlGaN, interface between GaN and AlGaN). This allows different approximations in different regions for different solution variables.

When execution of a device condition is requested, the strings are parsed into a partial differential equation class. The class is made of temporal terms (d/d$t$), spatial operations (grad, field), and area integrations (recombination, generation, reaction). Each of these types of term can include prefactors (diffusivity, mobility) as well as an argument to the operator (concentration, potential). The prefactors can be terms that are evaluated node by node or are things dependent on a derivative (field-dependent mobility). An explicit sum of products of the terms is implied, so these terms can be separated for assembly of the equation. A class chart is in Fig. 16.1.

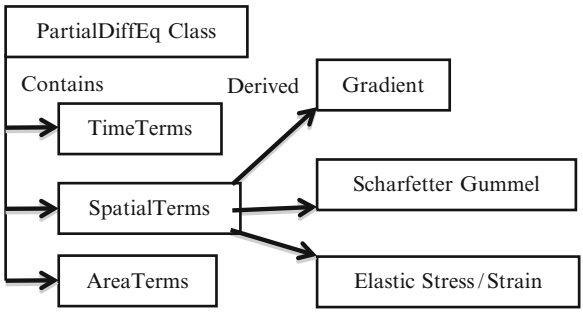



**Fig. 16.1** Class and container chart for the partial different equation expressions

The PartialDiffEq class contains one or more terms that have various derived pieces. Every portion of the term is parsed into a binary expression tree, with the usual numerical operators, simple functions, and logical expressions supported.

For assembly of the equations, both the value and the derivative of the expressions need to be computed for the Newton's method linearization. Each expression knows how to differentiate itself, and the chain rule is used liberally to construct the derivative expressions. For small simulations, the overhead of computing the derivatives can become a substantial fraction of the computation time. For most simulations, however, this computation is negligible.

The equations are then assembled using finite element techniques, as mentioned previously. The assembly is performed over 128 elements at a time. This allows most operations to work on 128 long vectors of values, which makes efficient use of modern pipelined architectures. To further save computation time, results are cached. Expressions that are evaluated multiple times are computed just once for each batch of 128 elements. This allows an automated efficiency in reusing results that requires no user intervention. Initial results on reasonably complicated problems (point-defect-mediated diffusion in two dimensions) show that this approach is comparable in CPU requirements as a hand-coded approach.

The great advantage of this is that physics can be adapted and improved quickly. In a research environment in which the dominant physics is not quite well understood, it allows the simulator to be adapted quickly and efficiently. Examples of adding complicated physics are contained in later sections.

### *16.2.3 Calibration Example: GaN HEMT*

Device simulations using the Florida object-oriented reliability simulator (FLOORS) were calibrated by comparing results with those from commercially available software, Synopsys Sentaurus, under identical device structure and electrical conditions. The GaN/AlGaN HEMT structure used in both simulators consisted of an undoped 1.0-μm SiC substrate, an undoped 0.20-μm AlN layer, a 1.775-μm GaN layer with an acceptor doping level of 6.5e16 $cm^{-3}$, and an undoped 0.025-μm $Al_{0.26}Ga_{0.74}N$ layer (Fig. 16.2). The structure was 4.0 μm wide. A 0.025-μm t-gate was contacted at the center of the surface of the AlGaN layer, and source and drain were contacted to the AlGaN at the left and rightmost regions of the layer. An insulating oxide layer was created on top of the AlGaN layer that also surrounded the contacts. An interface charge of 1.17e13 was placed between the AlGaN and GaN layers to reflect the polarization of the crystal due to the material offset at the interface.

A gate voltage, $V_G$, was applied in 1-V steps from 0 to −4 V, while the drain voltage, $V_D$, was swept from 0 to 10 V at each gate voltage step. The initial device temperature of the simulations was 300 K. A wafer backside temperature of 300 K was also incorporated in the simulations. Contact resistance for both source and drain was set at 0.5 Ω.

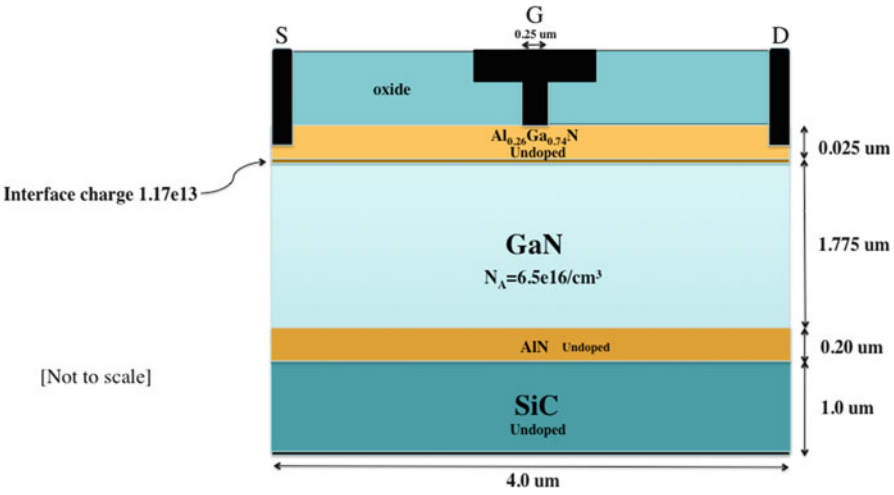



**Fig. 16.2** Cross-sectional diagram of the device structure used in FLOORS simulations

Development of a robust simulation model relied on incorporation of accurate expressions for mobility, electric field, thermal transport, and trap activity within the device. Bearing in mind that these simulations were focused on a p-type HEMT, with electrons flowing in the channel within the GaN layer, the equations for electrons and the physical parameters of GaN will be specifically emphasized here due to their overwhelming impact within the device.

Low and high field electron mobility expressions for GaN that identically matched those in the Sentaurus simulations were used in FLOORS:

$$\mu_{\text{low}} = \frac{1,630}{(T/300)^{1.88}} \tag{16.8}$$

$$\mu_{\text{high}} = \frac{\mu_{\text{low}}}{\left(1 + ((\mu_{\text{low}}{}^{*}E)/V_{\text{sat}})^{\beta}\right)^{\frac{1}{\beta}}} \tag{16.9}$$

with the values for the saturation velocity, $V_{\text{sat}}$, and $\beta$ shown in Table 16.1. The expression for the electric field, $E$, was the first derivative of the electrostatic potential $\Psi$.

The electron mobility was used in the thermal transport equation for GaN (eII. a.3), which accounts for changes in the device temperature profile as a result of carrier movement and lattice thermodynamics [8]. The heat capacity, $c$, and thermal conductivity, $k$, of GaN are given in Table 16.1. The final term of Eq. 16.3 is the heat generation term, $Q$, which represents self-heating within the device due to power dissipation and other carrier/lattice interactions. From the several models for $Q$ that have been proposed [9–11], FLOORS currently uses Eq. 16.7 for the heat

**Table 16.1** Expressions for GaN parameters used in thermal transport equations

| | |
|---|---|
| $V_{sat}$ ($10^7$ cm/s) | $3.3^{*}10^{-7} - ((3.0*10^6)*(T/300))$ |
| β | $0.85*(\ln(T/300))^{1.88}$ |
| $c$ (J/°K*cm$^3$) | $2.6725 + ((-4.25*10^{-3})*T) + ((3.0*10^{-6})*T^2)$ |
| $k$ (W/cm*°K) | $1.395 + ((5.14*10^{-3})*T) + ((-3.67*10^{-6})*T^2)$ |

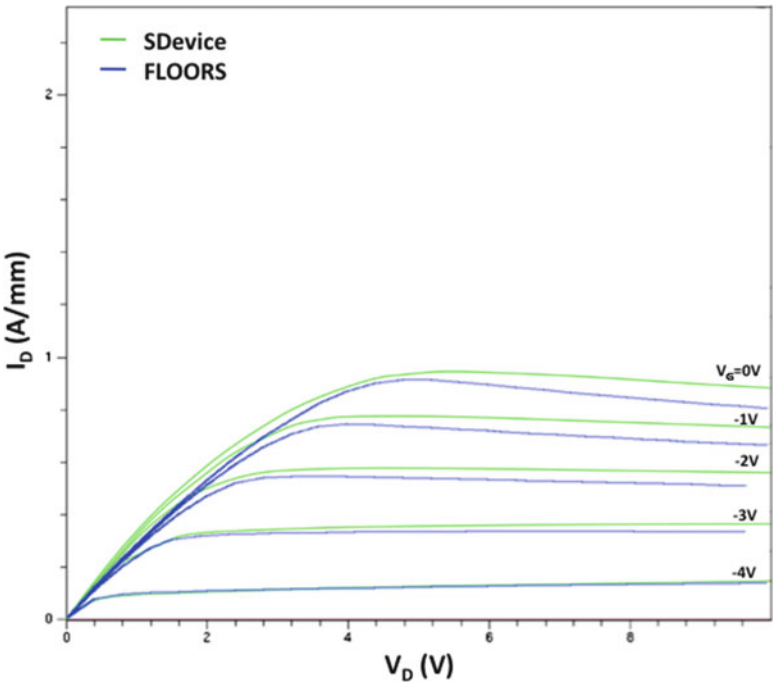



**Fig. 16.3** $I_D$ vs. $V_D$ for FLOORS (*blue*) and Sentaurus (*green*) for $V_G = 0$ to $-4$ V and $V_D = 0$–10 V

generation [12], which is simplified in the FLOORS to include only the n type due to negligible hole concentration for the simulated HEMT structure. The definition of $J_n$, the electron current density, is given in Eq. 16.6. Other parameters for the model are in Table 16.1.

Employing these thermoelectric models within FLOORS resulted in the *IV* curves shown in Fig. 16.3. The saturation voltages, $V_{Dsat}$, for the FLOORS and Sentaurus curves matched well for all $V_G$ values. The saturation voltages, $I_{Dsat}$, for both simulators were also almost identical for each gate voltage, with the largest discrepancies (~5%) occurring at the lowest $V_G$ values. A noticeable difference between the FLOORS and Sentaurus curves occurred at high $V_D$ values, where $I_D$ for FLOORS was lower than Sentaurus by approximately 0.10 A/mm (11%) at $V_D = 10$ V for $V_G = 0$ V. As $V_G$ decreased, the difference lessened, such that at $V_G = -4$ V, the curves were identical. The discrepancy at high $V_D$ can likely be

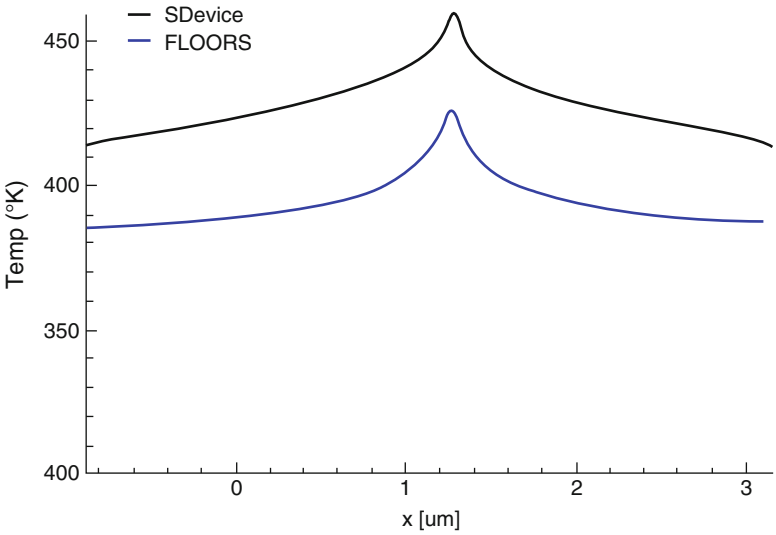



**Fig. 16.4** Temperature profiles of FLOORS (*blue*) and Sentaurus (*black*) device simulations at $V_G = 0$ V and $V_D = 10$ V in the channel at 0.26 μm depth

attributed to differences in the heat generation model (eII.c.6) used in each simulator, which can result in disparate thermal profiles in this region of device operation.

Another noticeable difference between the FLOORS and Sentaurus curves occurred at low $V_D$ values, where the initial slopes of the FLOORS curves at $V_D = 0$ to $-3$ V were slightly less than those of the Sentaurus curves for $V_G$ values ranging from 0 to $-2$ V. This can most likely be attributed to differences in the contacting metal blocks within the two simulators, which results in small differences in contact resistance calculations.

Comparison of the temperature profile at $V_G = 0$ V and $V_D = 10$ V for FLOORS and Sentaurus is shown in Fig. 16.4. The temperature in the FLOORS simulation was consistently ~30 K lower (~6.5%) throughout the entire width of the channel compared to the Sentaurus simulation, with the shape of the curves very similar. This small difference in temperature profiles supports the conclusion that the heat generation model contributed to a correspondingly small difference in IV simulation results.

### *16.2.4 Multiphysics: Electrical, Thermal, Mechanical*

Multiphysics modeling is needed for a variety of devices. For example, the modeling of MEMS thermal actuators requires physics in the electrical, thermal, and mechanical domains to simulate the effect of Joule heating on mechanical strain. Similarly, performance prediction of AlGaN/GaN HEMT devices

necessitates an electrothermomechanical (ETM) simulation due to their propensity for self-heating at normal operating temperatures and the piezoelectric nature of AlGaN and GaN.

Various commercial tools handle aspects of ETM modeling. COMSOL, Silvaco Atlas, and Sentaurus SDevice have fully coupled electrothermal solvers; however, they have limitations with mechanical simulation. In COMSOL, mechanical strain due to thermal expansion is a one-way couple. Atlas does not support mechanical simulations. SDevice does not support the inverse piezoelectric strain or temperature dependence on Young's modulus.

To date, researchers who need all three domains and a rigorous treatment of electronic properties in semiconductors use a combination of commercial software or write their own code. Venkatachalam et al. used a combination of Sentaurus for the electrical simulation and COMSOL for the thermal-mechanical simulation to model GaN-based HFETS [13]. Gao et al. used Silvaco Atlas and in-house software to simulate ETM effects on a GaN HEMT [14]. In both cases, the mechanics were one-way coupled to the electrical part. Thus, any strain-induced change on electrical performance could not be captured. One fully coupled finite element-based simulation couples electronic properties, strain, and transport in nanostructured devices [15]. However, thermal effects are not considered.

Recent work has shown that a mechanical-to-electrical couple is needed for the reliability assessment of AlGaN/GaN HEMTs. Horton et al. show that mechanical strain in the device enhances impurity diffusion on the drain side of the gate in the off state, which will ultimately affect *I–V* characteristics [16].

2-D simulation results from FLOORS are provided on examples of a MEMS thermal actuator and an AlGaN/GaN HEMT in the subsequent text. Equations 16.1, 16.2, 16.3, and 16.4 are implemented within the simulations. Mechanical strain caused by a mismatch of the thermal coefficient of expansion between two materials is simulated in both examples by a definition for initial stress given in Eq. 16.10:

$$\varepsilon_0 = \alpha_{\text{mismatch}} \Delta T, \tag{16.10}$$

where $\alpha_{\text{mismatch}}$ is the difference in the coefficients of thermal expansion. Strain due to the inverse piezoelectric effect in the HEMT example is modeled by Eq. 16.11:

$$\varepsilon_0 = \nabla \Psi \cdot d_{\text{pz}}, \tag{16.11}$$

where $d_{\text{pz}}$ is the matrix of inverse piezoelectric coefficients.

Figure 16.5 shows the mechanical deformation caused by a 3-V bias on the resistive heating element inside a bimorph beam. Dirichlet boundary conditions for the displacement and temperature were assigned to the bottom of the structure (bulk Si). Figures 16.6 and 16.7 compare the displacement and temperature results from FLOORS and COMSOL along the top of the beam. At most, there is a 2% difference in the simulation results. This demonstrates that the basic piezoelectric behavior is captured in FLOORS and that this can be used in more complex structures and devices with confidence.

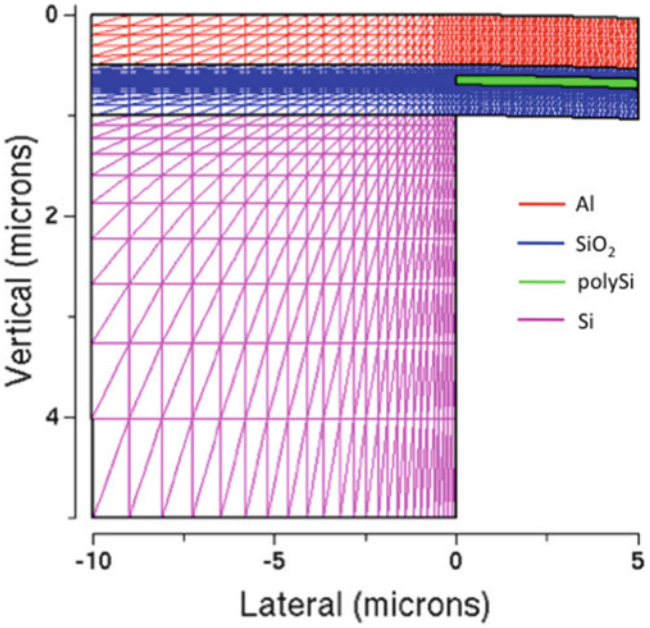



**Fig. 16.5** FLOORS electrothermomechanical simulation of a 2-D MEMS thermal actuator showing the resulting beam deformation with 3-V bias across the polysilicon heater

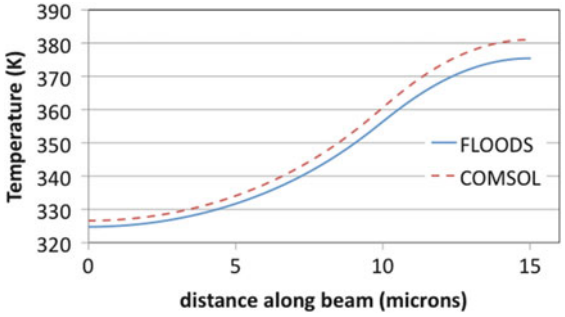



**Fig. 16.6** Electrothermomechanical simulation validation on a 2-D MEMS thermal actuator showing the temperature at the top of the beam. The FLOORS results show at most a difference of 2% to COMSOL results

Using the same device as shown in Fig. 16.2, the influence of the inverse piezoelectric effect and self-heating can be seen in the plot of vertical strain shown in Fig. 16.8. The compressive strain peaks at the drain edge of the gate and is in good qualitative agreement with simulation results from Gao et al. [14]. For the HEMT simulation, the Young's modulus was made dependent on temperature. Thus at elevated lattice temperatures, the decreased Young's modulus and addition of tensile strain due to the thermal expansion mismatch act to decrease the

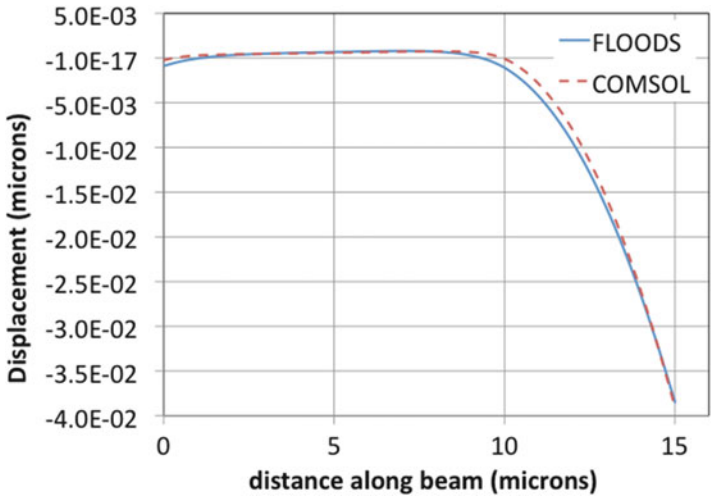



**Fig. 16.7** Electrothermomechanical simulation validation on a 2-D MEMS thermal actuator showing vertical displacement at the top of the beam. The FLOORS results are compared to COMSOL results

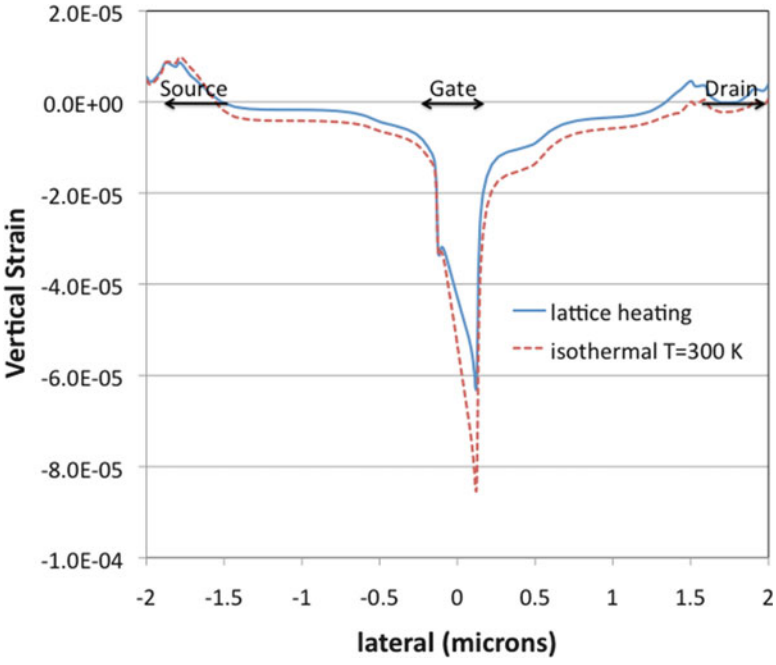



**Fig. 16.8** Effect of lattice temperature on the vertical strain along the AlGaN layer of a HEMT simulated in FLOORS. The full electrothermomechanical model shows a decrease in compressive strain from the isothermal case, relating relaxation of the lattice. $V_{\mathrm{G}} = 0.0$ V. $V_{\mathrm{D}} = 3.0$ V

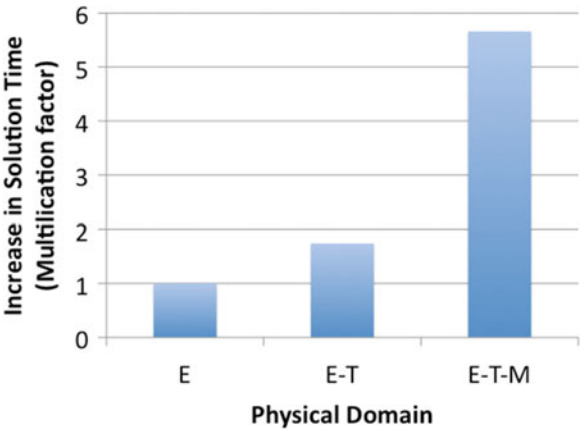



**Fig. 16.9** Comparison of solution time for the electrical-only, electrothermal, and electrothermomechanical simulations in FLOORS

compressive strain as shown in Fig. 16.8. Figure 16.9 compares the increase in simulation time with increase in domains. It takes about twice as long to run an electrothermal simulation than just an electrical simulation and about six times as long to complete an electrothermomechanical simulation.

## 16.3 FLOORS Example: Oxide Degradation Under Radiation

### *16.3.1 Introduction*

In device modeling, $SiO_2$, or any oxide, is usually treated as an ideal insulator – no mobile or fixed charges exist in the oxide, and the electric field or electrostatic potential drop across the inside of the oxide, as determined by Poisson's equation with charge = 0, is constant. The electrostatic potentials on either side of the insulator, i.e., in the gate and semiconductor, are the boundary conditions for full solution of the problem.

One-way oxide degradation occurs when charge is generated in or injected into the oxide, or onto the oxide/semiconductor interface [17]. Ionizing radiation is one such environmental phenomenon that is energetic enough to create charge, in the form of electron–hole pairs, across even the 9-eV bandgap of $SiO_2$. These electrons and holes are known to interact with oxygen vacancy complexes and hydrogenous species present in the oxide, eventually leading to fixed charge centers (deep hole traps) [18, 19] and interface trap buildup [20, 21]. To numerically account for this charge transport and defect interactions, $SiO_2$ can no longer be treated as a homogeneous insulator. Instead, $SiO_2$ is treated as a wide-bandgap semiconductor;

drift-diffusion equations are specified in the oxide, using appropriate values of diffusivity and mobility for each mobile particle being simulated in the oxide. The concentrations of preexisting neutral defects, such as oxygen vacancies ($V_o$) and hydrogenated oxygen vacancies ($V_oH$), will probably not be known and will have to be estimated to fit resultant changes in device characteristics. Similarly, the concentration of molecular hydrogen ($H_2$) may have to be estimated.

### 16.3.2 The Problem

Consider an MOS capacitor that has been irradiated by Co-60 gamma rays to a total dose of 30 krad and at a dose rate of 26 rad($SiO_2$)/s [22]. Experimental evidence suggests that many of the holes that are generated by the radiation are captured in "deep traps," close to the interface, and that protons, liberated from the oxide as a result of irradiation, are responsible for interface trap buildup [23, 24]. Theoretical, quantum mechanical, first principles calculations on the molecular scale suggest that reactions 12 and 13 are feasible in thermally grown $SiO_2$ at room temperature [25]:

$$V_o + h^+ \leftrightarrow V_o^+ \quad E_{f0} = 0.0\,\text{eV}, \quad E_{r0} = 4.5\,\text{eV}. \tag{16.12}$$

$$V_o^+ + H_2 \leftrightarrow V_oH + H^+ \quad E_{f1} = 0.4\,\text{eV}, \quad E_{r1} = 0.8\,\text{eV}. \tag{16.13}$$

First (16.12), a hole ($h^+$) can be captured on an oxygen vacancy ($V_o$) with a 0.0-eV energy barrier ($E_{f0}$). This hole is in a deep trap, since 4.5 eV ($E_{r0}$) is halfway into the $SiO_2$ bandgap. Holes in this trap will probably not be able to get enough thermal energy ($k_BT$) at room temperature to escape. Once the oxygen vacancy has captured a hole, it may "crack" molecular hydrogen ($H_2$), which can be left over in $SiO_2$ interstitial regions from processing steps involving $H_2$, forming a hydrogenated oxygen vacancy ($V_oH$) and a free proton ($H^+$) in the process.

First principles calculations also suggest that hydrogenated oxygen vacancies, formed during processing steps, prior to irradiation, also capture holes with a 0.0-eV barrier (16.14) and that once these precursors have captured a hole, protons may be released directly:

$$V_oH + h^+ \leftrightarrow V_oH^+ \quad E_{f2} = 0.0\,\text{eV}, \quad E_{r2} = 4.5\,\text{eV}, \tag{16.14}$$

$$V_oH^+ \leftrightarrow V_o + H^+ \quad E_{f3} = 0.4\,\text{eV}, \quad E_{r3} = 0.8\,\text{eV}. \tag{16.15}$$

It is possible to test whether these mechanisms can quantitatively explain the effects observed by experiment using FLOORS. The next section describes how.

### *16.3.3 FLOORS Implementation*

#### 16.3.3.1 Poisson and Continuity Equations

In silicon, the solution variables are electrostatic potential, electrons, and holes. The ionized dopant atoms in silicon are fixed point charges and as such must be included in Poisson's equation. However, the ionized dopant concentrations are known and do not change in concentration, and so these point defects will not have a continuity, or `ddt()` equation. Electrons and holes are charged and mobile, and so both drift and diffusion must be accounted for. In FLOORS, the `sgrad()` operator accounts for both drift and diffusion together.

Under irradiation, a nonequilibrium condition, electron–hole pairs will be generated according to a radiation generation term [26]:

$$U_{\text{radiation}} = Y_{\text{f}} R_{\text{d}}, \tag{16.16}$$

where $Y_{\text{f}}$ is the yield fraction of electron–hole pairs in $SiO_2$ that survives geminate recombination [27], an initial recombination step that is not modeled in FLOORS, and $R_{\text{d}}$ is the dose rate of the radiation. Thus, the equation set defined in silicon is

$$\nabla \cdot (\varepsilon \nabla \Psi) = N_{\text{d}}^{+} - N_{\text{a}}^{-} + n - p \tag{16.17}$$

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{1}{q} \nabla \cdot \vec{J}_n - U_{\text{radiation}} \tag{16.18}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{1}{q} \nabla \cdot \vec{J}_p - U_{\text{radiation}} \tag{16.19}$$

where $\Psi$ is electrostatic potential, and $N_{\text{d}}^{+}$, $N_{\text{a}}^{-}$, $n$, and $p$ are the concentrations of ionized donors, ionized acceptors, electrons, and holes. The FLOORS implementation is

```
pdbSetString Silicon DevPsi Equation "eps*grad(DevPsi) +
  Ndminus - Naplus - Elec+Hole"
pdbSetString Silicon Elec Equation "ddt(Elec) - MOB*sgrad
  (Elec, Vt) + Uradiation_Si"
pdbSetString Silicon Hole Equation "ddt(Hole) + MOB*sgrad
  (Hole, Vt) + Uradiation_Si"
```

In $SiO_2$, the solution variables are electrostatic potential, electrons, holes, protons, molecular hydrogen, uncharged oxygen vacancies ($V_o^{+}$ and $V_oH^{+}$, which are uncharged point defects), and charged oxygen vacancies ($V_o^{+}$ and $V_oH^{+}$, which are charged point defects). All the charged species must be included in Poisson's equation:

$$\nabla \cdot (\varepsilon \nabla \Psi) = N_{\text{d}}^{+} - N_{\text{a}}^{-} + p - n + \text{H}^{+} + \text{V}_{\text{o}}^{+} + \text{V}_{\text{o}}\text{H}^{+} \tag{16.20}$$

and the FLOORS implementation is

```
pdbSetString Oxide DevPsi Equation "eps*grad(DevPsi)-
  Hole + Elec-Hplus-Voplus-VoHplus"
```

Any species changing concentration with time must have its own continuity equation. The creation and consumption, or recombination/generation, of species due to the chemical reactions above must be included in the continuity equations. The formulation of these terms will be discussed later and for now will be represented under the term $U_{\text{net_reactions}}$. Species which are both charged and mobile undergo both drift and diffusion transport. For example, protons will have a continuity equation:

$$\frac{\mathrm{dH}^+}{\mathrm{d}t} = \frac{1}{q}\nabla \cdot J_{\mathrm{H}^+} - U_{\text{net_reactions}} \tag{16.21}$$

and will use the `sgrad()` operator:

```
pdbSetString Oxide Elec Equation "ddt(Elec) - MOB*sgrad
  (Elec, +DevPsi/Vt)+Urad+UnetR_Elec"
pdbSetString Oxide Hole Equation "ddt(Hole)+MOB*sgrad
  (Hole, -DevPsi/Vt)+Urad+UnetR_Hole"
pdbSetString Oxide Hplus Equation "ddt(Hplus)+MOB*sgrad
  (Hplus, -DevPsi/Vt)+UnetR_Hplus"
```

Species which are mobile, but not charged, have only diffusion in their continuity equation, which is accounted for in FLOORS by the `grad()` operator:

```
pdbSetString Oxide H2 Equation "ddt(H2)+D*grad(H2)+
  U_net_reactions_H2"
```

Finally, point defects are not mobile, so have neither drift nor diffusion, but still have net recombination/generation due to the reactions:

```
pdbSetString Oxide Vo Equation "ddt(Vo)-U_net_
  reactions_Vo"
pdbSetString Oxide VoH Equation "ddt(VoH)-U_net_
  reactions_VoH"
pdbSetString Oxide Voplus Equation " ddt(Voplus)-U_net_
  reactions_Voplus"
pdbSetString Oxide VoHplus Equation "ddt(VoHplus)-U_net_
  reactions_VoHplus"
```

Inputting equations by hand for more than three species gets very cumbersome and can lead to careless errors. Instead, we use a Tcl procedure that will take the name of each variable, and some information about the variable, such as whether it is charged and mobile, or fixed and uncharged, to build a continuity equation for you. For example, the procedure `AddSpecies`

```
proc AddSpecies {Name Mat Const Val Charge} {
   set eqn "ddt($Name)"
   if {$Const == "D"} {
      append eqn "-$Val*grad($Name)"
      pdbSetBoolean $Mat $Name mobile_yn 1
   }
   if {$Const == "MOB"} {
      if {$Charge == "-"} {append eqn "-$Val*Vt*sgrad
         (($Name), +DevPsi/Vt)"}
      if {$Charge == "+"} {append eqn "-$Val*Vt*sgrad
         (($Name), -DevPsi/Vt)"}
      pdbSetBoolean $Mat $Name mobile_yn 1
      }
    #Add dharged terms to Poisson Equation
    if {$Charge !="0"} {
        if {[pdbIsAvailable $Mat DevPsi Equation]} {
           set eqnP [pdbGetString $Mat DevPsi Equation]
           set eqnP [append eqnP " $Charge $Name"]
           pdbSetString $Mat DevPsi Equation $eqnP
           puts "Amending Poisson Eqn in $Mat to: $eqnP"
      }
   }
   #Store the transport equation in the pdb
   puts "Adding $Name Equation in $Mat: $eqn"
   pdbSetString $Mat $Name Equation $eqn
   #for understanding the type of species for the k-calc
   pdbSetString $Mat $Name Const $Const;#probably don't
   need this
   pdbSetString $Mat $Name Charge $Charge
}
```

takes the species name, its transport type (MOB for drift-diffusion, D for diffusion, or 0 for fixed), the value of that constant (mobility in $cm^2/V/s$, diffusion in $cm^2/s$, or 0 for fixed), and its charge type (+/−/0) and builds the appropriate continuity equation based on this information and also adds any charged species to Poisson's equation. In this way, you can add many species, easily keeping track of them, their equations, and their parameters:

```
#Example:
#Procedure   Name   Mat      Const   Val    Charge
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
AddSpecies   H2     Oxide    D       1.7e-15    0
```

#### 16.3.3.2 Generation and Recombination Due to Chemical Reactions

Each $U_{\text{net_reactions}}$ term is built based on energy balance equations [28]. Consider Eqs. 16.12 and 16.13 under equilibrium. From the law of mass action, the concentrations of all the species are related in equilibrium via

$$k_{f0} \cdot [\mathrm{V_o}] \cdot [\mathrm{h^+}] = k_{r0} \cdot \left[\mathrm{V_o^+}\right] \tag{16.22}$$

$$k_{f0} \cdot \left[\mathrm{V_o^+}\right] \cdot [\mathrm{H_2}] = k_{r1} \cdot [\mathrm{V_oH}] \cdot [\mathrm{H^+}] \tag{16.23}$$

where the bracket indicates concentration of the species and $k_f$ and $k_r$ are reaction rates, related by the reaction's equilibrium constant. Equations (16.22) and (16.23) above are true in equilibrium. For nonequilibrium conditions, there will be either net recombination or generation of each species involved in a reaction. For each reaction, a net recombination term can be assembled. This term is then added to continuity equation for species that are being generated and subtracted from continuity equations for each species experiencing recombination. For example, in Eq. 16.12, $\mathrm{V_o^+}$ is being generated according to

$$G_{\mathrm{V_o^+,R_0}} = k_{f0} \cdot [\mathrm{V_o}] \cdot [\mathrm{h^+}] - k_{r0} \cdot \left[\mathrm{V_o^+}\right] \tag{16.24}$$

and $\mathrm{V_o}$ and $\mathrm{h^+}$ are recombining according to

$$G_{\mathrm{h^+},R_0} = k_{r0} \cdot \left[\mathrm{V_o^+}\right] - k_{f0} \cdot [\mathrm{V_o}] \cdot [\mathrm{h^+}] \tag{16.25}$$

Since these terms are the opposite of each other, we can simply create one recombination/generation term for each reaction:

$$\mathrm{GR_0} = k_{f0} \cdot [\mathrm{V_o}] \cdot [\mathrm{h^+}] - k_{r0} \cdot \left[\mathrm{V_o^+}\right] \tag{16.26}$$

which must then be subtracted from the continuity equation for each species on the left hand side of the reaction and added to the continuity equation for each species on the right hand side of the equation. In other words, $\mathrm{GR_0}$ must be subtracted from the continuity equations for $\mathrm{V_o}$ and $\mathrm{h^+}$ and added to the continuity equation for $\mathrm{V_o^+}$. We can follow the same procedure for 13:

$$\mathrm{GR_1} = k_{f1} \cdot \left[\mathrm{V_o^+}\right] \cdot [\mathrm{H_2}] - k_{r1} \cdot [\mathrm{V_oH}] \cdot [\mathrm{H^+}] \tag{16.27}$$

$\mathrm{GR_1}$ must be subtracted from the continuity equations for $\mathrm{V_o^+}$ and $\mathrm{H_2}$ and added to the continuity equations for $\mathrm{V_oH}$ and $\mathrm{H^+}$. This may seem simple with only two reactions to consider, but the process can get confusing and prone to error as more reactions are added. However, the process of adding the recombination/generation terms due to reactions for each appropriate continuity equation can be automated as well. [29] used an automated script that parsed reactions inputted as simple text,

built the RG terms, and then added or subtracted them from the appropriate continuity equations, allowing easy changing of reactions or reaction energetics and eventually 20 reactions all at once.

#### 16.3.3.3 Formulation of the Reaction Rates

Following [30], the reaction rates used above are calculated based on the energetics of 12 and 13 as calculated by first principles calculations. There are three cases of interest: (1) a mobile species combines with a static species, overcoming a reaction barrier that is larger than diffusion barrier; (2) a mobile species combines with a static species, but the reaction barrier is lower than the diffusion barrier; and (3) a mobile species escapes from a static species. The potential energy diagrams for the three cases are shown in Fig. 16.10.

The equations for each case are as follows:

$$(\text{case } 1) \quad k = L_C D \exp\left(-\frac{E_{\text{bar}} - E_{\text{d}}}{k_{\text{B}} T}\right) \tag{16.28}$$

$$(\text{case } 2) \quad k = 2 L_C D \tag{16.29}$$

$$(\text{case } 3) \quad k = f \exp\left(-\frac{E_i}{k_{\text{B}} T}\right) \tag{16.30}$$

$L_C$ is an estimated "critical length" or critical distance beyond which the mobile particle cannot "see" the trap, but less than which the mobile particle will definitely be captured by the trap. This distance is on the order of a few atomic spacings for uncharged interactions. If one or more of the particles is charged, then this distance can be estimated to be larger. This value can be estimated because, as $L_C$ affects $k$ on a

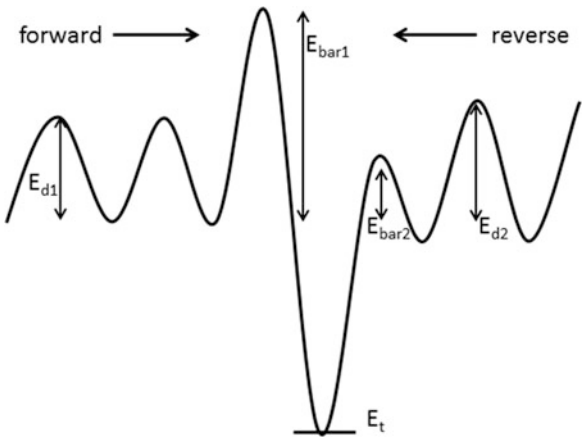



**Fig. 16.10** Energy diagram showing diffusion and reaction energies in the forward and reverse directions

linear scale, it is not nearly as important as $E$, which affects $k$ on an exponential scale. $D$ is the diffusivity of the mobile species. The Einstein relation is used to convert between mobility and diffusivity. $E_{bar}$ is the reaction barrier obtained from first principles calculations. $E_d$ is the diffusion energy of the mobile species. It is subtracted from $E_{bar}$ in order to avoid "double-counting" diffusion through both $D$ and the reaction kinetics. $E_i$ is the trap energy, or trap depth. $f$ is the "attempt-to-escape" frequency of the trapped particle. $k_B$ is Boltzmann's constant, and $T$ is temperature.

#### 16.3.3.4 Preliminary Solves: Poisson-Only Solution, DC elec/hole/DevPsi Only

Before a transient simulation can be carried out, following the experimental conditions of [22] and [31], the system is first solved in equilibrium (DC). Obtaining convergence of electrons and holes in the oxide can be a challenge due to the very small concentrations there. Solving Poisson's equation alone first, then using this solution as an initial guess for the full DC solve may be required. That is, first, solve only electrostatic potential as an independent variable, and then use the Boltzmann relation to directly calculate electrons and holes:

```
solution add name=DevPsi pde damp continuous
#using the Poisson equation as defined above for DevPsi
solution add Silicon name=Elec const val=(Nc*exp(-Ec-
0.0/kT))
solution add Silicon name=Hole const val=(Nv*exp(-0.0-
Ev/kT))
solution add Oxide name=Elec const val=(Nc*exp(-Ec-0.0/
kT))
solution add Oxide name=Hole const val=(Nv*exp(-0.0-Ev/
kT))
```

where Nc, Nv, Ec, and Ev are appropriately defined for each material. The 0.0 represents the value of the Fermi level when the silicon is biased at 0 V. If the 0 of potential is defined differently, or if the silicon is biased at a different value, then this value will have to be changed. A simple initial guess will suffice for a Poisson-only solution:

```
sel z=-4.1 name=DevPsi
device
```

The solutions in this simplified case are saved by FLOORS and can now serve as an initial guess for the full DC solution:

```
solution add name=DevPsi pde damp continuous
solution add name=Elec pde !negative
solution add name=Hole pde !negative
```

```
#using the Poisson and continuity equations as defined
above for DevPsi, Elec, and
Hole
device init
```

For the DC solution, no recombination/generation terms are included in the continuity equations.

### *16.3.4 Results and Modeling Capabilities*

Simulation can be very useful when it is not possible to decouple physical effects experimentally. Understanding the individual contributions of proposed mechanisms, and how results change due to their coupling, can lead to insights into the cause of experimentally observed effects such as oxide degradation.

For example, Fig. 16.11 shows how the two different proton production mechanisms discussed above contribute to interface trap creation in test devices

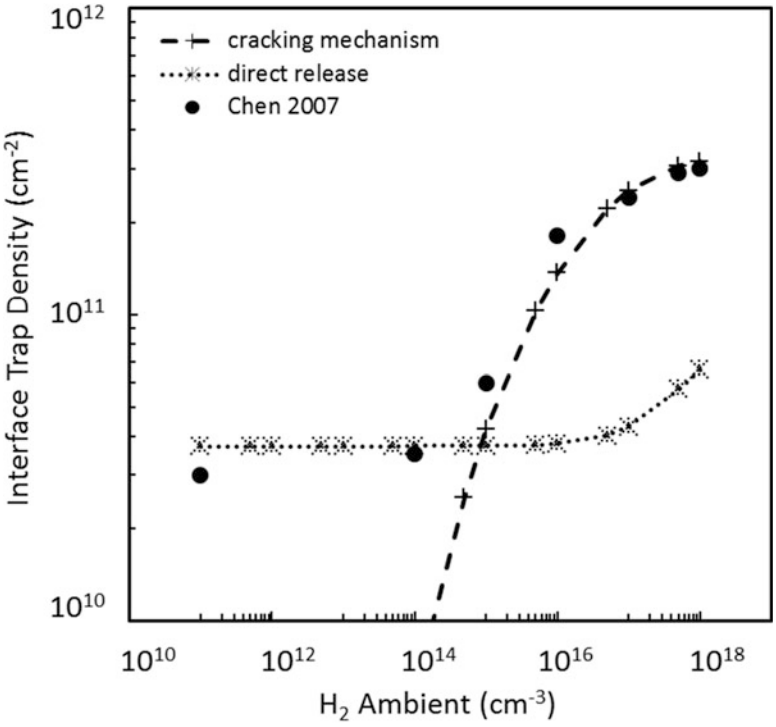



**Fig. 16.11** FLOORS simulations showing each mechanism simulated separately to show the individual contributions of the two mechanisms (The data are from [22]. Reprinted, with permission, from [29], Figure 3)

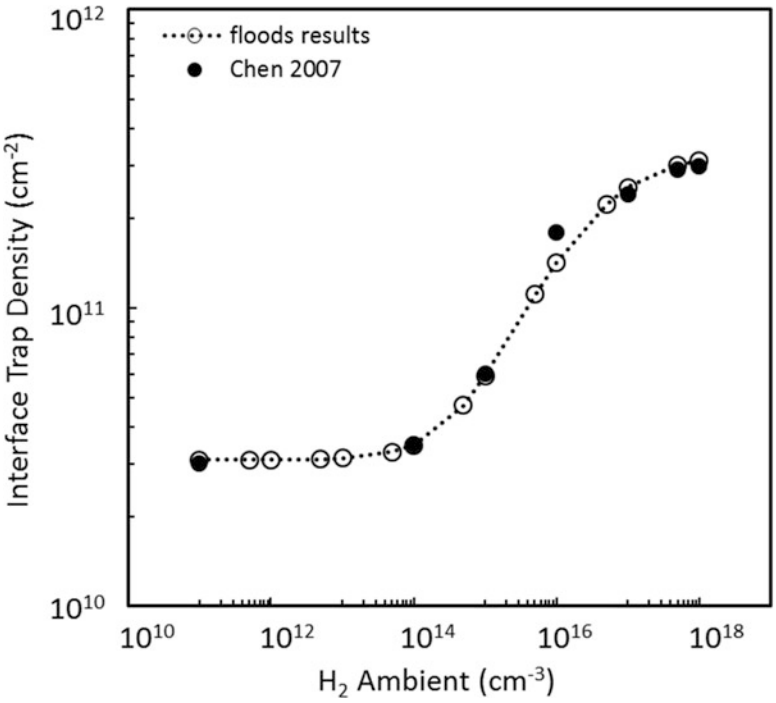



**Fig. 16.12** FLOORS results showing the two mechanisms simulated together (Results quantitatively match the data of [22] very well. Reprinted, with permission, from [29], Figure 1)

exposed to a wide range of ambient $H_2$ concentrations. When simulated alone, the direct release mechanism can account for interface trap buildup at low $H_2$ concentrations. The cracking reaction can account for the increased interface trap buildup at high $H_2$ concentrations but is not significant at low $H_2$ concentrations. Simulated together, these mechanisms account for the buildup of interface traps over a comprehensive range of hydrogen environments (Fig. 16.12). The concentration of preexisting hydrogenated oxygen vacancies controls the concentration of interface traps generated at low $H_2$ concentrations. The concentration of preexisting nonhydrogenated oxygen vacancies accounts for the concentration of interface traps generated at the high-$H_2$ plateau [29]. Both the barrier for proton release and the depth of each hole trap play a role in the concentration of interface traps generated [32].

First principles calculations, such as those used to obtain the forward and reverse reaction energies discussed above, contain inherent statistical error and also depend on assumptions made about the physical system. FLOORS can be used to narrow down this error and to test certain physical assumptions made. A sensitivity analysis can be performed, in which each parameter supplied by density functional theory is varied slightly, within the statistical error bars of the first principles calculation. The effect of this variation on the FLOORS simulation results, in this case, buildup of

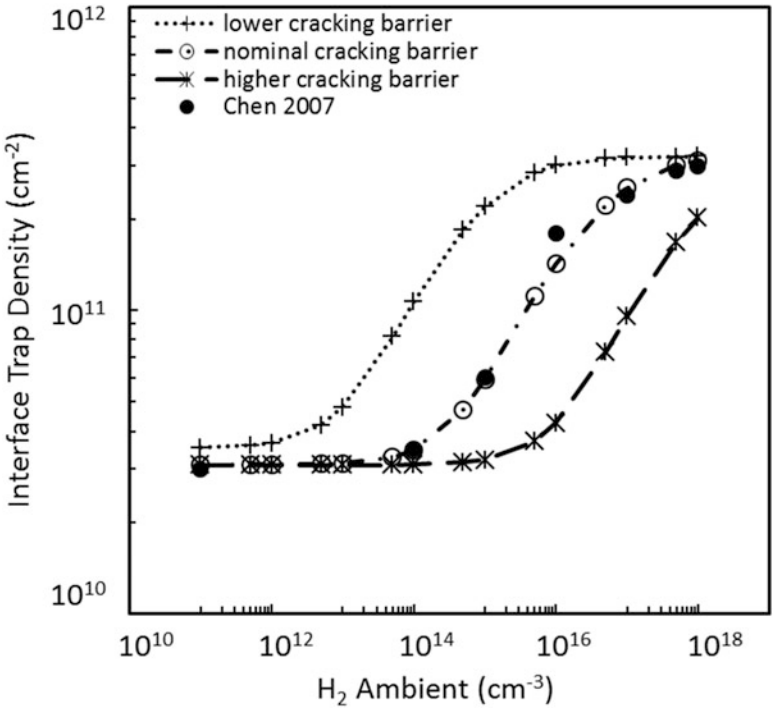



**Fig. 16.13** FLOORS results showing how the energy barrier to the cracking reaction controls the H2 concentration at which the shift from the direct release to the cracking mechanism occurs (The data are from [22]. Reprinted, with permission, from [29], Figure 9)

interface trap density, can be observed. In this way, it can be determined which parameters affect the results very strongly and which do not have any effect on the results, i.e., which parameters are sensitive and which are not.

For example, the energy barrier to $H_2$ cracking that was originally obtained from first principles calculations was 0.4 eV. This value produces the simulation results that make up the leftmost curve in Fig. 16.13, which does not match the data. The sensitivity analysis showed that raising this barrier to 0.5 eV, which is within the statistical error bounds of the first principles calculations, shifts the results to the right to match the data [32].

The model presented here for radiation-induced interface trap buildup, and the more complicated one investigated in [32], contain many parameters to vary, and a separate simulation is required for each point on the curves in Figs. 16.11 through 16.13. But this can be scripted in a tool such as FLOORS, allowing for systematic investigation of each parameter and mechanism and for efficient organization of results. Most parameters will be known within a range of uncertainty, either from first principles calculations or from experiment. Then, most of these parameters will not be sensitive within this range. For example, varying the energy in Fig. 16.13, only 0.1 eV affects the $H_2$ concentration at which the cracking mechanism takes

over from the direct release mechanism by over an order of magnitude. Other parameters had negligible effect on the simulation results. Using this method of combining first principles calculations with a sensitivity analysis in FLOORS, key parameters and key controlling mechanisms can be identified in degradation systems. The results presented here were for MOS regions of specialized test structures in which $H_2$ concentrations could be controlled and trap densities could be easily and reliably extracted. However, this method is applicable to similar systems in other materials and interface.

## 16.4 FLOORS Example: Gate Degradation and Diffusion Under Piezoelectric Strain

GaN high electron mobility transistors, although having impressive RF power densities, currently suffer observed and documented output degradation. This long-term reliability is currently being addressed, and a number of mechanisms have been proposed which attempt to theorize the problem. Diffusion of metal and other impurities from under the gate, enhanced by mechanical strain from the inverse piezoelectric effect, has been proposed as a mechanism for electrical degradation of AlGaN/GaN HEMTs in the off state [33]. Ni diffusion from Ni/Au gates has been observed in degraded areas of devices, and characteristic bulk diffusion models have been matched to increases in trap concentration over time. These observations have been linked to documented unrecoverable decreases in drain current and corresponding increases in gate current.

Recent electron energy-loss spectroscopy (EELS) line scans performed by [34] show clear signs of increased concentration of Ni and O in the AlGaN layer under the Au/Ni gate after off-state biasing conditions. These areas exhibit a distinctive diffusion-like shape as shown in Fig. 16.14. UV light-assisted trapping analysis has been employed by [33] to determine trap densities and activation energies. The amplitude, indicating density, of the dominant trap was measured as a function of time and temperature and yielded classic diffusion behavior [35]. This diffusion behavior was also observed at room temperature and indicates an additional potential driving the diffusion of these impurities other than temperature. This additional potential is the increasing mechanical strain in the AlGaN layer due to the inverse piezoelectric effect.

Off-state fields were computed with FLOORS, and the piezoelectric effect was incorporated to generate mechanical strain in a full elastic computation. The strain results were used to simulate the diffusion that would occur in a step-stress experiment. The altered structure was then used to sweep the *I–V* current to predict the corresponding drop in drain current, $I_{DS}$.

The mechanical strain due to the inverse piezoelectric effect, $\varepsilon_i$, is incorporated via a first-order expression relating strain and electric field. See Eq. (16.31) in Table 16.2. The corresponding inverse piezocoefficients for GaN [36] are listed

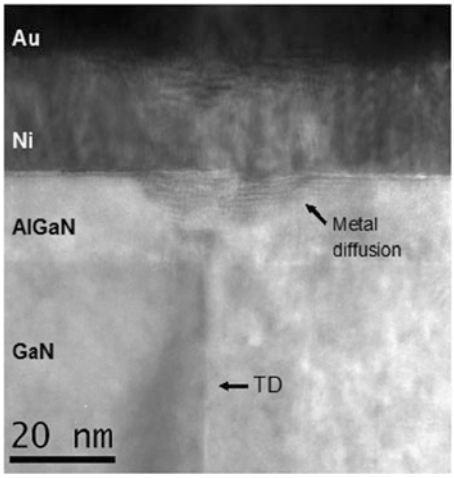



**Fig. 16.14** TEM image of the region with Ni and oxygen diffusion and the associated threading dislocation

**Table 16.2** Governing equations and constants used in simulation

| | |
|---|---|
| 16.31 | $\varepsilon_i = d_{ij} E_j$, where $j = 1, 2$, $i = 1, 2, 3$ in 2D formulation |
| 16.32 | $\frac{\mathrm{d}^2 \psi}{\mathrm{d}x^2} = -\frac{q}{\varepsilon_{\text{material}}}[-n(x) + p(x) + \text{Doping}(x)]$ |
| 16.33 | $D' = D \exp\left(\frac{Qs}{kT}\right)$ |
| 16.34 | $D' = D_0 \left[\exp\frac{Q * \text{misfit}}{kT} + \exp\frac{Q * \text{inv piezo strain}}{kT}\right]$ |
| $d_{33}$ | 3.4 pm/V |
| $d_{31}$ | –1.7 pm/V |
| $d_{15}$ | 3.1 pm/V |

The impurity profile is included in Eq. 16.2 as an additive term expressed as "+ *Impurity* (*x*)". The relative permittivity of GaN, $\varepsilon_{\text{GaN}}$, was taken as a static value of 9.0. This value is frequency dependent and tends to decrease at high frequencies such as the typical cutoff frequencies, $f_{\text{T}}$, of industrial HEMTs of ~40 GHz. Subsequent models will account for this. In Eq. 16.3, "$D'$" represents the modified diffusivity due to imposed strain. In all off-state simulations, temperature, $T$, was held at 300 K

in Table 16.2. The electric field $E_j$, given by $-(\mathrm{d}\psi/\mathrm{d}x)$, is derived from the spatially varying electrostatic potential $\psi$, found as a solution to Poisson's equation given by Eq. 32 in t.IV.2. For a given set of bias conditions and device dimensions, the electrostatic potential within the device is determined, and electric field is subsequently derived. The strain distribution within the structure is then determined as a function of bias. Figures 16.15 and 16.16 respectively show the strain distributions around the gate for an off-state bias condition of $V_{\text{DS}} = 30$ V, $V_{\text{G}} = -5$ V, and a $V_{\text{DS}} = 0$ state ($V_{\text{D}} = V_{\text{S}} = 30$ V) where $V_{\text{G}}$ was held at $-5$ V.

Modified diffusivity of impurity species as a function of strain [37] is given by Eq. (16.33) in Table 16.2 where $Q$ is the species-dependent energy per unit strain.

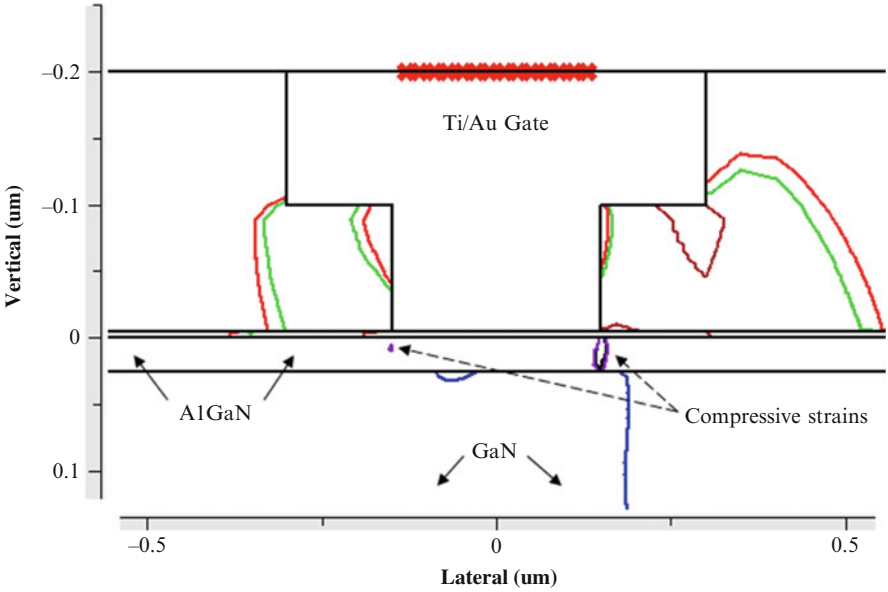



**Fig. 16.15** Simulation figure of GaN HEMT in the off state with $V_{GS} = -5$ V and $V_{DS} = 30$ V. Mechanical strain contours due to the inverse piezo effect show highest compressive strains of ~−7e−4 in the AlGaN layer at both edges under the gate but much more pronounced on the drain edge (*right side*)

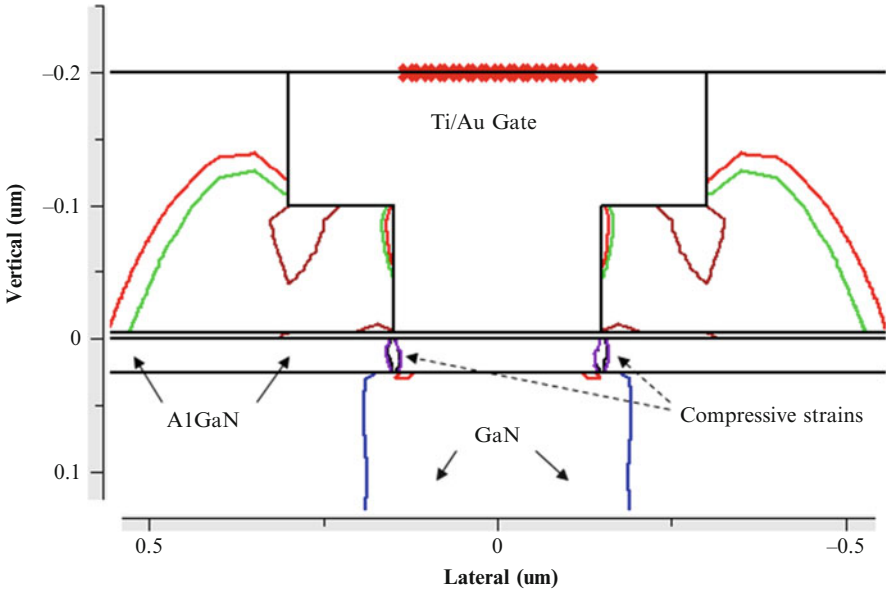



**Fig. 16.16** Simulation figure of GaN HEMT in the $V_{DS} = 0$ state with $V_G = -5$ V, $V_D/V_S = 30$ V. Mechanical strain contours due to the inverse piezo effect show highest compressive strains of ~−7e−4 in the AlGaN layer equally at both edges under the gate

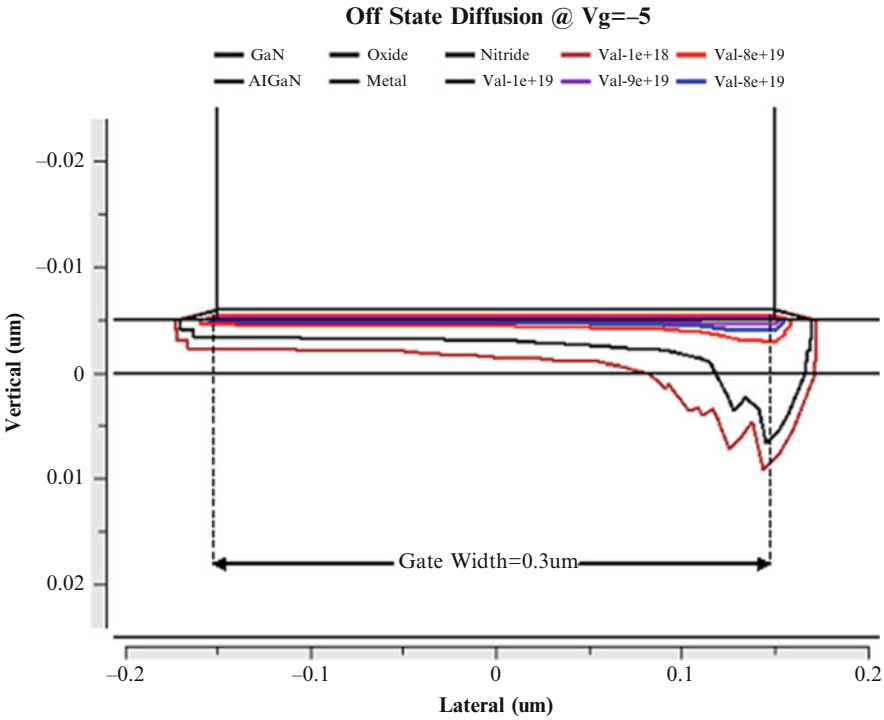



**Fig. 16.17** Simulation figure of impurity diffusion in the off state with $V_G = -5$ V, $V_{DS} = 30$ V. Concentration contours show enhanced diffusion at drain edge (*right side*) of gate where compressive strains are higher

Many researchers have consistently reported regions of material diffusion, cracking, and pitting at both gate edges. These features coincide with areas of increasing and concentrated compressive strains while in the off state, as seen from the strain simulation results. For GaN, it is then assumed that compressive strains correspond to increased diffusivity in these areas. For the AlGaN/GaN device, the modified diffusion is given as Eq. (16.34) in Table 16.2 where "misfit" is the lattice mismatch strain of the AlGaN on GaN, and the inverse piezoelectric strain has been previously described. The misfit strain is constant for a particular AlN concentration; therefore, the first part can be lumped as a constant, leaving the diffusion to be modulated by the strain from the inverse piezoelectric effect. Values of −150 eV/unit strain and 4.5e-18 $cm^2/s$ were used for $Q$ and $D_0$, respectively, in these simulations. Experimentally determined values for these particular terms have not yet been published, although the chosen diffusion coefficient $D_0$ is consistent with experimentally derived values for bulk diffusion of the measured trap Tp1 in AlGaN [35]. Figures 16.17 and 16.18 show the impurity diffusions around the gate, resulting from the strain distributions for the same two bias conditions previously outlined. Figure 16.19 shows impurity diffusion from a bias condition of $V_G = -10$ V, $V_{DS} = 30$ V.

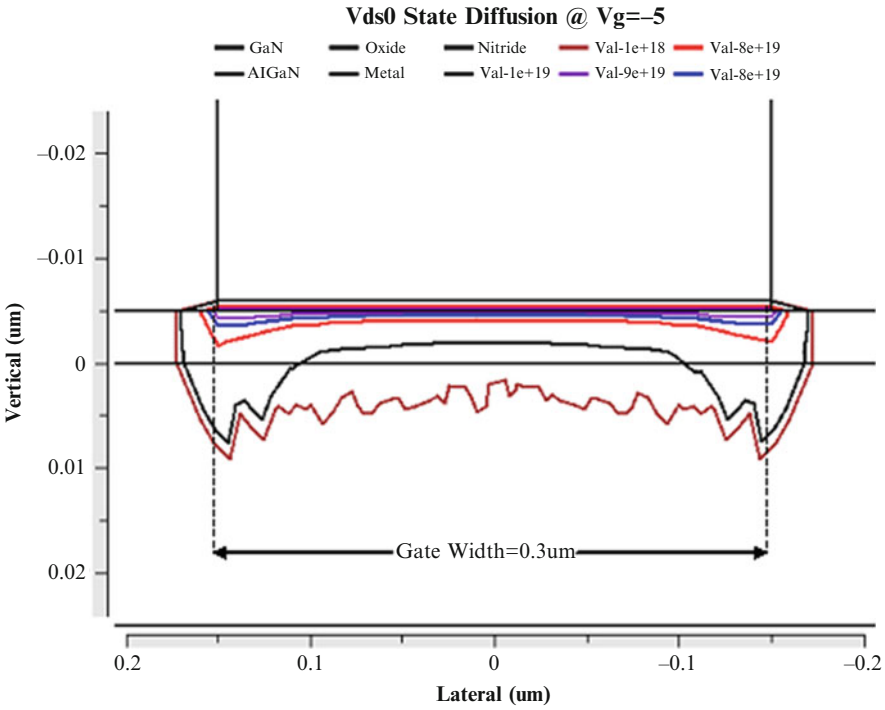



**Fig. 16.18** Simulation figure of impurity diffusion in the $V_{DS} = 0$ state with $V_G = -5$ V, $V_D = V_S = 30$ V. Concentration contours show enhanced diffusion on both sides of gate where compressive strains are higher

The resultant metal impurity profile is treated as profile of charged trapping sites that are added to Poisson's equation. The assumption of permanent trapped charge as used in the model is based on the observation of nonrecoverable changes in $I_{DS}$ during off-state step-stress experiments performed by [37] and nonrecoverable changes in trap density observed by [33]. Figure 16.20 shows the simulation of the altered structures IV characteristics. For a 10-V negative gate bias with $V_{DS}$ equal to 35-V stress, a 10.6% reduction in gate current is observed. Increasing current degradation is observed with increasing voltage stress on the device, in accordance with observed device behavior.

## 16.5 Conclusion

The FLOORS platform offers a capable way to simulate degradation in III-V devices. A variety of physics can be included. This chapter shows examples of self-heating simulation in HEMTs, adjustable MEM mirrors responding to heat

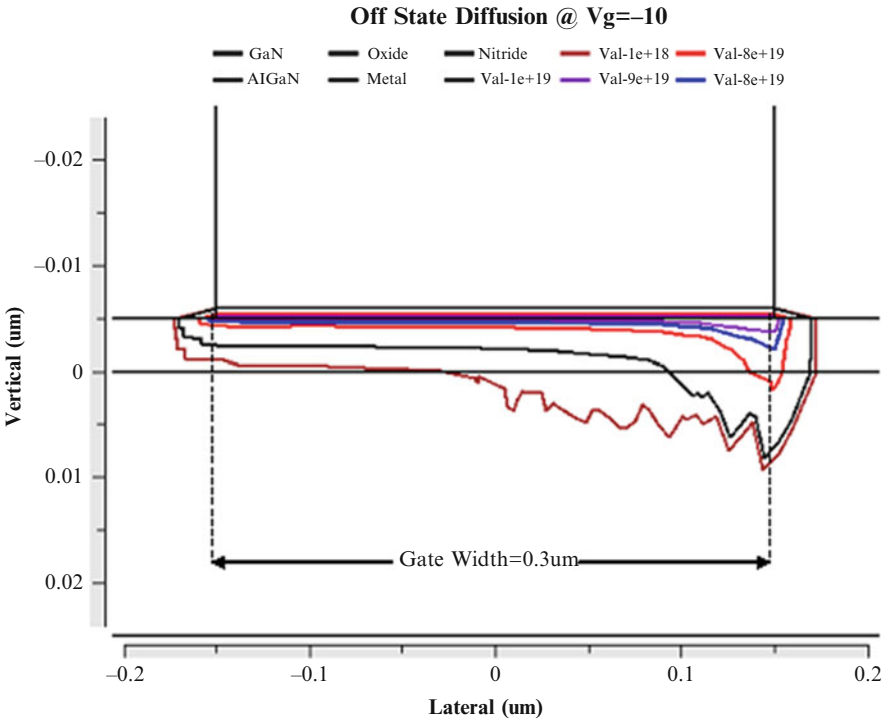



**Fig. 16.19** Simulation figure of impurity diffusion in the off state with $V_G = -10$ V, $V_{DS} = 30$ V. Concentration contours show deeper enhanced diffusion at drain edge(*right side*) of gate where compressive strains are higher

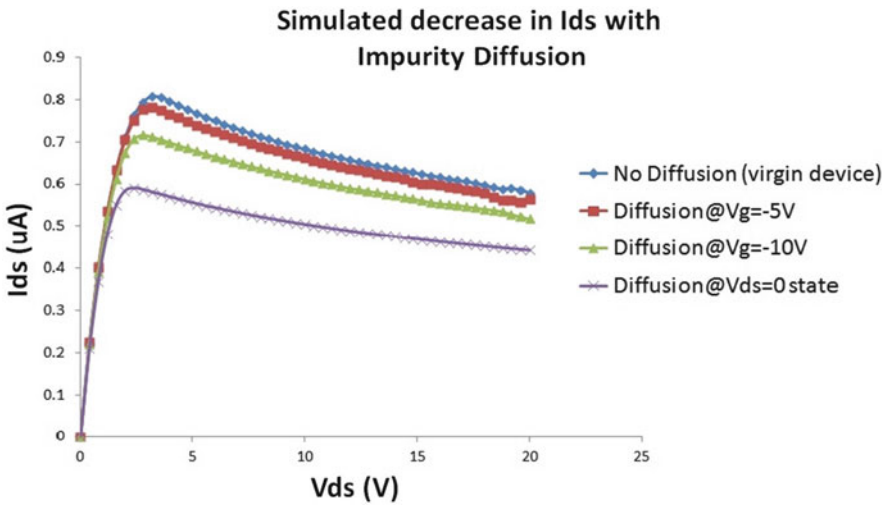



**Fig. 16.20** Current degradation in a HEMT for devices showing metal in diffusion in the gate region of the device

generation from current flow, and fully coupled thermal, electrical, and mechanical simulation of device structures.

Using FLOORS, two degradation examples were simulated. First, a simulation of the degradation of oxide materials under radiation as a function of hydrogen ambient was shown. Full models of hole motion, molecular and atomic hydrogen were shown fully coupled with reactions occurring at oxide vacancy defects. Second, a full simulation of the degradation behavior of GaN HEMTs in the off state has been performed. The degradation was modeled as compressive strain-driven impurity diffusion into the AlGaN layer and has accounted for both a critical voltage and a subsequent decrease in $I_{DS}$.

## References


1. D. Scharfetter, H. Gummel, IEEE Trans. Electron Devices **16**(1), 64–77 (1969)
2. M.R. Pinto, PhD thesis, Stanford University (1990)
3. R. Bank, D. Rose, W. Fichtner, IEEE Trans. Electron Devices **30**(9), 1031–1041 (1983)
4. J. Machek, S. Selberherr, IEEE Trans. Electron Devices **30**(9), 1083–1092 (1983)
5. S. Micheletti, Comput. Vis. Sci. **3**, 177–183 (2001)
6. D. Cummings, M.E. Law, S. Cea, T. Linton, *SISPAD* (September 2009)
8. G. Wachutka, IEEE Trans. Comput.-Aided Des. **9**(11), 1141–1149 (1990)
9. S.P. Gaur, D.H. Navon, IEEE Trans. Electron Devices **ED-23**, 50–57 (1976)
10. M.S. Adler, IEEE Trans. Electron Devices **ED-25**, 16–22 (1978)
11. H. Elschner, Nachrichientech. Electron **29**, 415–418 (1979)
12. A. Chryssafis, W. Love, Solid-State Electron. **22**, 249–256 (1979)
13. A. Venkatachalam, W.T. James, S. Grahm, Semicond. Sci. Technol. **26**, 085027 (2011) (6 pp.)
14. F. Gao, H. Lo, R. Ram, T. Palacios, *Device Research Conference*, June 2010, pp. 127–128
15. M. der Maur, G. Penazzi, G. Romano et al., IEEE Trans. Electron Devices **58**(5), 1425–1431 (2011)
16. D. Horton, F. Ren, L. Lu, M.E. Law, *An Electro-Mechanical Simulation of Off State AlGaN/GaN Device Degradation*, Proceedings of the International Reliability Physics Symposium, April 2012
17. T.P. Ma, P.V. Dressendorfer (eds.), *Ionizing Radiation Effects in MOS Devices and Circuits* (Wiley, New York, 1989)
18. R.A. Weeks, Paramagnetic resonance of lattice defects in irradiated quartz. J. Appl. Phys. **27**(11), 1376–1381 (1956)
19. P.M. Lenahan, P.V. Dressendorfer, Hole traps and trivalent silicon centers in metal/oxide/silicon devices. J. Appl. Phys. **55**(10), 3495–3499 (May 1984)
20. J.F. Conley Jr., P.M. Lenahan, Room temperature reactions involving silicon dangling bond centers and molecular hydrogen in amorphous SiO2 thin films on silicon. IEEE Trans. Nucl. Sci. **39**(6), 2186–2191 (Dec. 1992)
21. J.F. Conley Jr., P.M. Lenahan, Molecular hydrogen, E0 center hole traps, and radiation induced interface traps in MOS devices. IEEE Trans. Nucl. Sci. **40**(6), 1335–1340 (1993)
22. X.J. Chen, H.J. Barnaby, B. Vermeire, K. Holbert, D. Wright, R.L. Pease, G. Dunham, D.G. Platteter, J. Seiler, S. McClure, P. Adell, Mechanisms of enhanced radiation-induced degradation due to excess molecular hydrogen in bipolar oxides. IEEE Trans. Nucl. Sci. **54**(6), 1913–1919 (2007)
23. N.S. Saks, D.B. Brown, Observation of H + motion during interface trap formation. IEEE Trans. Nucl. Sci. **37**(6), 1624–1631 (1990)

24. S.N. Rashkeev, D.M. Fleetwood, R.D. Schrimpf, S.T. Pantelides, Defect generation by hydrogen at the Si-SiO2 interface. Phys. Rev. Lett. **87**(16), 165506 (2001)
25. B.R. Tuttle, D.R. Hughart, R.D. Schrimpf, D.M. Fleetwood, S.T. Pantelides, Defect interactions of H2 in SiO2: Implications for ELDRS and latent interface trap buildup. IEEE Trans. Nucl. Sci. **57**(6), 3046–3053 (2010)
26. T.R. Oldham, *Ionizing Radiation Effects in MOS Oxides* (World Scientific, River Edge, 1999), pp. 13–17
27. M.R. Shaneyfelt, D.M. Fleetwood, J.R. Schwank, K.L. Hughes, Charge yield for cobalt-60 and 10-keV X-ray irradiations of MOS devices. IEEE Trans. Nucl. Sci. **38**(6), 1187–1194 (1991)
28. S.R. de Groot, P. Mazur, *Non-Equilibrium Thermodynamics* (North-Holland Publishing Company, Amsterdam, 1963)
29. N.L. Rowsey, M.E. Law, R.D. Schrimpf, D.M. Fleetwood, B.R. Tuttle, S.T. Pantelides, IEEE Trans. Nucl. Sci. **58**, 2940 (2011)
30. P.M. Fahey, P.B. Griffin, J.D. Plummer, Point defects and dopant diffusion in silicon. Rev. Mod. Phys. **61**, 289–384 (Apr 1989)
31. R.L. Pease, P.C. Adell, B.G. Rax, X.J. Chen, H.J. Barnaby, K.E. Holbert, and H.P. Hjalmarson, *The effects of hydrogen on the enhanced low dose rate sensitivity (ELDRS) of bipolar linear circuits*, IEEE Trans. Nucl. Sci., vol. 55, no. 6, pp. 3169–3173, Dec. 2008
32. N.L. Rowsey, M.E. Law, R.D. Schrimpf, D.M. Fleetwood, B.R. Tuttle, S.T. Pantelides, Radiation-induced oxide charge in low- and high-$H_2$ environments. *RADECS'11 Proceedings*, Sevilla, Spain, Sept. 2011
33. M. Kuball, M. Tapajna, R.J.T. Simms, M. Faqir, U.K. Mishra, Microelectron. Reliab. **51**(2), 195 (2011)
34. L.Liu, F. Ren, S.J. Pearton, R.C. Fitch, D.E. Walker, K.D. Chabak, J.K. Gillespie, M.Kossler, M. Trejo, David Via, and A. Crespo, Investigating the effect of off-state stress on trap densities in AlGaN/GaN high electron mobility transistors J. Vac. Sci. Technol., B. Microelectron. Nanom. Struct. (2011)
35. M. Ťapajna, U. K. Mishra, and M. Kuball Importance of impurity diffusion for early stage degradation in AlGaN/GaN high electron mobility transistors upon electrical stress Appl. Phys. Lett. **97**, 023503 (2010)
36. K.C. Hall, S.W. Leonard, H.M. van Driel, A.R. Kost, E. Selvig, D.H. Chow, Appl. Phys. Lett. **75**(26), 4210 (1999)
37. J. Joh, J.A. del Alamo, J. Electron Device Lett. **29**(4), 287 (2008)

# Chapter 17
# The Analysis of Wide Band Gap Semiconductors Using Raman Spectroscopy

**Sukwon Choi, Eric Heller, Donald Dorsey, and Samuel Graham**

**Abstract** One of the main issues that must be addressed in the development of AlGaN/GaN HEMTs is the concern over device reliability. Due to the heteroepitaxial growth and other processing methods, mechanical stresses are inherent in these devices that can negatively impact device performance. Stress-induced effects on device performance due to Al content, (Zhang et al, J Appl Phys, 87:7981–7987, 2000), passivation, and substrate-induced residual stress in the GaN layer have been reported. In addition to impacting the 2-DEG, excessive strain in the AlGaN layer can induce crystallographic defects which lead to degraded device performance and shortened lifetimes. Under high-voltage conditions, the large electric field in the vicinity of the drain side gate edge is responsible for generation of large in-plane tensile strains in the AlGaN layer. These strains can result in device degradation via the formation of electrically active defects or trap states near the gate in the AlGaN. The direct measurement of the residual strain in the AlGaN layer is not simple due to its thickness (on the order of 25 nm). Instead, observation of the through-thickness variation of residual strain in the GaN layer allows a first-order estimation of the imposed strains in the AlGaN layer. In this chapter, we detail the use of Raman spectroscopy for characterizing AlGaN/GaN HEMTs.

S. Choi • S. Graham (✉)
Woodruff School of Mechanical Engineering, Georgia Institute of Technology, 30332-0405 Atlanta, GA, USA
e-mail: sgraham@gatech.edu

E. Heller • D. Dorsey
Materials and Manufacturing Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base, 45433, OH, USA

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_17,
© Springer Science+Business Media New York 2013

## 17.1 Introduction

### *17.1.1 AlGaN/GaN HEMTs*

Gallium nitride (GaN)-based systems are under development due to their potential use in both advanced electronics and optoelectronic devices. GaN is a direct wide band gap (~3.4 eV at room temperature) semiconductor that exists normally in a wurtzite phase which exhibits strong polarization in the *c*-direction (or <0001> direction).

High electron mobility transistors (HEMTs) or heterostructure field effect transistors (HFETs) made from AlGaN/GaN facilitate the ability to overcome the inevitable trade-off between high frequency and high current operation for conventional metal semiconductor field effect transistors (MESFETs) or metal oxide semiconductor field effect transistors (MOSFETs). The absence of dopants within the channel region prevents ionized impurity scattering which suppresses the electron mobility. High sheet charge density ($>10^{13}$ cm$^{-2}$) is obtainable depending on the modulation doping in the large band gap material or the polar charge and conduction-band discontinuity at the heterointerface (polarization doping). High carrier density is maintained at very low temperatures which enables special applications such as deep space signal reception. Also, utilization of superior materials for the active channel which offer high mobility but are difficult to process is viable since only a thin layer is required.

AlGaAs/GaAs and AlGaN/GaN are typical structures employed to build HEMTs. The electrons in the two-dimensional electron gas (2-DEG) of AlGaAs/GaAs HEMTs are provided from a doner sheet in the AlGaAs layer, whereas for AlGaN/GaN, the strong pyroelectric and piezoelectric nature of wurtzite GaN (and AlGaN) results in the formation of the 2-DEG where majority of the electrons originate from ionized surface donor states. The net polarization charge created at the AlGaN/GaN heterointerface is the sum of polarization contributions from both the AlGaN and GaN layers. Generally, the AlGaN layer is under tensile strain because of the lattice mismatch between AlGaN and GaN. Due to the noncentrosymmetric crystal nature of III–V materials, the polarization contribution from the AlGaN layer is the sum of the spontaneous and piezoelectric polarizations. The net polarization charge is a positive number for a Ga(Al)-face heterostructure [1]. As a consequence, a vast amount of free electrons tend to compensate the polarization induced charge and form the 2-DEG in the GaN buffer near the heterointerface. This 2-DEG is used as the channel region of the HEMT. The 2-DEG sheet carrier density in an AlGaN/GaN system is above $10^{13}$ cm$^{-2}$ which is an order of magnitude higher than that of GaAs-based HEMTs. This translates into their high current-carrying capabilities. Furthermore, the electron mobility and electron velocity in GaN remains at fairly high values even at elevated temperatures. Although Si- and GaAs-based structures can offer very high mobility and electron velocity, they have low band gaps. GaN's wide band gap results in a large breakdown field (~3 MV/cm) which is about three times higher than that for Si or GaAs. These superior material properties enable high power RF operation (large

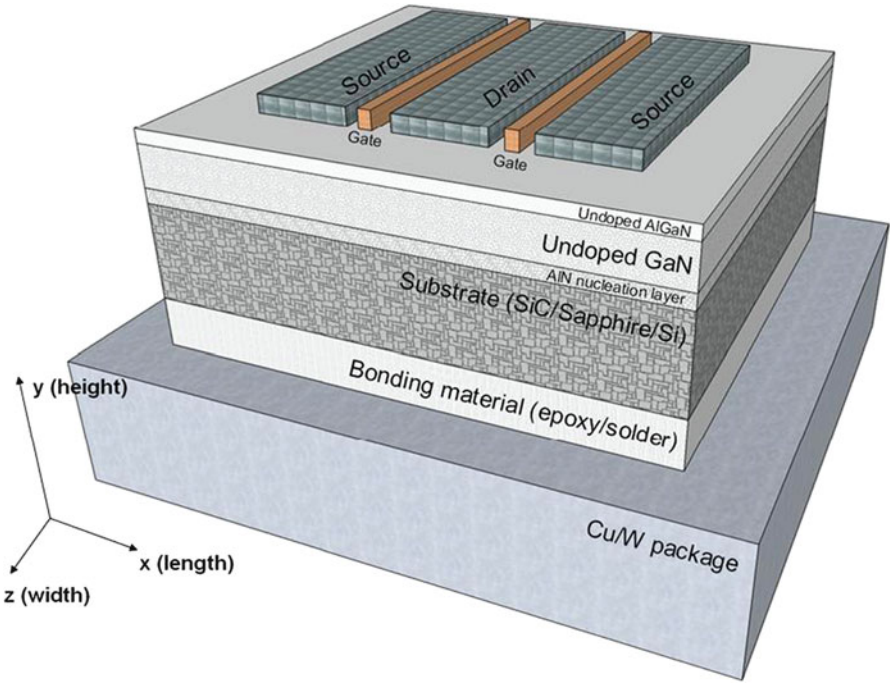



**Fig. 17.1** AlGaN/GaN HEMT structure

voltage and large current operation simultaneously) thus render AlGaN/GaN HEMTs as promising components for high temperature, high power, and high frequency electronics. They are planned for or already finding wide applications in RF sources and power amplifiers for satellite communications, high performance radars, and commercial ground base stations for advanced wireless communication schemes. There is an increasing demand in GaN electronics in automotive industry for developing more efficient power conditioning circuitry (inverter application) in hybrid electric vehicles. Active research is under progress for heterogeneous on-wafer integration of GaN and Si electronics as an alternative of direct competition with existing Si technologies [2].

Due to lack of availability of GaN substrates, GaN is typically grown by epitaxial techniques such as MBE or MOCVD on sapphire, Si, or SiC substrates. A heterostructure is formed by depositing AlGaN on top of the GaN layer. Normally the GaN and AlGaN layers need not be intentionally doped. Ohmic source and drain contacts are formed and the Schottky gate is deposited on the barrier material (AlGaN). A dielectric passivation layer covers the device which prevents charging of undesired surface states and protects the device from contaminants. Advanced design schemes can be implemented to control reliability issues in AlGaN/GaN HEMTs. For example, source connected field plates and T-gates can be added to mitigate the high electric field which develops at the drain edge of the gate in the presence of a large voltage bias. A typical AlGaN/GaN HEMT structure is depicted in Fig. 17.1.

Although these devices demonstrate great potential and superior performance, reliability issues limit the widespread use or commercialization of AlGaN/GaN HEMTs. Hot electrons induced by high electric fields can be trapped on the device surface, in AlGaN, or in the GaN buffer and even generate traps which degrade device performance [3]. The inverse piezoelectric effect, on the other hand, can drive a failure mechanism as a result of the electric field in the device where excessive tensile strain induced in AlGaN leads to formation of crystallographic defects in this layer [4]. The bias condition for operation is an important factor that determines which failure mode plays a dominant role in device degradation. This was shown in [3] comparing the impact of three bias conditions on device degradation, namely, pinch-off state ($V_{gs} = -7.7$ V, $V_{ds} = 20$ V: highest gate-to-drain electric field in absence of carriers in the channel), semi-on state ($V_{gs} = -5.5$ V, $V_{ds} = 20$ V: high electric field and existence of hot carriers in the channel), and on state ($V_{gs} = 0$ V, $V_{ds} = 20$ V: Lowest electric field and highest population of less energetic electrons in the channel). Hot-electron degradation prevails at semi-on conditions with moderate $V_{ds}$ (<30 V) since a fairly high current and electric field present in the device induces a large population of hot electrons. At pinch-off states with high $V_{ds}$ (>30–50 V), the inverse piezoelectric effect becomes prominent due to highest $V_{gd}$ in the device. Excessively high junction temperature has potential to degrade the stability of Schottky gate and ohmic contacts and form material defects (wear-out), thus causing device failure or reliability problems. Active research is being done on the degradation physics of AlGaN/GaN HEMTs. The foregoing statements are only a portion of possible degradation scenarios reported for these devices.

Understanding of device degradation physics is the first step to improve the device reliability. For this reason, not only conducting electrical measurements along with accelerated lifetime tests but also mechanical characterization is essential. Micro-Raman spectroscopy and micro-photoluminescence (PL) are noncontact, local, and nondestructive optical material characterization techniques with spatial resolutions on the order ~1 μm adequate for investigation of AlGaN/GaN HEMTs in terms of examining mechanical stress and temperature. In the following sections, a method to comparatively measure mechanical stress by means of micro-Raman spectroscopy and micro-photoluminescence will be developed. Also, various Raman measurement techniques for retrieving temperature information of GaN will be compared and discussed. Practical examples of implementing these stress and temperature measurement techniques to device investigation will be presented.

### *17.1.2 Raman Scattering*

Raman spectroscopy is an optical scattering technique that is used to probe the vibrational energies of the optical phonons in a material. Raman scattering is observed by the energy difference between the incident light and radiation scattered from the sample which corresponds to the molecular vibrational frequency. In the

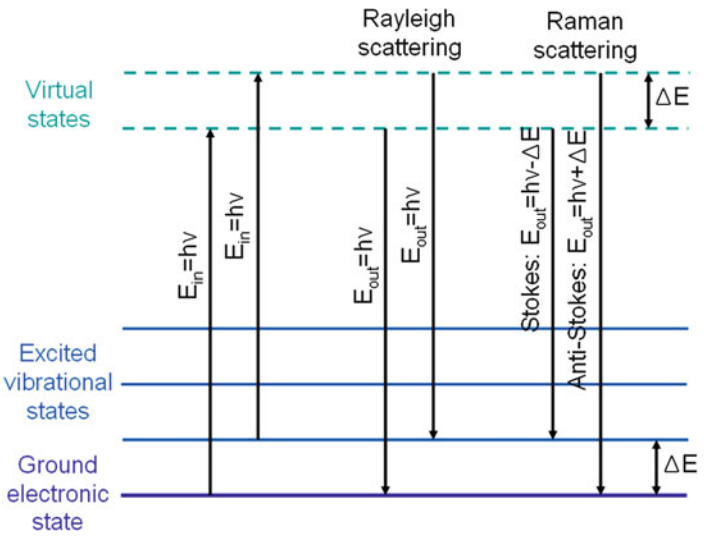



**Fig. 17.2** Energy level diagram for Rayleigh and Raman scattering processes

light scattering process, the electric field of the radiation interacts with a molecule by distorting the electron cloud around the chemical bonds in the molecule. Electrons are promoted to a so-called virtual state and return to their ground state by reradiating photons. Here the dominant process is elastic scattering, that is, no energy change, where the frequency of the light incident upon the sample and that of the scattered light are identical. This process is called Rayleigh scattering. However, there exists an inelastic scattering process which occurs for one out of $10^6$–$10^8$ photons such that quantized vibrational energy levels of the molecule get involved during the scattering process. As a consequence, this vibrational energy is transferred either from the incident photon to the molecule (Stokes) or from the molecule to the scattered photon (anti-Stokes). This process is Raman scattering. The foregoing elucidation of Raman scattering was based on noninteracting randomly oriented molecules in gas or liquid phase. When solid samples such as semiconductors are under the Raman scattering process, the energies of quantized lattice waves or so-called phonons are observable. The Raman scattering process is illustrated in Fig. 17.2.

Modern Raman instruments are equipped with lasers which are capable of irradiation on a sample with very high photon density. Thus, Raman spectroscopy has become a routine tool comparable to IR absorption spectroscopy. Usually, Raman scattering is recorded only on the low energy side from the energy of the source laser beam to give the Stokes Raman shift since the Stokes signal intensity is much higher than that of the anti-Stokes signal at room temperature. Using a harmonic approximation for the potential energy curve between atoms that constitute the molecule, it is easily deduced that lighter the atoms are and stronger the bond is, higher the vibrational frequency will be. Raman scattering is expressed as a

shift in energy from that of the exciting radiation usually in wavenumbers ($cm^{-1}$). In Raman spectroscopy, this linear wavenumber scale is commonly used because it has a direct proportionality with energy and frequency. The relationship among wavenumber ($\bar{\omega}$), energy ($E$), frequency ($v$), and radiation wavelength ($\lambda$) is $E = hv = hc/\lambda = h\bar{\omega}$, where $h$ is Planck's constant ($6.626 \times 10^{-34}$ J·s) and $c$ is the speed of light ($2.998 \times 10^{8}$ m/s).

Molecular vibrations that cause a change in the polarizability of the molecule which varies as a function of the distance between constituent nuclei result in efficient Raman scattering (Raman active). This contrasts to vibrations that induce a net change in dipole moment or charge distribution whereon absorption of the radiation takes place when the radiation frequency (mid-IR region) exactly matches the natural vibrational frequency of the molecule (IR active). This distinction is referred to as the selection rule. Raman spectroscopy and IR absorption are thus complementary techniques, especially for noncentrosymmetric molecules, to extract information of the vibrational characteristics of a molecule. An excellent presentation of a classical approach to the theory of Raman scattering is given in more detail in the literature [5].

### *17.1.3 Photoluminescence Spectroscopy*

As its name manifests, excitation of electrons within a material is brought about by absorption of photons in photoluminescence. Photoexcitation promotes electrons to permissible excited states. These electrons release the excess energy as they thermalize to their equilibrium state either by emission of photons, that is, radiative processes (fluorescence, phosphorescence), or nonradiative processes (vibrational relaxation, internal/external conversion, intersystem crossing). The favored deactivation route to the equilibrium or ground state is the one that minimizes the lifetime of the excited state. The energy of the emitted light corresponds to the difference in energy between the excited state and the equilibrium state involved in the electronic energy transitions.

Photoluminescence processes in semiconductors are illustrated in Fig. 17.3, which plots the $E$–$k$ diagrams for direct band gap semiconductors and indirect band gap semiconductors. Here $E$ and $k$ are the kinetic energy and wave vector of the electron or hole, respectively. When a semiconductor is illuminated with light, electron–hole pairs are created as photons are absorbed. As the electron–hole pairs undergo recombination, emission of photons takes place with a random direction phase and polarization. The most common radiative transition in semiconductors is between states in the conduction and valence bands, where this energy of interband transition is known as the band gap energy. Here it is assumed excitation is carried out with a photonic energy beyond the band gap of the material. Thus, photoluminescence spectroscopy gives a direct measure of the band gap of semiconductors. It should be noted that the phonon energy (~0.01 eV) is much smaller than the energy of emitted photons in the photoluminescence process. Thus, rough estimation of the

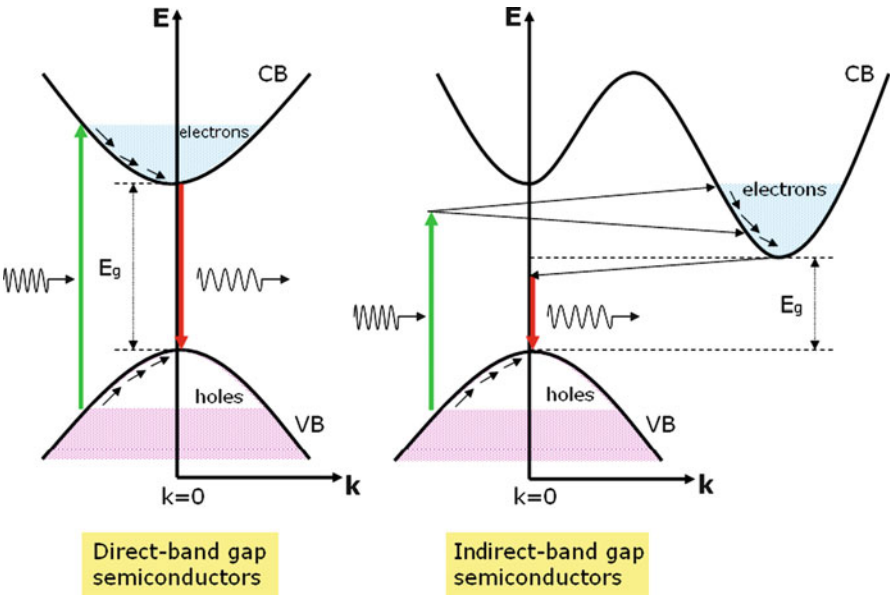



**Fig. 17.3** Photoluminescence processes in direct band gap and indirect band gap semiconductors. Electrons and holes created by absorption of photons are represented as the *shaded* states near the *bottom* of the conduction band and the *empty* states near the *top* of the valence band, respectively. Photon absorption and emission processes are represented by *vertical arrows* on the *E–k* diagrams according to energy and momentum conservation since momentum of the photon is negligible compared to that of the electron. Transitions within the conduction and valence bands represent rapid thermalization processes of the excited electrons and holes via phonon emission. For indirect band gap semiconductors, photon absorption process must be assisted with either phonon absorption or emission, and photon emission requires phonon emission to conserve momentum

band gap of indirect band gap semiconductors is also possible. Special features of photoluminescence spectroscopy on semiconductors other than band gap determination are impurity level/defect detection, recombination mechanism analysis, material quality measurement, etc.

## 17.2 Stress Measurement

### *17.2.1 Overview*

One of the main issues that must be addressed in the development of AlGaN/GaN HEMTs is the concern over device reliability. Due to the heteroepitaxial growth and other processing methods, mechanical stresses are inherent in these devices that can negatively impact device performance. The 2-DEG in AlGaN/GaN HEMTs is responsible for the high current capability and is created in response to the large

spontaneous/piezoelectric polarization in AlGaN which causes electrons to drift towards the heterointerface and aggregate into the quantum well. The electronic performance of these devices is mainly limited by the sheet carrier density and electron mobility in the 2-DEG. These electrical properties are a strong function of the mechanical strain built in the AlGaN layer. For example, imposing tensile in-plane strain in the AlGaN layer increases the piezoelectric field thus results in higher 2-DEG sheet carrier density in Ga(Al)-face heterostructures. Stress-induced effects on device performance due to Al content [6, 7], passivation [8], and substrate-induced residual stress in the GaN layer [9] have been reported. In addition to impacting the 2-DEG, excessive strain in the AlGaN layer can induce crystallographic defects which lead to degraded device performance and shortened lifetimes [4]. Simply operating AlGaN/GaN HEMTs leads to a vertical electric field in the device structure which generates inverse piezoelectric strain. Under high voltage conditions, the large electric field in the vicinity of the drain side gate edge is responsible for generation of large in-plane tensile strains in the AlGaN layer. These strains can result in device degradation via the formation of electrically active defects or trap states near the gate in the AlGaN [4]. Mitigation of the inverse piezoelectric stress can be realized by smoothing the peak electric field by introducing field plates. From the mechanics point of view, minimizing the built-in residual strain in the AlGaN layer at nonbiased conditions is important but accompanies a trade-off in less output power due to a smaller 2-DEG sheet carrier density.

The direct measurement of the residual strain in the AlGaN layer is not simple due to its thickness (on the order of 25 nm). Instead, observation of the through-thickness variation of residual strain in the GaN layer allows a first-order estimation of the imposed strains in the AlGaN layer. At present, most of commercial GaN layers implemented in AlGaN/GaN HEMTs are obtained by heteroepitaxy on substrates with lattice and symmetry similarities due to limitations in commercially available free-standing GaN substrates. Overall, thin GaN heteroepitaxial layers inevitably possess a residual strain (or stress) due to the growth process on lattice-mismatched substrates and post-growth cooling with difference in thermal expansion coefficients between the substrate material and the GaN epitaxial layer. This stress is commonly biaxial across the thin GaN film and stress free in the growth direction. Strain from defects, which is usually tensile, arises due to point defects, dislocations, stacking faults, and twins [10]. Thus, details of the buffer layer also modify the stress in the GaN layer significantly [11]. It has been reported that GaN grown on sapphire contains compressive stress, whereas GaN on 6H-SiC and Si is under tensile stress [12–14]. Excessive residual stress induces formation of dislocations, stacking faults, and cracking near the GaN–substrate interface which leads to degraded performance and shortened lifetime in GaN-based electronic devices [15].

A consistent description of the biaxial and hydrostatic strain in GaN thin films utilizing Raman spectroscopy, photoluminescence, and x-ray diffraction was reported [16]. A combined analysis employing Raman spectroscopy and photoluminescence to investigate biaxial stress in GaN epilayers grown on Si,

6H-SiC, and sapphire was presented. Roles of the AlN buffer, lattice, and thermal mismatch between the GaN epilayers and the substrates to residual strain in the films were discussed [17]. A critical review of discrepancies between experimental and theoretical results in literature with regard to elastic properties, electronic as well as phonon deformation potentials, and the respective strain and stress coefficients of uniaxially and biaxially strained GaN and AlN was given based upon ab initio studies using a pseudopotential plane-wave method [18]. The relationship between band gap and biaxial stress of wurtzite ZnO and GaN was investigated by side-inclination x-ray diffraction, optical absorbance spectrum, and ab initio calculation [19]. Raman spectroscopy was used to study the inverse piezoelectric strain in AlGaN/GaN HEMTs via observation of the $E_2$(high) phonon frequency shift between biased (pinch-off) and unbiased states. The inverse piezoelectric strain in GaN was shown to be related with the magnitude of the vertical component of the electric field in GaN layer [20]. The magnitude of stress loads that evolve during AlGaN/GaN HEMT operation, that is, operational thermoelastic stress, inverse piezoelectric stress, and residual stress, was quantified by simultaneously utilizing the Stokes peak position and linewidth of the Raman signal. It was shown that GaN on Si devices exhibited more tensile residual stress at room temperature and more compressive thermoelastic stress under operation compared to GaN on SiC devices with similar structures, whereas the magnitude of inverse piezoelectric stress were comparable (compressive) [21].

### *17.2.2 Theory and Methodology*

Among the methods used to examine stress in epitaxial GaN, micro-Raman spectroscopy, and micro-photoluminescence (PL) are noncontact, local, and nondestructive optical material characterization techniques adequate for measuring strain in AlGaN/GaN HEMTs. Lateral resolutions are on the order ~1 μm, while the depth resolution depends on the excitation wavelengths and optical properties of the AlGaN/GaN layers. Micro-Raman spectroscopy probes the energy associated with Raman active phonon modes. The zone-center (Γ point, $k \approx 0$) phonons are responsible for the first-order Raman scattering due to the energy and momentum conservation rules in the light scattering process. The material response to applied strain is observed through the Stokes peak position shift of the observed Raman active modes from its true strain-free value [22]. With the laser incidence along the normal direction of the basal plane (i.e., parallel to the *c*-axis of the GaN layer) and a 180° backscattering configuration, the allowed zone-center optical phonon modes detectable for hexagonal wurtzite GaN ($C_{6v}{}^4$ space group symmetry) are $A_1$(LO) (~734 $cm^{-1}$), $E_2$(low) (~143 $cm^{-1}$), and $E_2$(high) (~568 $cm^{-1}$), unless there are disoriented microstructures [23]. The $E_2$(high) mode of GaN corresponds to the atomic oscillation in the *c*-plane which has a nonpolar character as shown in Fig. 17.4. Its mode frequency is therefore sensitive to the lattice strain in the

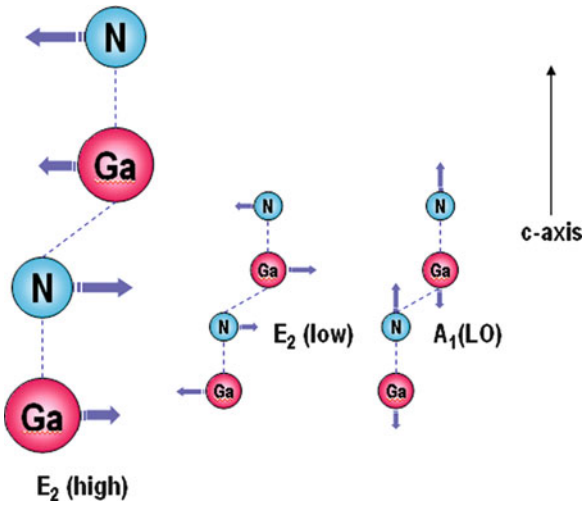



**Fig. 17.4** Graphical description of atomic vibration modes in hexagonal GaN

*c*-plane and suitable for examining the in-plane biaxial stress in GaN epifilms [24]. In general, the $E_2$(high) phonon frequency decreases in comparison to that of strain-free GaN as tensile biaxial stress is applied. This phenomenon can be explained by the elongation of the bonds perpendicular to the *c*-axis followed by relaxation of the force constants. The change in interatomic forces between atoms results in change in vibrational frequency.

PL is another optical characterization technique that measures the variation of strain level in GaN since the energy band gap is strongly sensitive to the state of strain. Alteration of interatomic distances as well as relative positions of atoms by mechanical strain alters the band structure of semiconductors followed by a change in band gap energy. In fact, strain engineering in semiconductor thin films has been employed to obtain desired physical and electronic properties. A decrease in the energy band gap of GaN results from tensile strain, whereas it increases with an applied compressive strain. It was shown by experiments followed by ab initio calculation that for wurtzite GaN, under biaxial tensile stress, the valence-band maximum (VBM) and the conduction-band minimum (CBM) offset to lower energies while the VBM offset is smaller than the CBM offset resulting in decrease in the energy band gap [19].

GaN with its energy band gap of about 3.4 eV at room temperature is transparent to visible laser sources. Thus, in utilizing a sub-band gap laser for Raman measurements, Raman data represents a volume average of the stress through the GaN layer (with thickness in the order of 1–2 μm). On the other hand, micro-PL shows the stress near the surface of the GaN since the photon penetration depth of the He–Cd 325-nm laser is of the order of 80–90 nm [25–27]. The combination of two techniques offers complementary information to reveal the through-thickness variation of stress in the GaN epilayer.

## *17.2.3 Measurement of Stress in Thin GaN Epilayers*

### 17.2.3.1 Motivation

Stress in GaN devices is not a directly observable quantity. Rather, measured strain is translated into stress through the knowledge of an ideal strain-free reference as well as knowledge of the elastic properties of the material under consideration. The degree of biaxial stress in GaN is known to have a linear relationship with both the $E_2$(high) phonon frequency and the luminescent band gap shift at a given temperature (for moderate stress ranges, e.g., from −1.5 to 0.5 GPa) [17]. With Raman and PL measurements, an accurate assessment of the slope of strain or stress versus frequency or energy shift is feasible, but the absolute magnitude of strain or stress depends on the choice of the strain-free reference. The strain-free $E_2$(high) mode frequency and energy band gap of GaN at room temperature are reported in literature over a wide range [11, 15–17]. This variation may stem from difference in the choice of GaN samples that would represent a strain-free perfect GaN crystal, details of the initial spectrometer calibration, and excitation laser power conditions utilized which have potential to locally heat the sample, especially for PL measurements. Stress is estimated by utilizing the slope of biaxial stress versus frequency or energy shift, and these stress conversion factors solely depend on the choice of elastic stiffness constants which suffer from a large scatter in published values [12, 16, 17, 28, 29]. Therefore, studies attempting to correlate biaxial stress with Raman and/or PL measurements show contradictory results. Therefore, to apply these stress measurement techniques to actual devices, it is highly desirable to use suitable parameters and accurate stress–strain relations to correlate a known stress state (i.e., uniaxial, biaxial, hydrostatic, etc.) to phonon frequency and/or band gap energy shift.

A unified framework incorporating x-ray diffraction (XRD), Raman spectroscopy, and PL was conducted to obtain the $E_2$(high) Raman peak and the PL band-edge emission peak representing zero-strain reference values of fully relaxed GaN as well as Raman and PL biaxial stress conversion factors. The relationships between optical phonon frequencies and band gap with biaxial stress were precisely determined.

### 17.2.3.2 Experimental

To determine the relationships between optical phonon frequencies, band gap energy, and biaxial stresses in the *c*-plane of unintentionally doped GaN films, a range of MOCVD and HVPE GaN films (bare epi) were explored. The films had a nominal thickness in the range of 0.5–5 μm and were grown on different substrates (two GaN/sapphire, one GaN/6H-SiC, one GaN/4H-SiC, and two GaN/Si templates) with epitaxial relationships described in Table 17.1. The samples were

**Table 17.1** Room temperature properties of substrate materials for epitaxy of GAN [30]

| Material | Crystal structure | Epitaxial relationship | In-plane direction GaN//Substrate | Lattice constant (Å) | | Lattice mismatch (%) | Thermal expansion coefficient ($10^{-6}$ $K^{-1}$) | |
|---|---|---|---|---|---|---|---|---|
| | | | | *a* | *c* | | ‖*a*-Axis | ‖*c*-Axis |
| GaN | Wurtzite | – | – | 3.1878 | 5.1850 | – | 4.997 | 4.481 |
| AlN | Wurtzite | (0001)/(0001) | $[10\bar{1}0]//[\bar{1}0\bar{1}0]$ | 3.1129 | 4.9819 | 2.41 | 5.411 | |
| Sapphire | Hexagonal | (0001)/(0001) | $[2\bar{1}\bar{1}0]//[01\bar{1}0]$ | 4.7589 | 12.9910 | −22.65 | 8.31 | 8.5 |
| 6H-SiC | Hexagonal | (0001)/(0001) | $[10\bar{1}0]//[10\bar{1}0]$ | 3.0806 | 15.1173 | 3.48 | 4.76 | 4.46 |
| 4H-SiC | Hexagonal | (0001)/(0001) | $[11\bar{2}0]//[11\bar{2}0]$ | 3.0800 | 10.0830 | 3.50 | 4.75 | |
| Si | Diamond | (0001)/(111) | $[11\bar{2}0]//[1\bar{1}0]$ | 5.4309 | | −16.99 | 3.9 | |

characterized using x-ray diffraction (XRD), micro-Raman spectroscopy, and micro-photoluminescence (PL). In Table 17.1, lattice mismatch is defined as

$$\frac{d_{\text{GaN}} - d_{\text{substrate}}}{d_{\text{substrate}}} \times 100(\%) \tag{17.1}$$

where $d_{\text{GaN}}$ and $d_{\text{substrate}}$ are the inter-planar spacings of the GaN film and substrate along the in-plane direction, respectively [30].

X-ray diffraction (XRD) measurements, the most direct approach for quantifying strain in epitaxial GaN, were carried out to measure lattice constants of the GaN samples using a PANalytical X'Pert Pro system equipped with a Cu Kα x-ray source. The *c*-axis lattice constants were measured by examining the symmetric reflections from the $(0002)$ and $(0004)$ planes. With this, together with the asymmetric reflections from the $(10\bar{1}1)$ and $(20\bar{2}2)$ planes, the *a*-axis lattice constants were calculated. The strains along the *a* and *c* axes are given by

$$\begin{aligned} \varepsilon_a &= \frac{a - a_0}{a_0} \\ \varepsilon_c &= \frac{c - c_0}{c_0} \end{aligned} \tag{17.2}$$

where *a* and *c* are the lattice spacing along the a and c axes and $a_0$ and $c_0$ are the relaxed, unstrained true lattice constants of pure GaN which must be known a priori. Bulk GaN crystals grown under high pressure have been accepted as the definition of the strain-free situation of GaN for a number of studies ($a_0 =$ 3.1880 A and $c_0 = 5.18561$ A [12], $a_0 = 3.1890$ A and $c_0 = 5.1864$ A [31], $a_0 =$ 3.1912 A [28]). However, it was reported that such bulk crystals contain high concentrations of oxygen (unintentionally doped) affecting lattice parameters (hydrostatic stress – expansion); thus, they are inappropriate as strain-free reference samples [32]. Also, residual strain can build up in bulk GaN crystals as the Ga-face and N-face have different growth rates and different dopant incorporation rates [33]. Instead, relaxed lattice constants ($a_0 = 3.1878$ and $c_0 = 5.1850$ A) estimated from homoepitaxial layers grown on bulk GaN [32, 34] were suggested to be representative of the true lattice parameters of a hypothetical perfect crystal of GaN. Lattice parameters ($a_0 = 3.1876$ and $c_0 = 5.1846$ A) of bulk crystals grown from solutions containing Mg with low Mg and O concentrations below $10^{-19}$ cm$^{-3}$ were shown to approach these ideal lattice constants [32]. Therefore, in this study the true lattice constant for pure GaN was assumed to be $a_0 = 3.1878$ and $c_0 = 5.1850$ A.

To relate strain to stress, the elastic constants of GaN must be known. A critical comparison of elastic stiffness constants as well as related moduli and coefficients of GaN in literature was reported by comparing values determined by measurements and ab initio calculations [18]. Elastic stiffness constants in [35] (C11 $= 390 \pm 15$ GPa, C33 $= 398 \pm 20$ GPa, C44 $= 105 \pm 10$ GPa, C66 $= 123 \pm 10$ GPa, C12 $= 145 \pm 20$ GPa, and C13 $= 106 \pm 20$ GPa) obtained by

Brillouin scattering measurements were employed as the elastic constants of wurtzite GaN. In case of wurtzite GaN films under homogeneous biaxial stress in the $c$-plane ($\sigma_a = \sigma_b$, $\sigma_c = 0$), the biaxial relaxation coefficient $R^{\mathrm{B}}$ and the Poisson ratio $v$ are expressed in terms of elastic stiffness coefficients as follows:

$$v = \frac{C_{13}}{C_{11} + C_{12}} \tag{17.3}$$

$$R^{\mathrm{B}} = -\frac{\varepsilon_c}{\varepsilon_a} = \frac{2C_{13}}{C_{33}} \tag{17.4}$$

The biaxial stress is related to the in-plane strain $\varepsilon_a$ by

$$\sigma_a = Y\varepsilon_a \tag{17.5}$$

where the biaxial modulus $Y$ is defined as

$$Y = C_{11} + C_{12} - 2C_{13}^2/C_{33} \tag{17.6}$$

The relationship between the in-plane biaxial stress and the $c$-axis strain can be easily derived as

$$\sigma_a = -\left(\frac{Y}{R^{\mathrm{B}}}\right)\varepsilon_c \tag{17.7}$$

In general, thin films having a wurtzite crystal structure exhibit an anisotropic behavior. Fortunately, however, the relation $C_{11} + C_{12} \approx C_{13} + C_{33}$ holds well for the elastic stiffness constants adopted here as well as others published [18, 19, 36]. This renders a quasi-cubic or isotropic elastic assumption valid for GaN [16, 19], whereas this relation does not hold for other wurtzite structured nitrides such as AlN and InN [37]. For elastically isotropic materials under homogeneous biaxial stress in the $c$-plane, Poisson's ratio $v$ along [0001] is related to the biaxial relaxation coefficient or strain ratio $R^{\mathrm{B}}$ by

$$R^{\mathrm{B}} = -\frac{\varepsilon_c}{\varepsilon_a} = \frac{2v}{1-v} \tag{17.8}$$

The biaxial stress is expressed in terms of Poisson's ratio $v$, Young's modulus $\hat{E}$ (we use this notation to distinguish Young's modulus from the luminescent band gap energy $E$), in-plane strain $\varepsilon_a$, and out-of-plane strain $\varepsilon_c$ by

$$\sigma_a = -\frac{1}{2v}\hat{E}\varepsilon_c = \frac{1}{1-v}\hat{E}\varepsilon_a \tag{17.9}$$

Since the elastically isotropic approximation holds for wurtzite GaN under the in-plane biaxial strain case under consideration, the biaxial modulus $Y$ and Young's modulus $\hat{E}$ are related as

$$\hat{E} = \frac{2v}{R^{\mathrm{B}}} Y = \frac{C_{33}Y}{C_{11} + C_{12}} \tag{17.10}$$

The elastic stiffness constants employed in this study yield $R^{\mathrm{B}} = 0.533 \pm 0.10$, $v = 0.198 \pm 0.039$ ($0.210 \pm 0.032$ from (17.8)), $Y = 479 \pm 33$, and $\hat{E} = 356 \pm 46$ GPa.

Micro-Raman spectroscopy and PL measurements were performed through the use of Horiba Jobin Yvon LabRAM HR800. A 532-nm diode-pumped solid-state laser and a 325-nm He-Cd laser were used as excitation sources for Raman and PL measurements, respectively. The spectrometer had a focal length of 800 mm and equipped with a back-illuminated UV-coated liquid N2-cooled 2,048 × 13.5 μm$^2$ pixels charge-coupled device (CCD) detector. A 1,800 groove/mm diffraction grating was used and the confocal hole size was adjusted to 100 μm. This setup was established to obtain high enough spectral resolution suited for stress measurements while having a reasonable acquisition time and spectral range for each measurement. Raman measurements using 532-nm excitation had a spectral resolution of 0.83 cm$^{-1}$/pixel of CCD defined following the rule that it takes 3 pixels minimum to define a peak. For PL measurements utilizing the 325-nm laser, a spectral resolution of 4.75 cm$^{-1}$/pixel ($5.89 \times 10^{-4}$ eV/pixel) of CCD was obtained. It should be noted that the typical peak widths were 2–3 cm$^{-1}$ or larger for the $E_2$(high) modes, and higher resolution using our 2,400 groove/mm grating did not improve seeing a sub-wavenumber shift. Software curve fitting algorithms provide sub-pixel peak position resolution, thus the precision to peak shifting is more than an order higher than the resolution mechanically limited by the spectrometer. Experiments were carried out with the laser light perpendicular to the basal plane of GaN in a 180° backscattering configuration and unpolarized detection. The $E_2$(high) mode was investigated among the Raman active modes observable for this configuration, namely, two $E_2$(TO) and the $A_1$(LO) modes [10]. The 532 and 325-nm laser powers incident on the GaN samples were adjusted via neutral density filters to ~25.2 mW (except for GaN/Si samples where 0.2 mW was used to prevent substrate heating) and ~0.003 mW, respectively. The sub-band gap laser wavelength (532 nm) used for Raman measurements prevents localized heating of GaN by laser light absorption. The 325-nm laser power used for PL measurements was determined by taking into account of minimized local heating, reasonable acquisition time, and prevention of material damage. A 50× LWD objective (NA = 0.5) was used for Raman measurements. 40× and 39× NUV objectives (NA = 0.5) were employed to focus the laser on the sample surface and collect the PL signatures. The theoretical diffraction-limited laser focused spot sizes (diameter) were 1.3 and 0.79 μm for the Raman and PL measurements, respectively, using the relation $d = 1.22\lambda/\mathrm{NA}$, where $d$ is the spot diameter, $\lambda$ is the laser wavelength, and NA is the numerical aperture. Before conducting Raman

and PL measurements, the system was calibrated at 25°C with a 4H-SiC piece to 776.599 $cm^{-1}$ which was equivalent to 520.46 $cm^{-1}$ for high quality single-crystal silicon. Si is a common Raman calibration source especially good for GaN investigation utilizing a 488-nm laser. However, for lasers used in this study (532 and 325 nm), acquiring a Si Raman peak with a reasonable intensity or signal to noise ratio at a reasonable acquisition time is not easy when laser powers are properly adjusted not to induce localized heating in Si. Therefore, 4H-SiC together with a plasma line of a fluorescent light source was used instead of Si. The long focal length offers higher spectral resolution but also renders the spectrometer more susceptible to thermal drift. To correct for the drift, the plasma lines of the external fluorescence source was employed as an additional reference during sample measurement. The shift from its initial peak value (482.215 ± 0.005 $cm^{-1}$) in the Raman spectra provided means to track the drift (systematic error) of the spectrometer itself in real time with each measurement which allows a much more time efficient measurement compared to the standard method of calibrating the spectrometer using Si before and after every sample measurement. In order to determine the peak position and full width at half maximum (FWHM), Raman and PL spectra were fitted using a linear combination of Gaussian–Lorentzian line-shapes. The spectral range for PL measurement was 3.15–3.6 eV. In order to determine the peak position of the near band-edge peak at room temperature, the PL peaks for each sample were resolved into two peaks, namely, the principal line (zero-phonon line) and its phonon replica line. The energy spacings between two peaks were close to a predicted value of 53.5 meV at 298 K estimated from the temperature dependence of the energy spacing between the 1LO phonon sideband and its principal line [38]. Best estimates and uncertainties (95% confidence bounds) were calculated based on at least 20 measurements for each sample or condition. Data acquisition times were adjusted to obtain consistent Raman (~8,000) and PL peak (~4,000) intensity counts. Samples were measured at room temperature while mounted on a heating stage (INSTEC HCC324S) maintained at 25°C (±0.06°C) to isolate stress measurement results from temperature effects, namely, shift in peak position and line broadening.

#### 17.2.3.3 Results and Discussion: Substrate Effects on Stress

The relationship between $a$ and $c$ lattice constants of the GaN epifilms are displayed in Fig. 17.5. The linear relationship between the lattice parameters means the epitaxial layers are elastically deformed by the built-in stress. Also shown are the lattice parameters of a 350-μm-thick free-standing GaN sample. Its lattice parameters are indeed larger than the assumed strain-free lattice constants [32, 34] showing the effect of unintentional doping. It is shown in Fig. 17.6 that biaxial stress is dominant in the GaN films since a nearly constant value of $\varepsilon_c/\varepsilon_a$ (the slope of the plot) is observed. Thus, hydrostatic strain introduced by the presence of point defects [16] in addition to biaxial strain induced through the substrate is not considered in the following stress analysis.

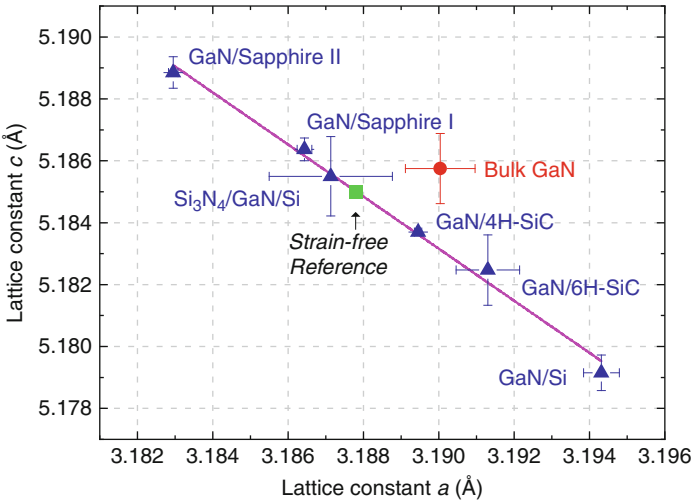



**Fig. 17.5** Lattice constant $a$ versus lattice constant $c$. The *triangles* indicate measured GaN templates. The *solid line* is the fitting result. The *circle* and *square* correspond to a GaN substrate and strain-free ideal lattice constants, respectively

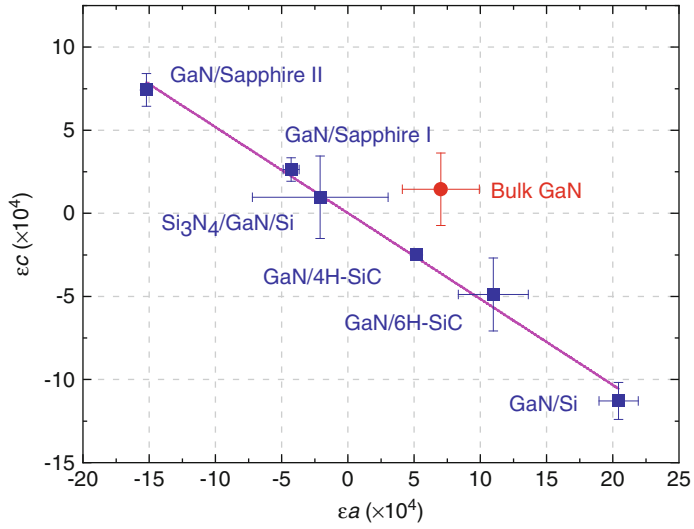



**Fig. 17.6** In-plane strain versus out-of-plane strain. The *squares* correspond to measured GaN templates. The *solid line* is the fitting result. The *circle* is indicative of a GaN substrate

As mentioned above, strain-free lattice constants were assumed to be $a_0 = 3.18$ 78 and $c_0 = 5.1850$ A. Strain perpendicular and parallel to the $c$-axis were obtained using the lattice parameters of the GaN epilayers measured by XRD. The biaxial relaxation coefficient or strain ratio obtained from the slope of the plot in Fig. 17.3

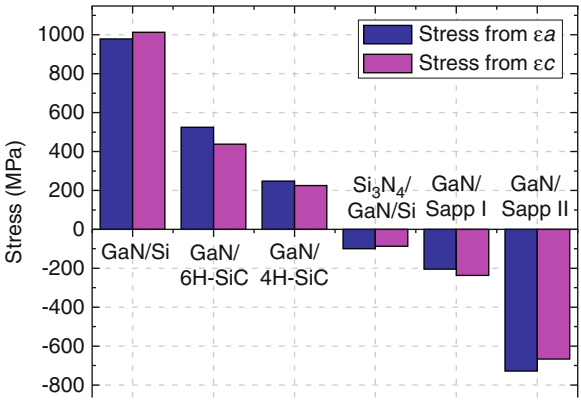



**Fig. 17.7** Biaxial stress in the basal plane of the measured GaN epifilms

or (17.8) is $R^{\mathrm{B}} = 0.519 \pm 0.06$ which is close to the theoretically predicted value of $R^{\mathrm{B}} = 0.533 \pm 0.104$ according to (17.4). The experimental Poisson's ratio is determined by (17.8) to be $v = 0.203 \pm 0.018$. This value agrees well with the theoretical value $v = 0.198 \pm 0.039$ predicted from (17.3) using the elastic stiffness constants employed in this study as well as a typically accepted value in literature of $v = 0.205$ [36, 37]. The biaxial stresses in the basal plane of the GaN epifilms were calculated by (17.5) and (17.7) which are displayed in Fig. 17.7. Here we adopt the convention that a positive value corresponds to tensile stress. The average of these stress values were assigned to represent the residual stress in each GaN epifilm and were used to extract the biaxial stress conversion factors and estimate the $E_2$(high) phonon frequency and band-edge peak corresponding to the strain-free case at room temperature. The stress uncertainties in each GaN epifilm were excluded in the ensuing statistical (error propagation) analysis for calculating the final uncertainties (95% confidence intervals) in the biaxial stress conversion factors and strain-free reference values.

The magnitude of the residual biaxial stress in an epitaxial GaN layer depends on type of substrate, thickness of buffer layer and epitaxial film, growth temperature, generation of misfit dislocations, material relaxation by cracking, etc. The different magnitude of biaxial stress in GaN epilayers (Fig. 17.7) can be explained by referring to the material parameters in Table 17.1. For the GaN/Si sample, the lattice constant of GaN is smaller than that of Si where the atomic spacing on the (111) plane is 3.84 Å (−16.99% lattice mismatch). The thermal expansion coefficient of GaN parallel to the a-axis is larger than that of Si. Thus, the lattice of GaN would contract more than that of Si during post-growth cooling. Both of these mismatches lead to a high tensile stress in the GaN epilayer although a buffer layer is implemented. For the sample GaN/6H-SiC, both the lattice mismatch (+3.5%) and thermal expansion coefficient discrepancy are relatively small compared to GaN/Si. The lattice constant of GaN is slightly larger than that of SiC which is expected to induce a compressive stress in GaN. On the other hand, thermal expansion coefficient of GaN is slightly larger than that of SiC. A tensile residual stress is built in the GaN layer at room temperature when the residual thermal stress

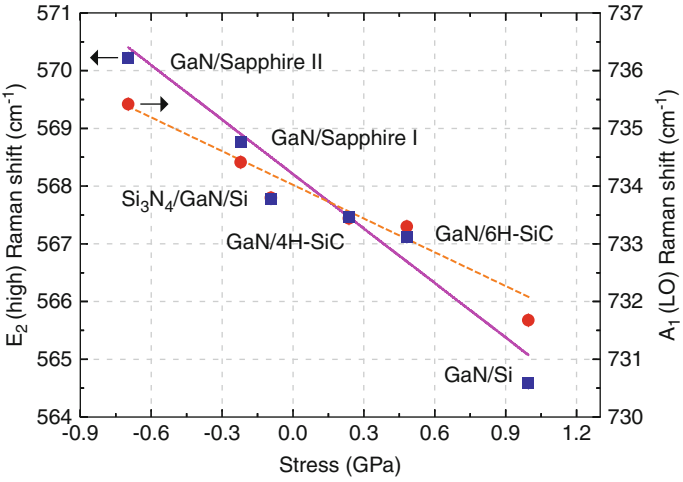



**Fig. 17.8** Room temperature Raman shifts versus corresponding residual stresses obtained by XRD measurements. Both $E_2$(high) and $A_1$(LO) mode relations are displayed as *squares* and *circles*, respectively. The *solid line* is the fitting of $E_2$(high) frequencies whereas the *dashed line* is the result of $A_1$(LO) frequencies

built in the GaN layer after cooling from deposition temperature to room temperature overcompensates the influence of lattice mismatch. In case of not implementing a buffer layer in the structure, a relatively large stress value can remain in the epitaxial GaN layer. The lattice mismatch between GaN and sapphire is quite large (−22.65%). However, the sapphire thermal expansion coefficient is much larger than that of GaN which is responsible for the highly compressive residual stress in GaN at room temperature. The residual stress induced by lattice constant mismatch relaxes within less than several nanometers of the film thickness during growth [13, 24]. For all cases measured here (except bulk GaN), the substrate-induced stress would virtually all reside in the GaN epitaxial layer (and the buffer layer) since the GaN layer was much thinner than its substrate.

Figure 17.8 shows the correlation between room temperature Raman spectra and biaxial stresses of the measured samples. As expected, the $E_2$(high) phonon mode positions are stress dependent which are mainly caused by the substrate type and growth process. The linear relationship between the frequency shift of the $E_2$(high) phonon mode and in-plane biaxial stress $\sigma_a$ in GaN can be expressed as

$$\omega - \omega_0 = K_{\mathrm{RS}}^{E_2(\mathrm{high})}\sigma_a \tag{17.11}$$

where $\omega$ is the measured $E_2$(high) phonon frequency, $\omega_0$ is the strain-free $E_2$(high) mode frequency, and $K_{\mathrm{RS}}^{E_2(\mathrm{high})}$ is the Raman biaxial stress conversion factor.

A strain-free reference sample plays a crucial role in the determination of the magnitude of biaxial stress in the GaN layers by Raman spectroscopy. Reference samples used in literature vary, for example, undoped 300-μm-thick GaN grown on

a sapphire substrate by chloride-hydride vapor phase epitaxy [12], 25-μm-thick GaN on a Si substrate [29], the edge of the surface of an undoped ~4.6-μm-thick epifilm grown by metalorganic chemical vapor deposition on a sapphire substrate [24], GaN nanorods grown epitaxially on Si(111) substrates with a GaN buffer layer [39], single detached GaN nanocolumn samples on Si(111) and sapphire (0001) substrates [40], 400-μm-thick free-standing GaN [28, 41], and homoepitaxially grown GaN films by molecular beam epitaxy [16]. Also, details of initial spectrometer calibration conditions are rarely elucidated. Therefore, it is not surprising that there is a large scatter in published reference room temperature Raman $E_2$(high) mode peak positions. To illustrate, 566.2 $cm^{-1}$ [16], 567.3 $cm^{-1}$ [24], 567.5 $cm^{-1}$ [28, 41], 567.6 $cm^{-1}$ [15], 567.7 $cm^{-1}$ [39], and 568 $cm^{-1}$ [11]. The $E_2$(high) response of the GaN templates under study are plotted in Fig. 17.8 together with the corresponding biaxial stresses lying in the samples obtained from XRD measurements. The strain-free $E_2$(high) mode frequency $\omega_0 = 568.208 \pm 0.007$ $cm^{-1}$ and the Raman biaxial stress conversion factor $K_{RS}^{E_2(high)} = -3.145 \pm 0.021$ $cm^{-1}$/GPa were determined from the intersection with the $y$-axis and the slope of the plot in Fig. 17.8, respectively, based on the strain-free lattice parameters, elastic stiffness constants, and initial calibration conditions adopted in this study. The Raman response of the bulk GaN sample was $\omega = 567.402 \pm 0.002$ $cm^{-1}$ which indicates expansion (not purely biaxial) relative to the hypothetical strain-free crystal or that the crystal is chemically altered with different bond strengths where there is oxygen. This value for bulk GaN agrees well with results (567.5 $cm^{-1}$) in [41] and [28] even though their initial calibration conditions were not clearly explained. The unintentional doping effect, that is, increased free (n-type) carrier concentration, which is responsible for the dilatation, is also revealed by its $A_1$(LO) Raman signature where a blueshift in frequency ($\omega = 735.016 \pm 0.007$ $cm^{-1}$) and broadening occurs with respect to the expected strain-free peak. Previous work [42] has shown that biaxial stress of 1 GPa shifts the $E_2$(high) mode by 2.91 $cm^{-1}$ through loading of a single epilayer stack (GaN/6H-SiC) using a four-point bending stage where GaN was assumed to be elastically isotropic, and a Poisson's ratio of $\nu = 0.183$ and Young's modulus $\hat{E} = 388$ GPa were employed. Applying the same elastic assumptions to our XRD data, the biaxial stress conversion factor resulted in $K_{RS}^{E_2(high)} = -2.900 \pm 0.019$ $cm^{-1}$/GPa. Our value ($K_{RS}^{E_2(high)} = -3.145 \pm 0.021$ $cm^{-1}$/GPa) also shows satisfactory agreement considering the uncertainties of the experimental results with the value ($K_{RS}^{E_2(high)} = -2.7 \pm 0.3$ $cm^{-1}$/GPa) in [12] where identical elastic stiffness constants were adopted. Their biaxial stress conversion factor was mentioned to be reliable by comparing with ab initio calculations [18]. Another published value for $K_{RS}^{E_2(high)}$ was $-2.9 \pm 0.3$ $cm^{-1}$/GPa [43] which was later, by the same authors [44], recalculated to be $-3.19$ using same elastic stiffness constants employed in our study. They remeasured three GaN on Si(111) samples with various thicknesses using Raman and XRD resulting in $K_{RS}^{E_2(high)} = -2.43 \pm 0.6$ $cm^{-1}$/GPa. Both results agree well with our biaxial stress conversion factor. Some other reported values are $-3.4 \pm 0.3$ $cm^{-1}$/GPa [29], $-4.2 \pm 0.3$ $cm^{-1}$/GPa [16], and $-4.3$ $cm^{-1}$/GPa [28],

etc., where the main reason for deviation is inconsistency in employing elastic stiffness coefficients. For example, the elastic constants adopted in our study yields $\hat{E} = 356 \pm 46$ GPa under an elastically isotropic assumption, while the Young's modulus used in [16] is $\hat{E} = 290$ GPa [35]. When elastic stiffness constants used in this study are applied to their data, the biaxial stress conversion factor turns out to be $-2.9$ cm$^{-1}$/GPa. In [45], the biaxial strain dependence of the $E_2$(high) mode was investigated using room temperature Raman and XRD measurements on MOVPE-grown GaN films on sapphire and 6H-SiC substrates. If the elastic stiffness constants used in our study are applied to their data, a biaxial conversion factor of $-2.824$ cm$^{-1}$/GPa is obtained which agrees well with our value.

Conducting a similar approach illustrated above, the relationship between biaxial stress in GaN and the $A_1$(LO) Raman shift was investigated and is displayed in Fig. 17.8. The strain-free mode frequency and the Raman biaxial stress conversion factor were determined to be $\omega_0 = 734.022 \pm 0.007$ cm$^{-1}$ and $K_{\mathrm{RS}}^{A_1(\mathrm{LO})} = -1.947 \pm 0.014$ cm$^{-1}$/GPa, respectively. Since the longitudinal optical polar phonon frequency is affected by free carrier concentration or different doping levels, a single GaN/6H-SiC template was individually tested under uniaxial bending on the basal plane with a setup using a pivot pin and press screws similar to that described in [46] to confirm the aforementioned results. The $E_2$(high) and $A_1$(LO) modes were measured simultaneously. The stress in the GaN layer was estimated from the $E_2$(high) mode frequency and its corresponding biaxial stress conversion factor assuming that the stress conversion factors for uniaxial and biaxial loading on the basal plane differ by a factor of 2. The relationship between stresses in the layers (0.78–1.5 GPa, uniaxial) and the $A_1$(LO) mode frequency shifts resulted in $\omega_0 = 733.899 \pm 0.016$ cm$^{-1}$ and $K_{\mathrm{RS}}^{\mathrm{A1(LO)}} = -1.923 \pm 0.028$ cm$^{-1}$/GPa . The nice agreement between results from two different methods provides confidence in the analyzed results. Again, the derived biaxial stress conversion factor shows an excellent agreement with the theoretically calculated value $K_{\mathrm{RS}}^{A_1(\mathrm{LO})} = -1.91$ c m$^{-1}$/GPa using the same elastic stiffness constants [18]. The value in [43] $K_{\mathrm{RS}}^{A_1(\mathrm{LO})} = 0.8 \pm 0.4$ cm$^{-1}$/GPa suffers from a large imprecision and was derived using elastic stiffness constants which were pointed out that their usage should be avoided [18]. Again, recalculation made by the same authors [44] adopting the same elastic stiffness constants used in our study resulted into $K_{\mathrm{RS}}^{A_1(\mathrm{LO})} = -0.88$ cm$^{-1}$/GPa where they explained high unintentional n-type doping, and low carrier mobility in the GaN layers investigated were responsible for obtaining the erroneous value in their previous study [44]. By measuring again three GaN on Si(111) samples, they deduced $K_{\mathrm{RS}}^{A_1(\mathrm{LO})} = -1.91 \pm 0.58$ cm$^{-1}$/GPa . The wurtzite structure of GaN consists of alternating hexagonal close-packed (0001) Ga and N planes. Each Ga (N) atom is equidistant to its four nearest N (Ga) atoms which form a tetrahedron. This structure is depicted in Fig. 17.9. As seen in Fig. 17.4, the $A_1$(LO) phonon mode corresponds to atomic oscillations along the $c$-axis. Tensile in-plane biaxial stress causes an increase in the a-lattice constant. In this case, the three bond lengths ($L_1$) nearly perpendicular to the $c$-axis of the four bonds in the tetrahedron shown in Fig. 17.9 and the $\alpha$-bond

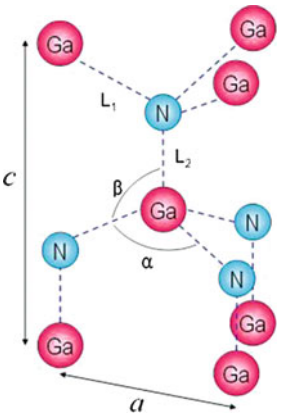



**Fig. 17.9** Tetrahedral atomic bonding arrangement within wurtzite GaN

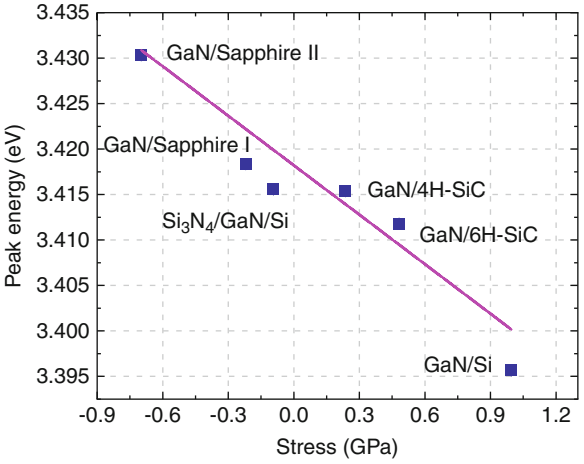



**Fig. 17.10** Room temperature luminescent band gaps of GaN epilayers versus corresponding residual stresses obtained by XRD measurements. The *solid line* is the fitting result

angles increase as well. This effect of force constant softening is revealed by the redshift in the $E_2$(high) mode frequency. Counterintuitively, the $A_1$(LO) mode frequency also shifts to a lower value when tensile biaxial stress is present. This phenomena is understood based on the fact that the decrease of the $c$-lattice constant ensuing from tensile biaxial stress is accompanied by an increase in the bond length ($L_2$) along the $c$-axis since the bond angle $\beta$ decreases [18, 19].

Figure 17.10 shows a set of room temperature PL spectra for the six samples analyzed by XRD. The residual stress in each GaN film reveals itself by the shift in the energy band gap which is represented by the PL peak position. A linear dependence of energy band gap on biaxial stress is observed.

$$E - E_0 = K_{\mathrm{PL}}\sigma_a \tag{17.12}$$

Here $E$ is the measured PL peak, $E_0$ is the strain-free PL peak, and $K_{\mathrm{PL}}$ is the PL biaxial stress conversion factor. $E_0$ and $K_{\mathrm{PL}}$ are found to be $3.41822 \pm 0.00002$ eV and $-0.01813 \pm 0.00004$ eV/GPa, respectively.

In [17], it is shown that a biaxial stress of 1 GPa shifts the band gap by $0.0211 \pm 0.0032$ eV. However, the biaxial stresses in their GaN templates correlated with each PL measurement have been calculated based on diverse Raman biaxial stress conversion factors in literature depending on the substrate type of each sample. The strain-free value of the energy band gap of GaN at room temperature has been reported as $E_0 = 3.4285$ eV [17]. However, the reference value in [17] was estimated based on a Raman reference value of 568 $cm^{-1}$ without details of initial calibration described. Moreover, their laser (325 nm) output power used for PL was 40 mW which may induce significant local heating on GaN. These may have caused the discrepancy between their reference PL peak and ours ($E_0 = 3.41822 \pm 0.00002$ eV). The laser power incident on the samples in our study was 0.003 mW to minimize laser induced local heating. Even at this excitation power, minimal heating was unavoidable. When $E_0$ and $K_{PL}$ derived here were used with a laser power of 0.3 uW (10% of the laser power used in this work), an offset of stress of ~50 MPa (compressive) would result. However, using such a low laser power (0.3 uW) is impracticable in terms of device measurement due to prolonged accumulation time (more than 20 min for 1 measurement) which makes the experiment more susceptible to drift in measurement location. As the additional shift of the PL peak induced by local heating is rather a constant value, this portion is included in $E_0$; thus, using the derived $E_0$ and $K_{PL}$ together with the suggested laser power (0.003 mW) can still be regarded as appropriate. However, one should take precautions that this condition may not hold for stress measurement on devices under voltage bias since the level of carrier induced heating would vary with bias conditions. A study [24] combining room temperature cathodoluminescence (CL) and Raman spectroscopy and another study [11] employed 3.42 eV as a strain-free reference at room temperature, but this was not based on rigorous experiments. The PL biaxial stress conversion factor $K_{PL} = -0.020 \pm 0.003$ eV/GPa reported in [12] where XRD, Raman, and low temperature PL measurements were correlated meets with our value within the error-bar ranges. References [16] and [47] report $K_{PL} = -0.027 \pm 0.002$ eV/GPa and $K_{PL} = -0.023$ eV/GPa, respectively, using a Young's modulus of $\hat{E} = 290 \pm 20$ GPa under isotropic assumptions relating stress determined by room temperature XRD to low temperature PL measurements. Their Young's moduli are mainly responsible for higher values of biaxial stress conversion factors. When recalculation is made using the elastic stiffness constants adopted in this study, $K_{PL}$ for [16] becomes $-0.019$ eV/GPa. In [44], the authors made correction on a reported value of $K_{PL} = 0.0155$ eV/GPa by employing the same elastic constants used in our study. As a consequence, $K_{PL} = 0.0175$ eV/GPa was obtained.

It should be noted that Raman and PL measure local stress (beam diameter of ~1 μm), whereas XRD measures stress over a larger area (0.2 mm × 0.2 mm). As mentioned before, PL measures the surface of GaN (penetration depth is 80–90 nm), whereas Raman spectroscopy and XRD probe a depth of more than several microns, thus measuring the average strain throughout the whole of the layer [13]. A significant lateral or vertical strain gradient within the GaN layer will cause adverse effects on the correspondence of results obtained by each technique.

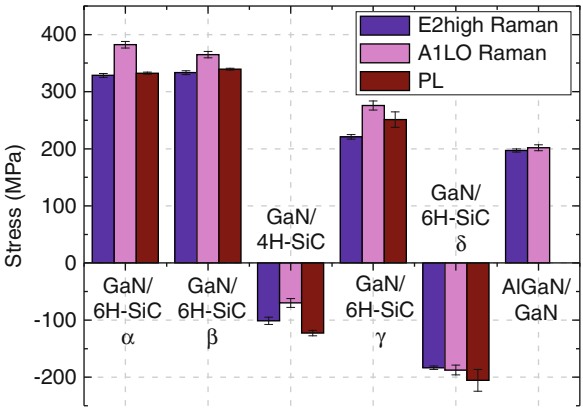



**Fig. 17.11** Biaxial stress of arbitrary GaN samples measured by Raman spectroscopy and PL based on parameters derived in this study. PL measurements were not done on the AlGaN/GaN sample due to low PL efficiency caused by a passivation layer

Reciprocal space mapping measurements indicated that strain present in the GaN films used in this fundamental study to correlate XRD, Raman spectroscopy, and PL were not perfectly uniform. Samples with poor quality were excluded in the analysis. Despite these limitations, a reasonable agreement is observed between stress results examined by Raman spectroscopy and PL on GaN templates which were not included in the foregoing analysis. Figure 17.11 shows the correspondence of biaxial stresses measured on five different GaN samples utilizing the Raman $E_2$(high), $A_1$(LO) mode, and PL peak shifts. The linear fit shown in Fig. 17.5 intersects the point corresponding to strain-free lattice constants. Moreover, the calculated dimensionless parameters such as $R^B$, $v$ as well as the derived biaxial stress conversion factors show consistency with acceptable literature values obtained by theoretical calculations or experimental investigation using individual techniques or a combination of two or more of them. The obtained values for strain-free reference phonon frequencies and luminescent band gap are logically explainable. Thus, stress in thin GaN epifilms under biaxial stress can be estimated with confidence by utilizing the aforementioned stress conversion factors and strain-free reference values for phonon frequencies and luminescent energy band gap at room temperature.

## *17.2.4 Application: Analysis of the Residual Stress Distribution in AlGaN/GaN HEMTs*

### 17.2.4.1 Motivation

Passivation and metallization structures can also impose strain in the AlGaN layer in AlGaN/GaN HEMTs. It has been shown that deposition and annealing of the ohmic contacts can induce tensile strain in the AlGaN layer [48]. Stress induced by the gate metal film, which forms a Schottky contact, can have an impact on device

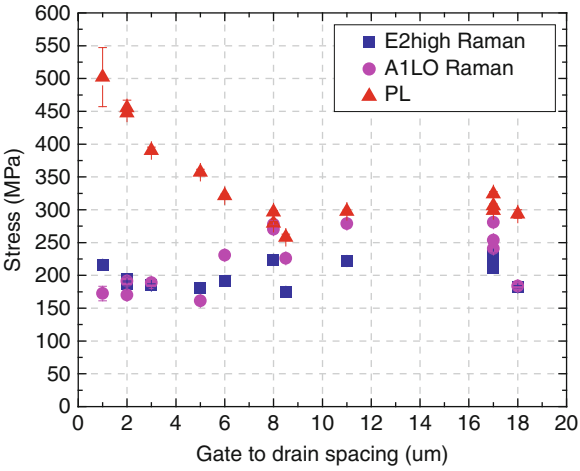



**Fig. 17.12** Residual stress at the center point between the gate and drain of devices with various gate-to-drain spacings measured by Raman spectroscopy and PL

performance and reliability. It has been reported for MOSFETs that there exists stress induced by the gate metal electrode which acts as a controlling factor for device performance [49], and for GaAs MESFETs, the additional strain and piezoelectric charge dipoles at the edges of the gate metallization has been modeled [50]. These local-induced strains in the AlGaN layer at the edges of the metal electrodes may have an impact on device performance by influencing the 2-DEG local sheet charge density. Furthermore, if there exist a significant amount of induced strain at the drain end of the gate which adds up the elastic energy in the AlGaN barrier at this region, a variation in critical voltage bias conditions to degrade the device under pinch-off or on-state conditions is expected [4].

To investigate the residual stress distribution in the conductive channel of AlGaN/GaN HFETs at room temperature, comparative analysis by means of micro-Raman spectroscopy and PL was done. Stress introduced by metallization structures is discussed.

#### 17.2.4.2 Results and Discussion of Residual Stress Distributions

With the framework described in Sect. 17.2.3, the stress profiles in electronic devices can be understood. The residual stress (of the GaN layer) at the center points between the gate and drain of 14 AlGaN/GaN HEMTs from the same subreticle with different gate-to-drain channel spacings were examined. The gate-to-drain spacings ranged from 1 to 18 μm, and the substrate used in these devices was 6H-SiC. Results are displayed in Fig. 17.12. As the gate-to-drain spacing decreases below 5 μm, a tensile shift of the stress near the heterointerface or GaN layer surface is observed whereas the average value of stress through the entire thickness of the GaN layer remains rather constant.

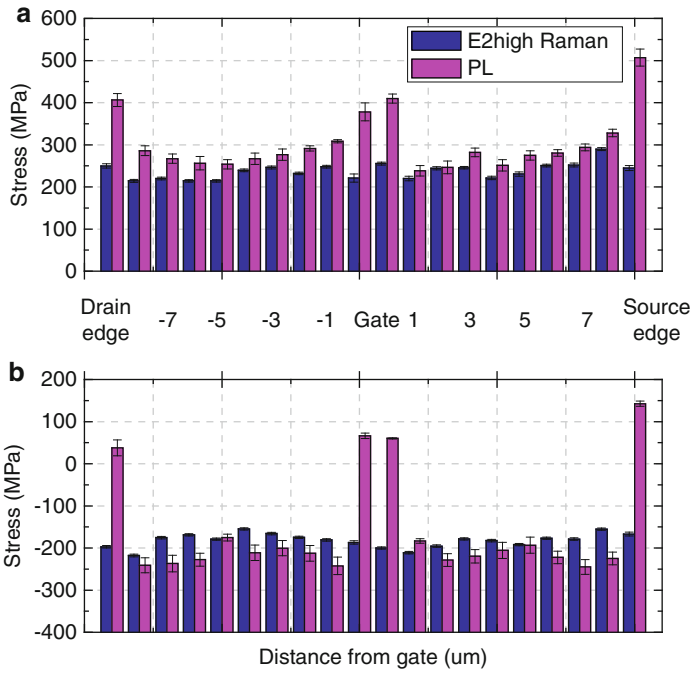



**Fig. 17.13** Residual stress distribution across the conductive channel in an AlGaN/GaN HEMTs measured by Raman spectroscopy and PL (**a**) and distance from gate (**b**)

To assure that the stress shift is induced near the patterned structures, residual stress distributions across the conductive channels of two types of non-field-plated devices having same structure (gate-to-drain and source-to-gate spacing of 9 μm) but fabricated on different GaN wafers (GaN/6H-SiC) were investigated. Raman spectroscopy and PL measurements were made at intervals of 1 μm between the source and drain metallizations. In Fig. 17.13, it is observed that for both device types, close to the metallizations, the residual stress of GaN near the heterointerface exhibits a significant tensile shift compared to the bulk average value. The residual stresses built in each device type exhibit an opposite sign. Across the conductive channel, the bulk average stress remains rather unchanged. Therefore, we conclude that the stress offset in Fig. 17.8 occurs due to metallization structures. This shift in residual stress near patterned structures may originate from both the stress in ohmic and Schottky gate electrodes [48, 49] and the stress concentration at the edge of discontinuities of the passivation overlayer as it basically delivers a force load to the edges of the film or edges of the metal electrodes [8].

A slight deviation between the Raman and PL measured stresses away from the metallization structures is observed (Fig. 17.13). In spite of this offset, the Raman and PL measured stresses tend to fall close into the error-bar range of each other. This small offset is believed to be caused by experimental errors.

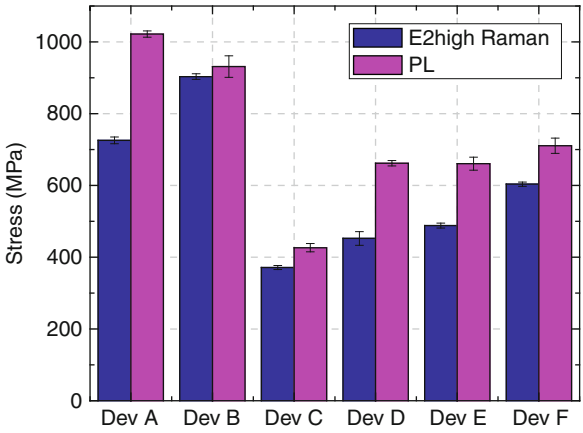



**Fig. 17.14** Tensile shift in residual stress near the AlGaN/GaN heterointerface caused by metallization structures in AlGaN/GaN/Si devices

Similar results of tensile residual stress shifts in the GaN layer close to the heterointerface compared its bulk average value near metallization structures were also observable in AlGaN/GaN HEMTs built on Si (Fig. 17.14). The center points between their gate-to-drain openings were measured where these devices had gate-to-drain spacings of 0.5–1 μm. Furthermore, the substrate-induced residual stress is more highly tensile in GaN/Si devices compared to GaN/6H-SiC devices discussed above.

From the discussions done so far, it is obvious that the gate, source, and drain metallization patterns induce tensile stress in the GaN nearest the surface, and there may be extension also in the thin AlGaN layer deposited over the GaN layer. Although the induced stress near the metallization patterns is most likely approximately uniaxial (but probably nonuniform because the ohmic metallizations on AlGaN/GaN are typically morphologically nonuniform), the Raman and PL signatures manifest that an induced tensile strain exists. As the gate pattern-induced tensile strain increases the elastic energy built in the AlGaN barrier in addition to the inverse piezoelectric contribution under operating conditions [4], the potential for defect formation may significantly increase. Also, stress-related reliability concerns in GaN/Si devices may be more pronounced compared to GaN/SiC devices as the sum of substrate and the gate metal-induced tensile stresses near gate edge is much higher than the case for GaN/6H-SiC devices.

### *17.2.5 Summary and Conclusions*

From XRD measurements, biaxial residual stresses present in various GaN films were determined using lattice constants for unstrained GaN and elastic stiffness constants from literature. Room temperature Raman and PL measurements on same samples allowed correlating biaxial stress with the $E_2$(high) and $A_1$(LO) phonon frequency shifts as well as energy band gap of GaN. It has been observed that

biaxial stress has a linear relationship with the Raman $E_2$(high), $A_1$(LO) mode, and the luminescent band gap of GaN. This linear relationship holds for the modest and realistic strain range considered in this study. At much higher strain levels, that is, ranges of strain of about two orders of magnitude higher, nonlinear behavior of change in band gap energy with strain as well as structural phase transition is predicted by density functional theory simulations [51]. Derived Raman and PL biaxial stress conversion factors ($K_{\mathrm{RS}}^{E_2(\mathrm{high})} = -3.145 \pm 0.021\ \mathrm{cm}^{-1}/\mathrm{GPa}$, $K_{\mathrm{RS}}^{A_1(\mathrm{LO})} = -1.947 \pm 0.014\ \mathrm{cm}^{-1}/\mathrm{GPa}$, $K_{\mathrm{PL}} = -0.01813 \pm 0.00004\ \mathrm{eV}$) showed consistency with reliable values reported in literature. Strain-free reference $E_2$(high) and $A_1$(LO) phonon frequencies as well as luminescent band gap ($\omega_{0,E_2(\mathrm{high})} = 568.208 \pm 0.007\ \mathrm{cm}^{-1}$, $\omega_{0,A_1(\mathrm{LO})} = 734.022 \pm 0.007\ \mathrm{cm}^{-1}$, $E_0 = 3.41822 \pm 0.00002\ \mathrm{eV}$) were obtained under explicit initial spectrometer calibration conditions (4H-SiC, 776.599 $\mathrm{cm}^{-1}$) and assumptions for elastic behavior of the GaN crystal.

In the AlGaN/GaN HEMTs observed, a tensile strain was induced on the AlGaN layer near the edge of metallization structures. Under operating conditions, the tensile strain induced on the AlGaN barrier at the drain side gate edge adds up to the built-in residual stress as well as the inverse piezoelectric stress. Such increase in elastic energy built in the AlGaN barrier will have an adverse effect on device reliability by leading to formation of crystallographic defects which may result in performance degradation such as current collapse and eventually device failure. This means metallization schemes that result in an induction of compressive strain in the AlGaN layer or proper passivation overlayer deposition that results in less tensile strain induced at its edge will be beneficial in terms of device reliability. Based on the framework presented in this study, Raman spectroscopy and PL can provide efficient means of process feedback in terms of stress monitoring between fabrication steps in the test and development stage of GaN devices.

## 17.3 Raman Thermometry

### *17.3.1 Overview*

As AlGaN/GaN devices operate at exceptionally high power densities (e.g., over 30 W/mm [6]), the overall power is large, and as a consequence, considerable heat generation in the active region becomes a crucial issue in terms of device performance and reliability. Considerable self-heating negatively impacts transport properties. The electron mobility and drift velocity reduce with increasing temperature mainly due to enhanced polar optical phonon scattering that in turn decreases the device output current and operational frequency (i.e., a decrease in DC and RF performance). Current saturation occurs at an applied electric field that is much lower than the electric field required (~200 kV/cm) for electron drift velocity to saturate in bulk GaN [52]. The self-heating effect is also observable in the *I–V*

characteristics, that is, negative DC output conductance at large $V_{ds}$ conditions. Excess high junction temperature has potential to degrade the stability of gate Schottky and ohmic contacts as well as form material defects, thus causing device failure or reliability problems.

Under normal operating conditions, a peak electric field is concentrated over a small region in the vicinity of the gate edge near the AlGaN/GaN interface [53]. This large electric field spike can lead to breakdown even for relatively small values of $V_{ds}$. An extremely high electric field can occur in conventional structures which do not employ field plates or recessed gates. The energetic (or hot) electrons generated by the high electric field lose energy to the lattice primarily by emitting longitudinal optical (LO) phonons via Fröhlich interaction ("a Coulomb interaction between electrons and the longitudinal electric field produced by the LO phonons [54]") due to the strong ionic nature of GaN. The group velocity of an optical phonon is near zero; thus, the emitted LO phonons remain in the channel until they either are reabsorbed by electrons or decay into transverse optical (TO) phonons and acoustic phonons. Acoustic phonons play a role to effectively transport the Joule heat away from the active region of the device towards the substrate. In the 2-DEG, the LO phonon decay time or lifetime (~350 fs) considerably exceeds the LO phonon emission time by a hot electron (~10 fs) [55]. Thus, at the drain end of the gate electrode where the lateral electric field is highest, a localized accumulation of nonequilibrium LO phonons or so-called hot phonons takes place which acts as a bottleneck in thermal energy transport. At this local region, a nanoscale "hot spot" (diameter of 10–100 nm [56]) in lattice temperature forms as a result of the combination of electric field spike, complex phonon interactions, and ballistic transport effects [57]. Conventional models that utilize Fourier's law that assume heat generation as a uniform source along a fixed area near the gate are inappropriate to estimate the channel temperature for devices under these conditions. This fact reflects needs for more advanced modeling schemes or experimental methods that estimate the channel temperature with reasonable accuracy.

### *17.3.2 Theory and Methodology of Raman Thermometry*

Among methods to examine operating temperature in GaN devices, micro-Raman spectroscopy is a noncontact, local, nondestructive, and fast optical material characterization technique adequate for monitoring temperature in AlGaN/GaN HEMTs with a suitable spatial resolution of ~1 μm. Micro-Raman spectroscopy probes the energy associated with Raman active vibrational modes. The zone-center ($\Gamma$ point, $k \approx 0$) phonons are responsible for the first-order Raman scattering due to the energy and momentum conservation rules in the light scattering process. In general, a temperature rise in GaN results in a decrease in phonon frequency (except the anomalous behavior of the $E_2$(low) mode) and increase in the phonon linewidth. Anharmonic effects are responsible for these phenomena. The downshift in phonon frequency at high temperatures originates from thermal expansion of the

lattice and phonon self-energy suppression (decay of phonons by dissipating the phonon energy to lower frequency modes such that the energy and momentum are conserved) [58]. Here the dominant contribution comes from the thermal expansion. The increase in interaction among optical phonons at high temperature results in increased phonon scattering and thus decreased phonon lifetime. The phonon linewidth $\Gamma$ ($cm^{-1}$) is related with the phonon lifetime $\tau$ via the energy–time uncertainty relation $\Gamma/\hbar = 1/\tau$ where $\hbar = 5.3 \times 10^{-12}$ $cm^{-1}$ s [59]. Therefore, rise in temperature accompanies phonon peak broadening. The typical method to measure temperature using Raman spectroscopy is to utilize the Stokes Raman peak shift. However, since phonon frequency is also sensitive to the lattice strain, the accuracy of this technique is subject to errors arising from stresses that stem from thermoelastic and inverse piezoelectric effects that develop during device operation. The Stokes to anti-Stokes intensity ratio provides means to measure temperature independent of stress. The ratio of anti-Stokes to Stokes intensities increases with temperature since a larger fraction of molecules are in the first vibrationally excited state rather than the ground state (refer to Fig. 17.2). However, in practice, it is not easy to implement this method due to intensive acquisition time and requirement of a holographic notch filter which is of high cost and has limited lifetime. Former work has shown that utilizing the change in linewidth of the $E_2$(high) phonon mode of GaN between the Raman spectra of the powered state and the pinch-off reference state provides means to perform stress-independent Raman thermometry for AlGaN/GaN HEMTs under operation with the speed of the Stokes peak position measurement [60].

The temperature dependence of first-order Raman scattering of hexagonal GaN was investigated by conducting Raman measurements on single crystalline GaN films grown on sapphire substrates. Downshifts of phonon frequency and broadening was observed for five Raman active phonon modes: $E_2$(high), $A_1$(LO), $A_1$(TO), $E_1$(LO), and $E_1$(TO). These phenomena were explained by adding up the contributions from thermal expansion, three- and four-phonon decay processes [58]. A similar study was conducted on undoped single crystalline bulk AlN where the temperature dependence of the $A_1$(LO) and $E_2$(high) phonon modes were discussed [61]. The measurement of operating temperature of AlGaN/GaN HEMTs using Raman spectroscopy was first performed in [62] utilizing the temperature dependent $E_2$(high) phonon frequency shift. Raman temperature measurements utilizing ultraviolet and visible laser excitations on operating AlGaN/GaN HEMTs based on the $E_2$(high) phonon frequency shift were conducted. This allowed demonstration of a temperature depth profile since UV Raman probes the temperature near the AlGaN/GaN interface whereas visible Raman sees the average temperature throughout the GaN layer and the substrate [63]. It was shown that peak shift based Raman temperature measurement on operating AlGaN/GaN HEMTs result in an underprediction of temperature due to complex stresses which arise during device operation from both the thermoelastic and inverse piezoelectric effects. In accord with this, a Raman technique capable of removing errors in AlGaN/GaN HEMT thermometry arising from the stress dependency of the Raman peak position shift was developed [60]. Electrothermal

simulations on AlGaN/GaN HEMTs grown on different substrates (SiC, Si, and sapphire) were performed and compared with experimental *I–V* characteristics as well as Raman temperature measurement results. Difference in current saturation levels for each structure was explained in terms of the self-heating effect [52].

### *17.3.3 Application: The Impact of Bias Conditions on Self-heating of AlGaN/GaN HEMTs*

#### 17.3.3.1 Motivation

It was reported that bias conditions change the ratio of power dissipated in the pinch-off region to that dissipated in the rest of the device thus having an influence on device peak temperature [64]. Utilizing the change in linewidth of the $E_2$(high) phonon mode of GaN between the Raman spectra of the powered state and the pinch-off reference state was shown to provide means to perform stress-independent Raman thermometry for AlGaN/GaN HEMTs under operation [60]. This method was implemented in this study to extract temperature information. The device channel temperature was examined via micro-Raman spectroscopy under various bias conditions (combination of $V_{gs}$ and $V_{ds}$) where power dissipation levels were identical.

#### 17.3.3.2 Experimental Method

Micro-Raman spectroscopy was performed through the use of Horiba Jobin Yvon LabRAM HR800. A 532-nm diode-pumped solid-state laser was used as the excitation source. A 1,800 groove/mm diffraction grating was used, and the confocal hole size was adjusted to 100 μm. This setup was established to obtain high enough spectral resolution suited for temperature measurements while having a reasonable acquisition time and spectral range for each measurement. Experiments were carried out with the laser light perpendicular to the basal plane of GaN in a 180° backscattering configuration and unpolarized detection. The $E_2$(high) mode was investigated among the allowed zone-center optical phonon modes detectable for this configuration, namely, two $E_2$(TO) and the $A_1$(LO) modes [55]. The 532-nm laser power incident on the devices was adjusted via neutral density filters to ~25.2 and ~2.4 mW. The sub-band gap laser wavelength (532 nm) used for Raman measurements prevents localized heating of GaN by laser light absorption. For devices under electrical bias, negligible carrier induced local heating at this setup was confirmed by comparing the Raman response acquired with the aforementioned excitation power levels. A 50× LWD objective (NA = 0.5) was employed to focus the laser on the sample surface and collect the Raman signatures. The theoretical diffraction-limited laser focused spot size (diameter) was 1.3 μm using

the relation $d = 1.22\lambda/\text{NA}$ where $d$ is the spot size, $\lambda$ is the laser wavelength, and NA is the numerical aperture. Before conducting Raman measurements, the system was calibrated at 25°C with a 4H-SiC piece to 776.599 $\text{cm}^{-1}$ which was equivalent to 520.46 $\text{cm}^{-1}$ for high quality single crystal silicon. Pinch-off state measurements were conducted before and after every on-state temperature measurement to take into account of the drift (systematic error) of the spectrometer. In order to determine the peak position and full width at half maximum (FWHM), Raman spectra were fitted using a linear combination of Gaussian–Lorentzian line-shapes. Data acquisition times were adjusted to obtain consistent Raman intensity counts (~8,000).

Six-finger non-field-plated AlGaN/GaN HEMTs fabricated on 6H-SiC with a gate width of 370 μm and gate-to-drain channel spacing of ~3.5 μm were examined. The thickness of the GaN layer was ~2 μm.

Raman measurements were conducted 1 μm away from the drain side gate edge at the midpoint along the third channel from the device edge. Output powers were controlled to 2, 4, and 6 W that correspond to 0.9, 1.8, and 2.7 W/mm, respectively, at various $V_{ds}$ conditions (5, 7, 10, 15, 20, 28, 48, and 64 V) by adjusting $V_{gs}$. The device base temperature under the package was maintained at 60°C (±0.15°C) while mounted on a heating stage (INSTEC HCC324S). To calibrate device Raman response, that is, peak shift and line broadening, to temperature rise, Raman Stokes signals were collected (20 spectra for each condition) while the device was isothermally heated without voltage applied to known temperatures (25.0, 45.8, 89.0, 135.0, 181.0, 229.5, and 278°C) using a LINKAM TS1500 temperature controlled stage with a temperature stability of ±1°C. Then the actual device temperature measurements for each bias condition were carried out by conducting a total of 44 measurements: 14 measurements at a pinch-off condition ($V_{gs} = -8$ V) for a given $V_{ds}$ to be investigated, then 16 measurements for that on-state condition at that $V_{ds}$, followed by 14 more measurements at the previous pinch-off condition. This pinch-off condition is chosen because (especially at higher $V_{ds}$) it is electrically similar to the corresponding on-state condition but with observed negligible power dissipation. Briefly, this method is discussed in detail elsewhere [60] but involves separating the voltage induced changes in the device that affect Raman from the thermal changes we want to measure; it is not sufficient to simply measure the change in the Raman signal at bias and ascribe all of the change to heating. Best estimates and uncertainties (95% confidence bounds) for temperatures were obtained by statistical analysis.

#### 17.3.3.3 Results for the Impact of Bias Conditions on Temperature Rise

Under negative gate bias, a depletion region is formed beneath the gate and the sheet carrier density reaches a minimum at the drain side gate edge which in turn causes an electric field concentration. Thus, power delivery by the electrons to the lattice takes place primarily in this narrow region of the channel. Among the bias conditions (combination of $V_{gs}$ and $V_{ds}$) with identical power dissipation, a lower carrier concentration is achieved for bias conditions with more negative $V_{gs}$, that is,

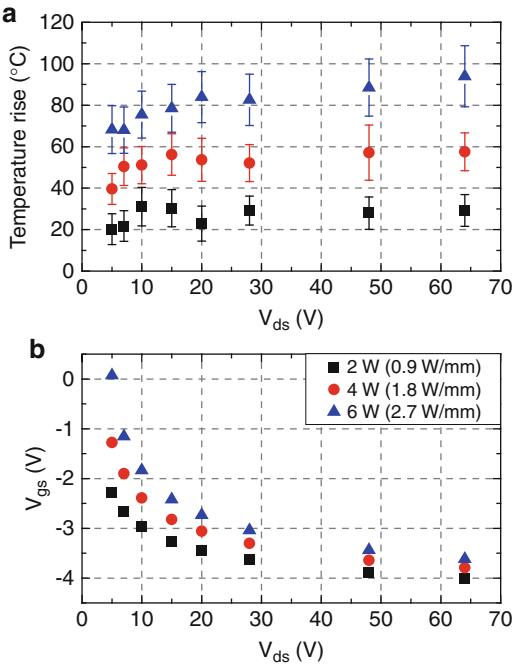



**Fig. 17.15** (**a**) Temperature measurement results. Power dissipation levels were maintained at 2, 4, and 6 W which correspond to 0.9, 1.8, and 2.7 W/mm, respectively. (**b**) Tested electrical bias conditions. To retain a certain power dissipation level, $V_{gs}$ is adjusted to less negative values when $V_{ds}$ is decreased. $V_{gs}$ is positive for $V_{ds} = 5$ V when 6 W is dissipated

higher $V_{ds}$, since the depletion region extends further down towards the GaN layer. The requirement for current continuity results in a higher longitudinal electric field and higher carrier velocity at the gate edge near the drain. Therefore, the distribution of Joule heating within the device varies with bias conditions even though the amount of power dissipation is identical. Under high $V_{ds}$ conditions (more negative $V_{gs}$), a higher portion of the applied $V_{ds}$ drops near the drain side gate edge with remainder of the voltage dropping across the rest of the 2-DEG and at source and drain ohmic contacts. The concentration of the heat generation region results in a higher channel temperature as shown in Fig. 17.15. As the channel is opened by applying a less negative gate bias, the voltage drop becomes more uniform across the channel. Under conditions where the channel is fully opened (positive gate bias), there is not a strong electric field region beneath the gate electrode which results in a lower channel temperature. The results for the foregoing discussion are shown in Fig. 17.15.

Figure 17.16 shows the systematic error that comes from not properly separating the effect of mechanical stress on the Raman peak position shift to acquire temperature information. Compressive thermoelastic stress causes a blueshift in the Raman frequency, whereas increased temperature redshifts it. This results in an underprediction of lattice temperature. The $E_2$(high) mode of GaN corresponds to the atomic oscillation in the $c$-plane which has a nonpolar character. Its mode frequency is therefore more sensitive to the lattice strain in the $c$-plane than the $A_1$(LO) phonon mode which corresponds to atomic oscillations along the

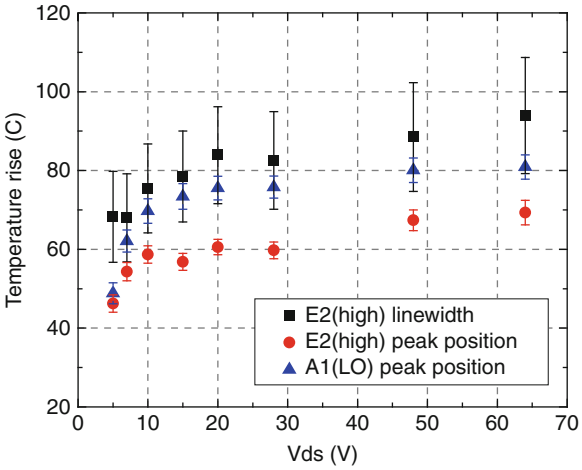



**Fig. 17.16** Temperature rise at 2.7 W/mm power dissipation acquired from Raman measurements. The Stokes linewidth (*black squares*)-based temperatures represent the true lattice temperatures. Considerable underprediction is shown for Stokes peak position-based temperatures due to their stress dependence

c-axis [13]. A graphical illustration of these phonon modes is given in Fig. 17.4. Figure 17.16 shows the channel temperature extracted by the $A_1$(LO) peak position is closer to the "true" channel temperature due to its lower sensitivity to mechanical stress.

The peak lattice temperatures are expected to be much higher than values obtained by (visible) Raman spectroscopy since the average temperature throughout the depth of the GaN layer across a lateral region with ~1 μm diameter is probed, whereas the hot spot dimension is on the order of 10–100 nm [56]. For this same reason, we also believe we have measured a lower bound for the magnitude of the bias dependence of the thermal resistance. This may promote localized damage to occur near the drain side gate edge depending on bias conditions even though thermal aging tests do not show thermal degradation occur at ambient temperatures close to operational junction temperatures predicted by conventional modeling results utilizing Fourier's law, assuming the heat generation region as a uniform source along a fixed area near the gate. This bias-dependent self-heating effect should be even more pronounced for devices grown on top of Sapphire or Si substrates due to their low thermal conductivities and for devices with longer channels. When setting up experimental and control groups for reliability test experiments to investigate certain failure mechanisms (e.g., see [65] where hot-electron-induced degradation was examined), the effect of bias-dependent self-heating should be taken into account, that is, one should not assume the channel temperature to be invariable among devices under different bias conditions although the amount of power dissipations are identical. Control of the electric field distribution across the channel can be performed through the implementation

of source connected field plates, T-gates, etc. The reduction in the peak electric field in these devices will change the distribution of the heat generation region and may reduce the effect seen here slightly but is not expected to eliminate it. In fact, additional field plating will make the effect we have observed harder to see and may impart a false sense of security when thermal resistance measurements are made without considering the possibility of bias dependence.

### *17.3.4 Summary and Conclusions*

Utilizing the change in linewidth of the $E_2$(high) phonon mode of GaN between the Raman spectra of the powered state and the pinch-off reference state provides means to perform stress-independent Raman thermometry for AlGaN/GaN HEMTs under operation with the speed of the Stokes peak position measurement.

The bias-dependent self-heating was examined from the analysis of experimental data acquired by micro-Raman spectroscopy. Under bias conditions where $V_{ds}$ is higher and $V_{gs}$ is more negative, the heat generation region becomes more concentrated at a nanoscale region at the drain end of the gate electrode. Consequently, a higher channel temperature results compared to that for conditions with lower $V_{ds}$ and less negative $V_{gs}$ even though the total power dissipation in the device is identical. This effect exerts local peculiarity in the device and should be taken into account in fundamental reliability studies and implementation of AlGaN/GaN HEMTs.

## References

1. O. Ambacher, J. Smart, J.R. Shealy, N.G. Weimann, K. Chu, M. Murphy, W.J. Schaff, L.F. Eastman, R. Dimitrov, L. Wittmer, M. Stutzmann, W. Rieger, J. Hilsenbeck, Two-dimensional electron gases induced by spontaneous and piezoelectric polarization charges in N- and Ga-face AlGaN/GaN heterostructures. J. Appl. Phys. **85**, 3222–3233 (1999)
2. T. Palacios, J.W. Chung, O. Saadat, F. Mieville, GaN and digital electronics: a way out of Moore's law? Phys. Status Solidi C: Curr. Top. Solid State Phys. **6**(6), 1361–1364 (2009)
3. G. Meneghesso, G. Verzellesi, F. Danesin, F. Rampazzo, F. Zanon, A. Tazzoli, M. Meneghini, E. Zanoni, Reliability of GaN high-electron-mobility transistors: state of the art and perspectives. IEEE Trans. Device Mater. Reliab. **8**, 332–343 (2008)
4. J.A. del Alamo, J. Joh, GaN HEMT reliability. Microelectron. Reliab. **49**, 1200–1206 (2009)
5. J.R. Ferraro, K. Nakamoto, *Introductory Raman Spectroscopy*, 2nd edn. (Academic, London, 2003)
6. R. Trew, G. Bilbro, W. Kuang, Y. Liu, H. Yin, Microwave AlGaN/GaN HFETs. IEEE Microw. Mag. **6**, 56–66 (2005)
7. Y.F. Zhang, I.P. Smorchkova, C.R. Elsass, S. Keller, J.P. Ibbetson, S. Denbaars, U.K. Mishra, J. Singh, Charge control and mobility in AlGaN/GaN transistors: experimental and theoretical studies. J. Appl. Phys. **87**, 7981–7987 (2000)
8. J. Bernat, P. Javorka, M. Marso, P. Kordos, Conductivity and hall effect measurements on intentionally undoped and doped AlGaN/GaN heterostructures before and after passivation. Appl. Phys. Lett. **83**, 5455–5457 (2003)

9. M. Azize, T. Palacios, Effect of substrate-induced strain in the transport properties of AlGaN/GaN heterostructures. J. Appl. Phys. **108**, 2157–2164 (2010)
10. A. Hushur, M.H. Manghnani, J. Narayan, Raman studies of GaN/sapphire thin film heterostructures. J. Appl. Phys. **106**, 054317 (2009)
11. W. Rieger, T. Metzger, H. Angerer, R. Dimitrov, O. Ambacher, M. Stutzmann, Influence of substrate-induced biaxial compressive stress on the optical properties of thin GaN films. Appl. Phys. Lett. **68**, 970–972 (1996)
12. V.Y. Davydov, N.S. Averkiev, I.N. Goncharuk, D.K. Nelson, I.P. Nikitina, A.S. Polkovnikov, A.N. Smirnov, M.A. Jacobsen, O.K. Semchinova, Raman and photoluminescence studies of biaxial strain in GaN epitaxial layers grown on 6H-SiC. J. Appl. Phys. **82**, 5097–5102 (1997)
13. K. Funato, S. Hashimoto, K. Yanashima, F. Nakamura, M. Ikeda, Residual strain dependence of optical characteristics in GaN layers grown on (0001) sapphire substrates. Appl. Phys. Lett. **75**, 1137–1139 (1999)
14. I.H. Lee, I.H. Choi, C.R. Lee, S.K. Noh, Evolution of stress relaxation and yellow luminescence in GaN/sapphire by Si incorporation. Appl. Phys. Lett. **71**, 1359–1361 (1997)
15. H. Park, C. Park, S. Yeo, S. Kang, M. Mastro, O. Kryliouk, T. Anderson, Epitaxial strain energy measurements of GaN on sapphire by Raman spectroscopy. Phys. Status Solidi (c) **2**, 2446–2449 (2005)
16. C. Kisielowski, J. Kruger, S. Ruvimov, T. Suski, J.W. Ager 3rd, E. Jones, Z. Liliental-Weber, M. Rubin, E.R. Weber, M.D. Bremser, R.F. Davis, Strain-related phenomena in GaN thin films. Phys. Rev. B: Condens. Matter **54**, 17745–17753 (1996)
17. D.G. Zhao, S.J. Xu, M.H. Xie, S.Y. Tong, H. Yang, Stress and its effect on optical properties of GaN epilayers grown on Si(111), 6H-SiC(0001), and c-plane sapphire. Appl. Phys. Lett. **83**, 677–679 (2003)
18. J.M. Wagner, F. Bechstedt, Properties of strained wurtzite GaN and AlN: Ab initio studies. Phys. Rev. B **66**, 115202 (2002)
19. Y.F. Li, B. Yao, Y.M. Lu, Y.Q. Gai, C.X. Cong, Z.Z. Zhang, D.X. Zhao, J.Y. Zhang, B.H. Li, D.Z. Shen, X.W. Fan, Z.K. Tang, Biaxial stress-dependent optical band gap, crystalline, and electronic structure in wurtzite ZnO: experimental and ab initio study. J. Appl. Phys. **104**, 083517 (2008)
20. A. Sarua, H.F. Ji, M. Kuball, M.J. Uren, T. Martin, K.J. Nash, K.P. Hilton, R.S. Balmer, Piezoelectric strain in AlGaN/GaN heterostructure field-effect transistors under bias. Appl. Phys. Lett. **88**, 123–126 (2006)
21. T. Beechem, A. Christensen, D.S. Green, S. Graham, Assessment of stress contributions in GaN high electron mobility transistors of differing substrates using Raman spectroscopy. J. Appl. Phys. **106**, 327–333 (2009)
22. I. DeWolf, Micro-Raman spectroscopy to study local mechanical stress in silicon integrated circuits. Semicond. Sci. Technol. **11**, 139–154 (1996)
23. Z. Feng, Micro-Raman scattering and microphotoluminescence of GaN thin films grown on sapphire by metal-organic chemical vapor deposition. Opt. Eng. **41**, 2022 (2002)
24. F.C. Wang, C.L. Cheng, Y.F. Chen, C.F. Huang, C.C. Yang, Residual thermal strain in thick GaN epifilms revealed by cross-sectional Raman scattering and cathodoluminescence spectra. Semicond. Sci. Technol. **22**, 896–899 (2007)
25. E.D. Haberer, C.H. Chen, M. Hansen, S. Keller, S.P. DenBaars, U.K. Mishra, E.L. Hu, Enhanced diffusion as a mechanism for ion-induced damage propagation in GaN. J. Vac. Sci. Technol. B **19**, 603–608 (2001)
26. J.M. Hwang, J.T. Hsieh, H.L. Hwang, O. Breitschadel, H. Schweizer, Defect depth profiling using photoluminescence and cathodoluminescence spectroscopy: the role of oxygen on reactive ion beam etching of GaN in O-2/Ar plasmas. Appl. Surf. Sci. **175**, 450–455 (2001)
27. D.G. Zhao, D.S. Jiang, J.J. Zhu, Z.S. Liu, S.M. Zhang, H. Yang, J.W. Liang, The influence of V/III ratio in the initial growth stage on the properties of GaN epilayer deposited on low temperature AlN buffer layer. J. Cryst. Growth **303**, 414–418 (2007)

28. S. Tripathy, S.J. Chua, P. Chen, Z.L. Miao, Micro-Raman investigation of strain in GaN and AlxGa1-xN/GaN heterostructures grown on Si(111). J. Appl. Phys. **92**, 3503–3510 (2002)
29. I. Ahmad, M. Holtz, N.N. Faleev, H. Temkin, Dependence of the stress-temperature coefficient on dislocation density in epitaxial GaN grown on alpha-Al2O3 and 6H-SiC substrates. J. Appl. Phys. **95**, 1692–1697 (2004)
30. K. Barghout, J. Chaudhuri, Calculation of residual thermal stress in GaN epitaxial layers grown on technologically important substrates. J. Mater. Sci. **39**, 5817–5823 (2004)
31. J. Pernot, E. Bustarret, M. Rudzinski, P.R. Hageman, P.K. Larsen, Strain relaxation in GaN grown on vicinal 4H-SiC(0001) substrates. J. Appl. Phys. **101**, 154–160 (2007)
32. C.G. Van de Walle, Effects of impurities on the lattice parameters of GaN. Phys. Rev. B **68**, 3409–3412 (2003)
33. M.A. Moram, M.E. Vickers, X-ray diffraction of III-nitrides. Rep. Prog. Phys. **72**, 457–462 (2009)
34. M. Leszczynski, H. Teisseyre, T. Suski, I. Grzegory, M. Bockowski, J. Jun, S. Porowski, K. Pakula, J.M. Baranowski, C.T. Foxon, T.S. Cheng, Lattice parameters of gallium nitride. Appl. Phys. Lett. **69**, 73–75 (1996)
35. A. Polian, M. Grimsditch, I. Grzegory, Elastic constants of gallium nitride. J. Appl. Phys. **79**, 3343–3344 (1996)
36. A.F. Wright, Elastic properties of zinc-blende and wurtzite AlN, GaN, and InN. J. Appl. Phys. **82**, 2833–2839 (1997)
37. M.A. Moram, Z.H. Barber, C.J. Humphreys, Accurate experimental determination of the Poisson's ratio of GaN using high-resolution x-ray diffraction. J. Appl. Phys. **102**, 658–665 (2007)
38. S.J. Xu, W. Liu, M.F. Li, Effect of temperature on longitudinal optical phonon-assisted exciton luminescence in heteroepitaxial GaN layer. Appl. Phys. Lett. **77**, 3376–3378 (2000)
39. H. Seo, Q. Chen, M. Iliev, L. Tu, C. Hsiao, J. Mean, W. Chu, Epitaxial GaN nanorods free from strain and luminescent defects. Appl. Phys. Lett. **88**, 153124 (2006)
40. N. Thillosen, K. Sebald, H. Hardtdegen, R. Meijers, R. Calarco, S. Montanari, N. Kaluza, J. Gutowski, H. Luth, The state of strain in single GaN nanocolumns as derived from micro-photoluminescence measurements. Nano Lett. **6**, 704–8 (2006)
41. M. Jamil, J. Grandusky, V. Jindal, F. Shahedipour-Sandvik, S. Guha, M. Arif, Development of strain reduced GaN on Si (111) by substrate engineering. Appl. Phys. Lett. **87**, 082103 (2005)
42. T. Beechem III, Metrology of gan electronics using micro-Raman spectroscopy. Solid State Electron **56**, 209–213 (2008)
43. F. Demangeot, J. Frandon, M.A. Renucci, O. Briot, B. Gil, R.L. Aulombard, Raman determination of phonon deformation potentials in alpha-GaN. Solid State Commun. **100**, 207–210 (1996)
44. F. Demangeot, J. Frandon, P. Baules, F. Natali, F. Semond, J. Massies, Phonon deformation potentials in hexagonal GaN. Phys. Rev. B **69**, 231–235 (2004)
45. M. Klose, N. Wieser, G.C. Rohr, R. Dassow, F. Scholz, J. Off, Strain investigations of wurtzite GaN by Raman phonon diagnostics with photoluminescence supplement. J. Cryst. Growth **190**, 634–638 (1998)
46. J. Kim, K. Baik, C. Park, S. Cho, S.J. Pearton, F. Ren, Measurement of external stress on bulk GaN. Phys. Status Solidi a-Appl. Mater. Sci. **203**, 2393–2396 (2006)
47. W.G. Perry, T. Zheleva, M.D. Bremser, R.F. Davis, W. Shan, J.J. Song, Correlation of biaxial strains, bound exciton energies, and defect microstructures in GaN films grown on AlN/6H-SiC(0001) substrates. J. Electron. Mater. **26**, 224–231 (1997)
48. A. Soltani, A. BenMoussa, S. Touati, V. Hoel, J.C. De Jaeger, J. Laureyns, Y. Cordier, C. Marhic, M.A. Djouadi, C. Dua, Development and analysis of low resistance ohmic contact to n-AlGaN/GaN HEMT. Diamond Relat. Mater. **16**, 262–266 (2007)
49. T. Matsuki, N. Mise, S. Inumiya, T. Eimori, Y. Nara, Impact of gate metal-induced stress on performance modulation in gate-last metal-oxide-semiconductor field-effect transistors. Jpn. J. Appl. Phys. Part 1-Regul. Pap. Brief Commun. Rev. Pap. **46**, 3181–3184 (2007)

50. T. Onodera, H. Nishi, Theoretical-study of the piezoelectric effect on Gaas-Mesfets on (100), (011), and (111 bar) Ga, and (111) as substrates. IEEE Trans. Electron Dev. **36**, 1580–1585 (1989)
51. S.P. Alpay, L. Dong, S.K. Yadav, R. Ramprasad, Band gap tuning in GaN through equibiaxial in-plane strains. Appl. Phys. Lett. **96**, 54–61 (2010)
52. B. Benbakhti, A. Soltani, K. Kalna, M. Rousseau, J.C. De Jaeger, Effects of self-heating on performance degradation in AlGaN/GaN-based devices. IEEE Trans. Electron Dev. **56**, 2178–2185 (2009)
53. S. Karmalkar, U.K. Mishra, Enhancement of breakdown voltage in AlGaN/GaN high electron mobility transistors using a field plate. IEEE Trans. Electron Dev. **48**, 1515–1521 (2001)
54. X.B. Zhang, T. Taliercio, S. Kolliakos, P. Lefebvre, Influence of electron-phonon interaction on the optical properties of III nitride semiconductors. J. Phys. Condens. Matter **13**, 7053–7074 (2001)
55. A. Matulionis, Hot phonons in GaN channels for HEMTs. Phys. Status Solidi a-Appl. Mater. Sci. **203**, 2313–2325 (2006)
56. S. Sridharan, A. Venkatachalam, P. Yoder, Electrothermal analysis of AlGaN/GaN high electron mobility transistors. J. Comput. Electron. **7**, 236–239 (2008)
57. A. Christensen, S. Graham, Multiscale modeling of hot spots in Gan high electron mobility transistors, in *Ipack 2009: Proceedings of the Asme Inter Pack Conference 2009*, Boston, MA, vol. 1, pp. 535–541 (2010)
58. W.S. Li, Z.X. Shen, Z.C. Feng, S.J. Chua, Temperature dependence of Raman scattering in hexagonal gallium nitride films. J. Appl. Phys. **87**, 3332–3337 (2000)
59. M. Kuball, J.M. Hayes, Y. Shi, J.H. Edgar, Phonon lifetimes in bulk AlN and their temperature dependence. Appl. Phys. Lett. **77**, 1958–1960 (2000)
60. T. Beechem, A. Christensen, S. Graham, D. Green, Micro-Raman thermometry in the presence of complex stresses in GaN devices. J. Appl. Phys. **103**, 332–336 (2008)
61. M. Kuball, J.M. Hayes, Y. Shi, J.H. Edgar, A.D. Prins, N.W.A. van Uden, D.J. Dunstan, Raman scattering studies on single-crystalline bulk AlN: temperature and pressure dependence of the AlN phonon modes. J. Cryst. Growth **231**, 391–396 (2001)
62. M. Kuball, J.M. Hayes, M.J. Uren, T. Martin, J.C.H. Birbeck, R.S. Balmer, B.T. Hughes, Measurement of temperature in active high-power AlGaN/GaN HFETs using Raman spectroscopy. IEEE Electron Device Lett. **23**, 7–9 (2002)
63. I. Ahmad, V. Kasisomayajula, D.Y. Song, L. Tian, J.M. Berg, M. Holtz, Self-heating in a GaN based heterostructure field effect transistor: ultraviolet and visible Raman measurements and simulations. J. Appl. Phys. **100**, 92–97 (2006)
64. E.R. Heller, A. Crespo, Electro-thermal modeling of multifinger AlGaN/GaN HEMT device operation including thermal substrate effects. Microelectron. Reliab. **48**, 45–50 (2008)
65. H. Kim, A. Vertiatchikh, R.M. Thompson, V. Tilak, T.R. Prunty, J.R. Shealy, L.F. Eastman, Hot electron induced degradation of undoped AlGaN/GaN HFETs. Microelectron. Reliab. **43**, 823–827 (2003)

# Chapter 18
# Reliability Study of InP-Based HBTs Operating at High Current Density

**Yoshino K. Fukai and Kenji Kurishima**

**Abstract** The high-speed and low-power operation of InP-based heterojunction bipolar transistors (HBTs) makes them very attractive for use in optical-communication ICs operating at 40 Gbit/s and more. An important requirement for InP HBTs is long-term stability in their electrical characteristics under high-current-density operation. This chapter describes the degradation behavior for devices operating at current densities of up to 10 mA/μm$^2$ under elevated ambient temperatures. The results of electrical measurements and microscopic analyses indicate that surface passivation in the external base, a refractory electrode on the emitter, and high crystal quality at the emitter–base junction are the keys to enhancing device reliability.

## 18.1 Introduction

High-speed electrical devices capable of terahertz-level performance are being developed with the aim of applying them to the next generation of communications systems and networks. Among several types of transistors, InP-based heterojunction bipolar transistors (HBTs) have been attracting a lot of attention for use in the ICs required for establishing over-40-Gbit/s optical communications systems. Owing to the good electron transport properties of InP and InGaAs, InP-based HBT technologies have already demonstrated ultrahigh-speed ICs that operate at above 100 GHz, even with moderate device scaling [1–3]. Additionally, InP HBTs have advantages for monolithic integration with long-wavelength optical devices in terms of material compatibility [4, 5].

Y.K. Fukai (✉) • K. Kurishima
NTT Photonics Laboratories, NTT Corporation, 3-1 Morinosato Wakamitya, Atsugi-shi, Kanagawa Pref. 243-0198, Japan
e-mail: fukai.yoshino@lab.ntt.co.jp

O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7_18,
© Springer Science+Business Media New York 2013

This chapter reviews the reliability studies of InP HBTs the authors have performed recently [6–10]. In general, it is necessary to increase the injection current density in order to boost the operation speed of InP HBTs [3]. Thus, long-term stability in electrical characteristics under high-current-density operation is ultimately required in order to apply these devices to practical high-speed ICs. In this work, accelerated life tests were conducted using stress current densities of up to 10 mA/μm$^2$. In the following section, we first briefly provide some background on InP-HBT reliability studies. Then, Sect. 18.3 explains the InP HBT structures used in the accelerated life tests. Sections 18.4, 18.5 and 18.6 discuss several kinds of degradation modes observed for the InP HBTs. Investigation of the underlying mechanisms for the degradation has revealed several ways to improve their reliability and prolong their lifetime. Among them, ledge structures with a SiN passivation film are prerequisites for suppressing the recombination current in the external-base region (Sect. 18.4). An electrode with a refractory metal is useful in preventing the generation of defects in the emitter-contact layer triggered by electrode-metal diffusion (Sect. 18.5). Further improvement of crystal quality at the emitter–base junction is very important for enhancing the current density available for practical ultrahigh-speed operation (Sect. 18.6).

## 18.2 Background of Reliability Study

### *18.2.1 Early Works for Development of InP HBTs*

In the early stage of InP HBT development, reliability issues were similar to those for GaAs HBTs. The diffusion of p-type dopants, such as Be and Zn, during epitaxial growth, fabrication process, or accelerated life tests was the first serious problem to be solved. This diffusion shifts the position of emitter–base junction into the wide-bandgap emitter, which increases current turn-on voltage in DC characteristics. The degradation of crystal quality due to the dopant diffusion enhances recombination current in the junction vicinity and thereby decreases current gain. The dopant diffusion may also decrease current-gain cutoff frequency, $f_t$, and increase maximum oscillation frequency, $f_{max}$, in RF characteristics [11]. As in the case of GaAs HBTs [12], reliability performance of InP HBTs was improved by replacing the p-type dopant with carbon, which has a very low diffusion coefficient. Early works showed inactivation of carbon acceptors due to hydrogen passivation, especially when metal-organic vapor phase epitaxy (MOVPE) was used for the growth [13, 14]. However, this problem was overcome by employing post-growth annealing to reverse the hydrogen passivation [15] or by developing molecular beam epitaxy (MBE) techniques for the growth of InP and related materials. Recently, high-quality InP-HBT epitaxial wafers have become commercially available. The stability of ohmic contacts was also investigated in order to suppress the increase in device series resistance. These issues are summarized elsewhere [16].

### *18.2.2 Reliability Performance Reported to Date*

Reliability studies of InP HBTs with emitter width ranging from a few micrometers to submicrometer are summarized in Table 18.1. Since early devices employed a Be-doped InGaAs base, they normally showed severe dopant diffusion. The main failure mode was a variation of current turn-on voltage, for which the activation energy was reported to be 1.5–1.9 eV [17–19]. The degradation of the DC current gain, $\beta$, was also observed with different values of activation energy [18, 19]. Presumably, this is because the degradation in current gain resulted from several kinds of mechanisms, including the decrease in the emitter injection efficiency and the increase in the recombination current in the space charge region and at the semiconductor surface. Furthermore, some devices showed initial transient change in current gain within a few hundred hours of stress [18, 20].

As mentioned before, replacing the p-type dopant with carbon successfully solved the problem of dopant diffusion. However, the degradation in current gain with relatively low activation energy (0.97 eV) was still observed, which is associated with the increase in the base recombination current [21]. Since defects and impurities were likely introduced during the fabrication process, surface treatments and passivation techniques were widely investigated [22, 23]. Recently, the stability of the current gain of InP HBTs has been dramatically improved by introducing InP-ledge structures (or an emitter edge-thinning design) on the external-base layer [6, 24]. For these HBTs, activation energies were reported to be 1.5 eV [24] and 1.7 eV [6], and sufficiently long device lifetime ($>10^7$ h at 125°C) was extrapolated against the moderate stress current density of around 1 mA/μm$^2$.

Besides the degradation in current gain, failure in InP HBTs related to ohmic contacts has also been investigated. For example, the thermal stability of electrode metals has been examined as a function of annealing temperature [25].

## 18.3 InP HBT Structures and Accelerated Life Tests

This chapter evaluates the reliability of InP HBTs with a 1-μm-wide emitter and those with a 0.5-μm-wide one. The former, the first generation of InP HBTs, was developed for applications to 40-Gbit/s optical communications systems, while the latter, the second generation, was developed for 100-Gbit/s applications. Before discussing the degradation behavior of tested devices in the following sections, we briefly explain the HBT structures.

The HBT-layer structures were grown by the MBE method on a 3-in. InP substrate. The baseline HBT with a 1-μm-wide emitter consists of an InP subcollector doped with Si, a 300-nm-thick InGaAs collector lightly doped with Si, a 50-nm-thick InGaAs base doped with C to $4 \times 10^{19}$ cm$^{-3}$, a 70-nm-thick InP emitter unintentionally doped, a 20-nm-thick InP contact layer doped with Si to

**Table 18.1** Reliability studies reported to date

| Reference | | Ref. [17] | Ref. [18] | Ref. [19] | Ref. [20] |
|---|---|---|---|---|---|
| Growth and structure | Dopant | Be | Be | Be | Be |
| | Growth | GSMBE | GSMBE | GSMBE | MBE |
| | Ledge | | | | |
| | Emitter area | $3 \times 3\ \mu m^2$ | $2 \times 5\ \mu m^2$ | $1 \times 5\ \mu m^2$ | $1.6 \times 5\ \mu m^2$ |
| Degradation in $\beta$ and $V_{BE}$ | Initial change | | $\beta$ increase for ~500 h | | $\beta$ increase for ~200 h |
| | $E_a$ for degradation | $V_{BE}$: 1.57 eV | $\beta$: 1.1 eV<br>$V_{BE}$: 1.5 eV | $\beta$: 1.75 eV<br>$V_{BE}$: 1.86 eV | Stable during the tests |
| | Criterion | $\Delta V = +30$ mV | $\beta < 20$<br>$\Delta V/V_0 = +5\%$ | $\beta < 20$<br>$\Delta V/V_0 = +5\%$ | |

| Reference | | Ref. [21] | Ref. [24] | Ref. [6] | Ref. [8] |
|---|---|---|---|---|---|
| Growth and structure | Dopant | C | C | C | C |
| | Growth | | MBE | MBE | MBE |
| | Ledge | | ○ | ○ | ○ |
| | Emitter area | $1 \times 3\ \mu m^2$ | $1.4 \times 5.9\ \mu m^2$ | $1 \times 4\ \mu m^2$ | $0.5 \times 3\ \mu m^2$ |
| Degradation in $\beta$ | Initial change | About 10% drop in $\beta$ | | | |
| | $E_a$ for degradation | 0.97 eV | 1.5 eV | 1.7 eV | 1.8 eV |
| | Criterion | $\Delta\beta/\beta_0 = -20\%$ | $\Delta\beta/\beta_0 = -3\%$ | $\Delta\beta/\beta_0 = -5\%$,<br>$\Delta\beta/\beta_0 = -15\%$ | $\Delta\beta/\beta_0 = -15\%$ |

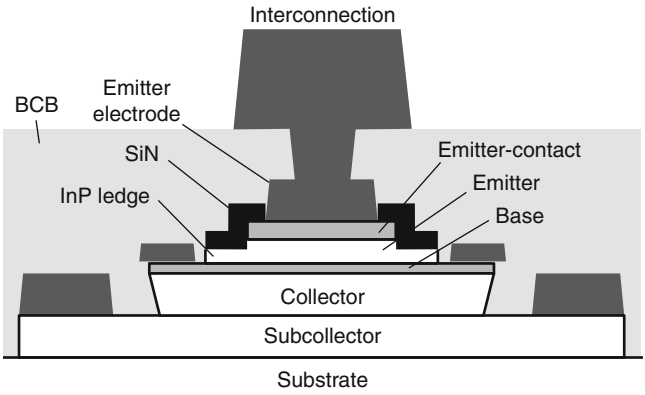



**Fig. 18.1** Cross-sectional view of the baseline HBT with a 1-μm-wide emitter used in the study

$2 \times 10^{19}$ cm$^{-3}$, and an InGaAs contact layer doped with Si to $3 \times 10^{19}$ cm$^{-3}$ [6, 9]. On the other hand, the HBT with a 0.5-μm-wide emitter has a 150-nm-thick InGaAs collector lightly doped with Si, a 35-nm-thick InGaAs base doped with C to $5 \times 10^{19}$ cm$^{-3}$, and a thinner (20–40 nm) InP emitter [7–10]. The collector current density required for achieving peak values of $f_t$ and $f_{max}$ depends on the collector thickness [3]. It is 1–2 mA/μm$^2$ for the 1-μm-emitter HBTs and 4–6 mA/μm$^2$ for the 0.5-μm-emitter HBTs.

The cross section of the baseline HBT with a 1-μm-wide emitter is schematically shown in Fig. 18.1. A triple-mesa structure was fabricated in a non-self-aligned manner using i-line lithography. After mesa formation, the entire mesa structure was spin-coated with benzocyclobutene (BCB). For the passivation of the external-base surface, the HBT has InP-ledge structures whose surface is covered by a SiN film. Typical distance between the emitter mesa and the base electrode is about 0.4 μm. The ledge width is 70–80% of the emitter-to-base spacing. The electrode-metal configurations are Ti(30 nm)/Pt(20 nm)/Au for the emitter and collector contacts and Pt/Ti/Pt/Au for the base contact. These electrodes were formed using a standard lift-off technique.

For the reliability study, accelerated life tests with electrical biases at an elevated ambient temperature were conducted under several conditions. DC bias voltages were continuously applied between the collector and emitter and between the base and emitter of each device so that collector current density, $J_C$, was fixed at 1, 2, 5, 7, or 10 mA/μm$^2$. The devices were mounted on ceramic packages, which were placed in a nitrogen-filled furnace heated at an ambient temperature, $T_a$, of 100–195°C. The bias-temperature (BT) stress was interrupted about ten times during a test in order to measure the current–voltage ($I$–$V$) characteristics for each device at room temperature. After the stress test, microscopic failure analyses were conducted using transmission electron microscopy (TEM), energy-dispersive X-ray spectroscopy (EDS), and transmission electron diffraction (TED) in order to investigate the mechanisms of the degradation observed.

## 18.4 Suppression of Surface Recombination in the External Base by Ledge Structures

Generally, the degradation related to the device surface is characterized by a rapid change in electrical characteristics with relatively low activation energy. Such degradation must first be minimized in order to investigate other degradation behaviors. In this section, it is shown that the ledge structures play an indispensable role in suppressing the surface-related degradation in the external base.

Figure 18.2 shows the two types of emitter mesa structures prepared for the HBTs with a 1-μm-wide emitter. One does not use any ledge passivation, and the other employs an InP-ledge layer with a SiN film. The thickness of the InP ledge is 40–50 nm. The HBT with the former mesa structure has the external-base surface exposed to BCB. On the other hand, the HBT with the latter one protects the external base with the thin InP layer whose surface is covered by SiN. Besides these two structures, an emitter mesa structure that has the InP-ledge layer without a SiN film was also prepared. In this structure, the InP-ledge surface is covered by BCB, not by SiN.

Figures 18.3 and 18.4 show typical time-wise change in the base current and in the current gain, $\beta$, for the HBTs, respectively. All of these devices were stressed with bias conditions of $V_{CE} = 2$ V and $J_C = 1$ mA/μm$^2$ at $T_a = 175$°C. Under the conditions, the junction temperature, $T_j$, was calculated to be 213°C. As seen from Figs. 18.3 and 18.4a, the HBTs without ledge passivation exhibit initial degradation of current gain (only within 20 h), associated with a rapid increase in base current. The current gain decreases gradually even after the initial degradation. On the other hand, the HBTs with ledge structures exhibit very little change in the current gain at the initial stage. Furthermore, the reduction of the current gain is only within 10% of the initial value, even after 10,000 h of stress. Also noteworthy is that the HBTs with ledge structures exhibit much higher initial values of $\beta$. These results clearly indicate that the HBTs without ledge passivation suffer from a rapid increase in surface-recombination

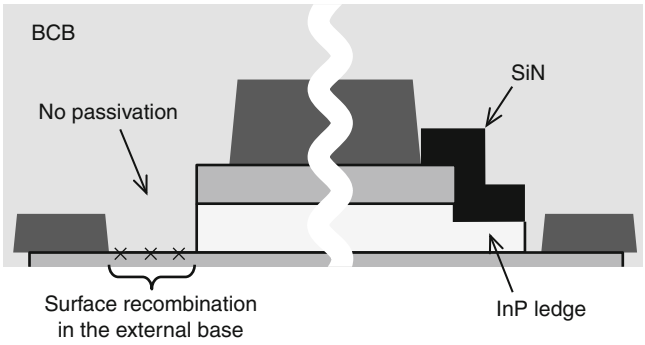



**Fig. 18.2** Cross-sectional view of the emitter mesa structures with no ledge passivation (*left*) and with an InP ledge and SiN passivation (*right*)

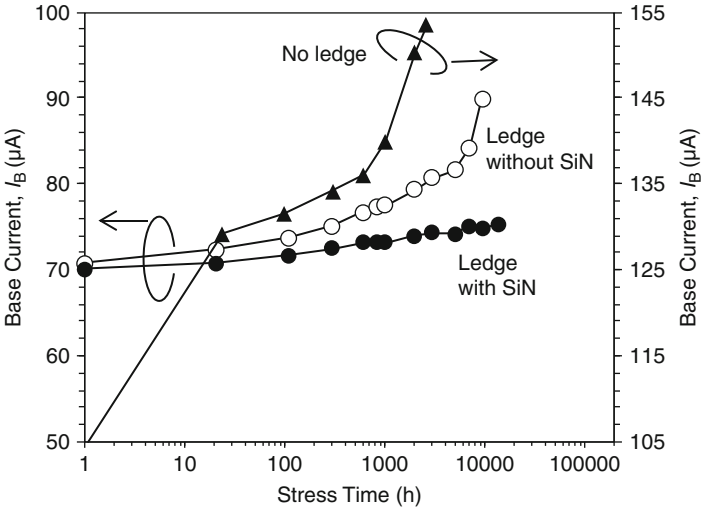



**Fig. 18.3** Time-wise change in base current, $I_B$, for the HBTs with and without ledge passivation. The HBTs were stressed at $J_C = 1$ mA/μm$^2$, $V_{CE} = 2$ V, and $T_a = 175°C$

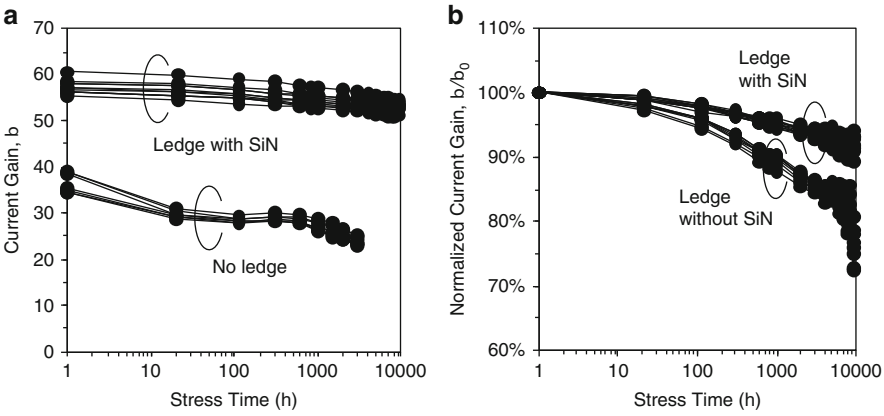



**Fig. 18.4** (**a**) Time-wise change in current gain, $\beta$, for the HBTs with no ledge passivation and for the HBTs with ledge and SiN passivation. (**b**) Time-wise change in normalized current gain, $\beta/\beta_0$, for the HBTs with and without a SiN film deposited on the ledge surface. The $\beta$ was measured at $J_C = 1$ mA/μm$^2$. The HBTs were stressed at $J_C = 1$ mA/μm$^2$, $V_{CE} = 2$ V, and $T_a = 175°C$. The junction temperature, $T_j$, is calculated to be 213°C

current in the external base. The ledge structures are prerequisites to achieving long-term stability in the electrical characteristics of InP HBTs.

As seen from Fig. 18.4b, the InP-ledge structures with a SiN film offer even better long-term stability than those without it. This result means that the SiN film is more effective than the BCB film in depleting the InP ledge layer and in further

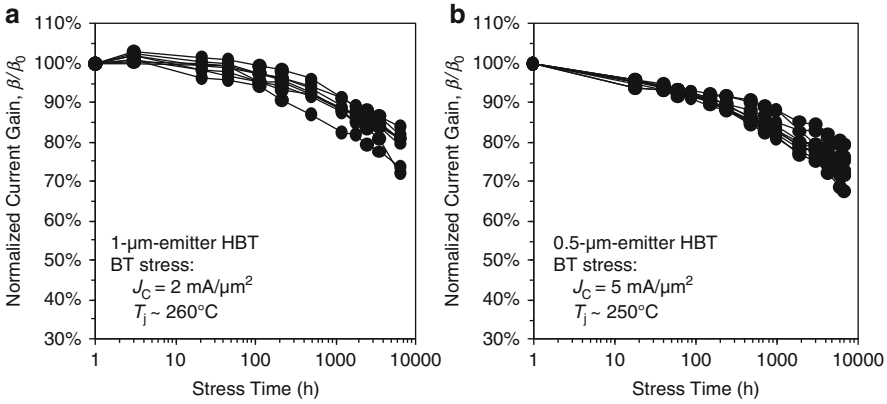



**Fig. 18.5** Time-wise change in normalized current gain, $\beta/\beta_0$, (**a**) for the 1-μm-emitter HBTs stressed at $J_C = 2$ mA/μm$^2$ and $T_j \sim 260$°C and (**b**) for the 0.5-μm-emitter HBTs stressed at $J_C = 5$ mA/μm$^2$ and $T_j \sim 250$°C. The $\beta$ was measured at $J_C = 1$ mA/μm$^2$

suppressing the ledge leakage current from the emitter mesa to the base electrode. In this chapter, all of the devices described hereafter use the InP-ledge structures covered with a SiN film.

For 0.5-μm-emitter HBTs operating at higher current density, a thinner (20–40 nm) InP emitter is used in order to fully deplete the ledge structures and thereby suppress the ledge leakage current more effectively [8]. Figure 18.5 compares the time-wise change in normalized current gain, $\beta/\beta_0$, for the 1-μm-emitter and 0.5-μm-emitter HBTs. The stress current density was 2 mA/μm$^2$ for the 1-μm-emitter HBTs, while it was 5 mA/μm$^2$ for the 0.5-μm-emitter HBTs. The junction temperatures were 260°C and 250°C, respectively. Apparently, the 0.5-μm-emitter HBTs also exhibit long-term stability in current gain even against much higher stress current density. The degradation behavior of $\beta$ is similar to that for the 1-μm-emitter HBTs. The Arrhenius plot in Fig. 18.6 indicates the temperature dependence of the median value of the critical time (i.e., mean time to failure, or MTTF) for each BT-stress condition. Here, critical time is defined as the time when the current gain decreases by 15% of the initial value. The activation energy for the 1-μm-emitter HBTs (stressed at $J_C = 1$–$2$ mA/μm$^2$) is estimated to be 1.7 eV. The activation energy for the 0.5-μm-emitter HBTs is nearly the same as that for the 1-μm-emitter HBTs, even under higher current-density operation (5 mA/μm$^2$).

To confirm the independence of the surface-related degradation of electrical bias, we also conducted high-temperature (HT) storage tests using the 1-μm-emitter HBTs. In these tests, the degradation of current gain was also observed, and the plot of the MTTF was found to be on the same line in Fig. 18.6. Therefore, the degradation is simply due to the thermal effect. Presumably, long-term thermal stress causes some degradation of crystal quality at the surface of the ledge structures, which increases the ledge leakage current from the emitter mesa to the base electrode.

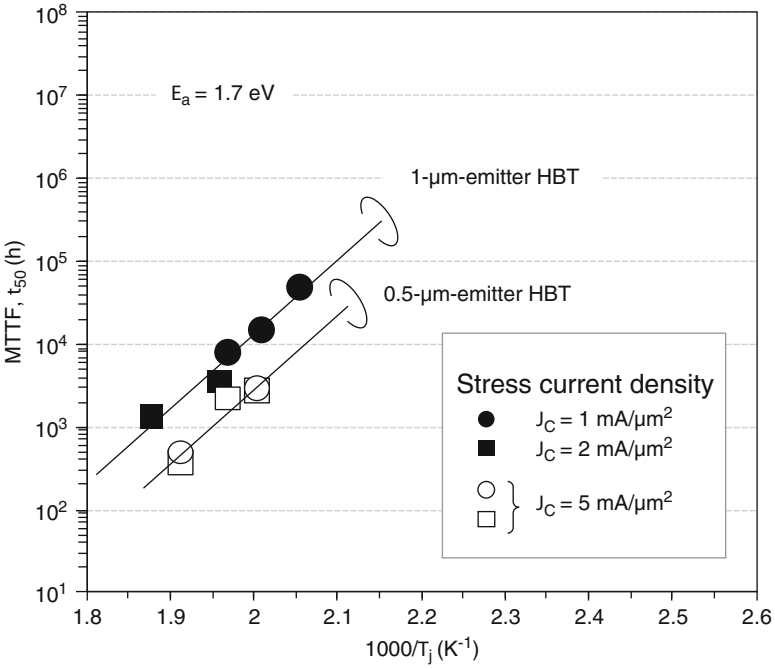



**Fig. 18.6** Arrhenius plot of MTTF for current gain, $\beta$. Failure criterion is defined as a 15% reduction in $\beta$ from the initial value. The activation energy, $E_a$, is estimated to be 1.7 eV

## 18.5 Degradation Triggered by Emitter-Electrode-Metal Diffusion and Its Suppression by a Refractory Metal

This section investigates the stability of the emitter-contact characteristics for the conventional Ti/Pt/Au electrode and newly developed electrodes with a refractory metal, such as Mo and W. It is shown that severe degradation of crystal quality in the emitter-contact layer takes place for the HBTs with the conventional electrode. Microscopic analyses by TEM, EDS, and TED were performed in order to understand the underlying mechanism. The electrodes with a refractory metal are shown to be very effective in suppressing such degradation.

### *18.5.1 Electrode-Metal Configurations*

The metal configurations for the electrodes with a refractory metal are shown in Fig. 18.7. The metal stacks are Ti(6 nm)/Mo(10 nm)-Ti/Pt/Au, Ti(6 nm)/W (150 nm)-Ti/Pt/Au, and W(150 nm)-Ti/Pt/Au. A thin Ti layer is inserted between

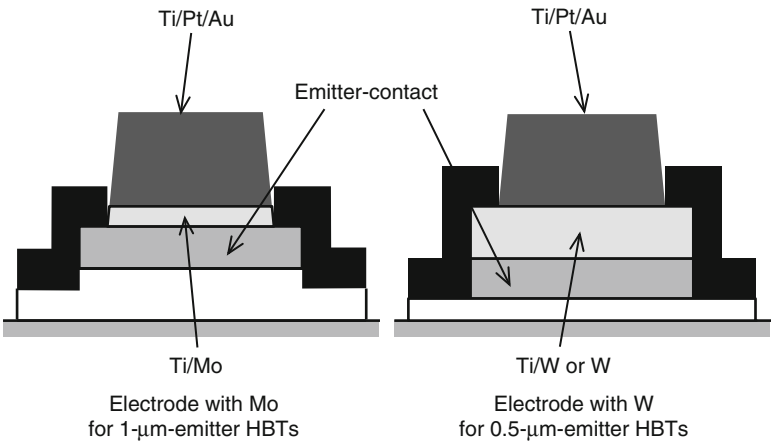



**Fig. 18.7** Cross-sectional view of the electrodes with Mo (*left*) and with W (*right*). The refractory metal of Ti(6 nm)/Mo(10 nm), Ti(6 nm)/W(150 nm), or W(150 nm) is inserted between the semiconductor layer and the Ti/Pt/Au electrode. The electrodes with Mo and W were applied to the 1-μm-emitter HBTs and 0.5-μm-emitter HBTs, respectively

the semiconductor layer and the refractory metal in order to obtain good adhesion and to keep the contact resistance at a minimum. The refractory metals are employed to block the undesirable influence from the Ti/Pt/Au electrode.

The conventional electrode and the new electrode with Mo were applied to the HBTs with a 1-μm-wide emitter. These electrodes are formed by electron beam evaporation after emitter mesa formation. The change in the metal configuration does not cause any notable influence in DC and RF performance of fabricated devices. For example, HBTs with the conventional electrode exhibit an average $\beta$ of 45, $f_t$ of 173 GHz, and $f_{max}$ of 263 GHz at $J_C = 1$ mA/μm$^2$ across a 3-in. wafer. On the other hand, HBTs with the refractory electrode show an average $\beta$ of 47, $f_t$ of 175 GHz, and $f_{max}$ of 275 GHz.

The electrodes with W were applied to the HBTs with a 0.5-μm-wide emitter operating at higher current density. The fabrication process starts with the formation of a Ti/W or W film on the wafer, where W is deposited by sputtering. The metal film and the semiconductor contact layer are then dry-etched successively to form emitter mesas. The Ti/Pt/Au electrodes are formed on the dry-etched metal film by evaporation. Fabricated HBTs with the Ti/W-Ti/Pt/Au electrode exhibit an average $\beta$ of 58, $f_t$ of 321 GHz, and $f_{max}$ of 301 GHz at $J_C = 4$ mA/μm$^2$ across a 3-in. wafer. On the other hand, HBTs with the W-Ti/Pt/Au electrode show a higher emitter series resistance because of the absence of the Ti contact layer. However, these HBTs also offer sufficiently high $f_t$ of over 300 GHz, which is acceptable for practical applications. Table 18.2 summarizes the electrode configurations, emitter area, and HBT performance.

**Table 18.2** Emitter electrodes, emitter area, and HBT performance

| Electrode metal | Ti/Pt/Au (conventional) | Ti/Mo-Ti/Pt/Au (refractory) | Ti/W-Ti/Pt/Au (refractory) | W-Ti/Pt/Au (refractory) |
|---|---|---|---|---|
| Metal deposition | Electron beam evaporation | | Sputtering for W, evaporation for others | |
| Emitter area | $1 \times 4\ \mu m^2$ | | $0.5 \times 3\ \mu m^2$ | |
| $\beta$, $f_t$, and $f_{max}$ | $\beta \sim 50$, $f_t$ ~175 GHz, $f_{max} \sim 270$ GHz | | $\beta \sim 60$, $f_t > 300$ GHz, $f_{max} \sim 300$ GHz | |
| $J_C$ for peak $f_t$, $f_{max}$ | $J_{C,peak} = 1-2\ mA/\mu m^2$ | | $J_{C,peak} = 4–6\ mA/\mu m^2$ | |

## *18.5.2 TEM Observation*

### 18.5.2.1 TEM Images of HBTs with a Ti/Pt/Au Electrode After BT Tests

Figure 18.8 shows cross-sectional TEM images for the 1-μm-emitter HBTs with the conventional electrode. The BT-stress conditions for the HBT shown in Fig. 18.8a were $J_C = 5\ mA/\mu m^2$ and $T_j = 250°C$ for 1,576 h, while those for the HBTs in Fig. 18.8b were $J_C = 2\ mA/\mu m^2$ and $T_j = 260°C$ for 2,445 h. Apparently, both HBTs exhibit the diffusion of Ti from the electrode and severe degradation of crystal quality in the InGaAs contact layer. Comparing the damaged regions in Fig. 18.8a and b, it is found that the damage spreads over the InGaAs layer with increasing stress time. Also noteworthy is that the InP layer is clearly visible in the TEM image although the damage spreads over the entire InGaAs layer (Fig. 18.8b). This means that the damaged region does not extend into the InP emitter layer.

Regarding the HBTs stressed at $J_C = 5\ mA/\mu m^2$, a "dark region" is also observed in the InGaAs contact layer (Fig. 18.8a). The growth directions of the dark region appear along the crystallographic orientations ([−100], [−1-11], and [−11-1]). Considering the junction temperatures during the tests were almost the same for the two HBTs, it would seem that current injection plays an important role in generating the dark region. In conclusion, the HBTs with the conventional electrode are significantly degraded with continuous operation, especially when they are subjected to a high current density for a long time.

### 18.5.2.2 TEM Images of HBTs with a Refractory Electrode After BT Tests

Figure 18.9 shows cross-sectional TEM images of the 1-μm-emitter HBTs having the electrode with Mo. The BT-stress conditions for the HBT shown in Fig. 18.9a were $J_C = 5\ mA/\mu m^2$ and $T_j = 260°C$ for 990 h, while those for the HBTs in Fig. 18.9b were $J_C = 2\ mA/\mu m^2$ and $T_j = 261°C$ for 2,445 h. Compared with the results for the conventional electrode, the diffusion of Ti and crystalline damage in the contact layer are very small, and the dark region does not appear even under 5-$mA/\mu m^2$ current injection. These results can most likely be explained in terms of

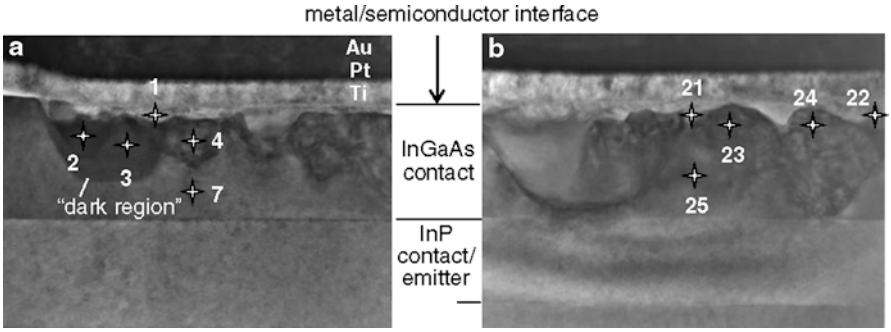



**Fig. 18.8** TEM image of the HBTs with a Ti/Pt/Au electrode. The HBTs were stressed at (**a**) $J_C = 5$ mA/μm$^2$ and $T_j \sim 250$°C for 1,576 h and (**b**) $J_C = 2$ mA/μm$^2$ and $T_j \sim 260$°C for 2,445 h

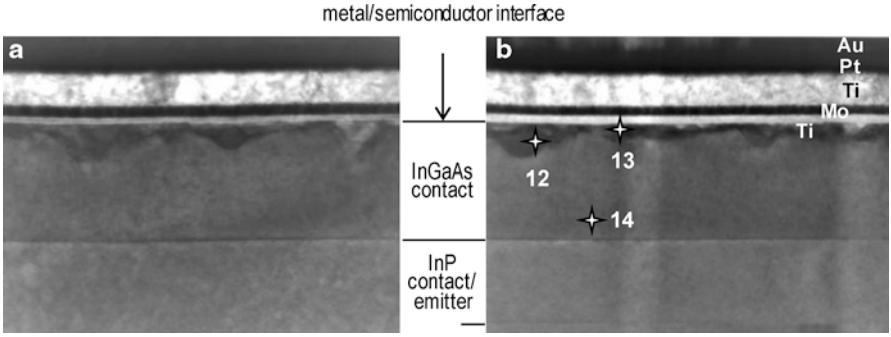



**Fig. 18.9** TEM image of the HBTs with a Ti/Mo-Ti/Pt/Au electrode. The HBTs were stressed at (**a**) $J_C = 5$ mA/μm$^2$ and $T_j \sim 260$°C for 990 h and (**b**) $J_C = 2$ mA/μm$^2$ and $T_j \sim 261$°C for 2,445 h

the small volume of the Ti contact layer and thermal stability of the Mo, which blocks the metal diffusion from the Ti/Pt/Au electrodes.

Figure 18.10 shows TEM images of the 0.5-μm-emitter HBTs having the electrode with W. These HBTs were stressed at $J_C = 5$ mA/μm$^2$ and $T_j = 270$–$280$°C for over 1,500 h. As in the case for the Ti/Mo-Ti/Pt/Au electrode, Ti diffusion into the emitter-contact layer is insignificant. In particular, for the HBT with the W-Ti/Pt/Au electrode, the interface between the electrode and semiconductor layer is very clear, and very little damage is observed in the contact layer. The W-Ti/Pt/Au is the most stable among the electrodes examined.

### 18.5.2.3 TEM Images of HBTs with a Ti/Pt/Au Electrode After HT Tests

Figure 18.11 shows a cross-sectional TEM image of the 1-μm-emitter HBT with the conventional Ti/Pt/Au electrode after a high-temperature (HT) storage test. The ambient temperature was 230°C, and the storage time was 2,000 h. Severe

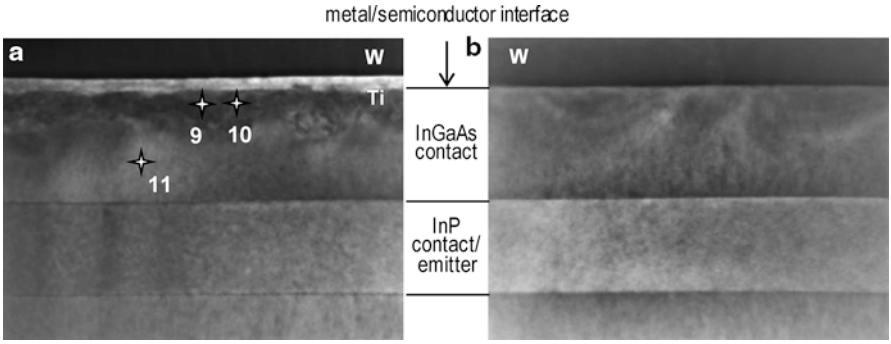



**Fig. 18.10** TEM image of the HBTs with (**a**) a Ti/W-Ti/Pt/Au electrode and (**b**) a W-Ti/Pt/Au electrode. The HBTs were stressed at (**a**) $J_C = 5$ mA/μm$^2$ and $T_j \sim 280$°C and (**b**) $J_C = 5$ mA/μm$^2$ and $T_j \sim 270$°C. The stress time was over 1,500 h

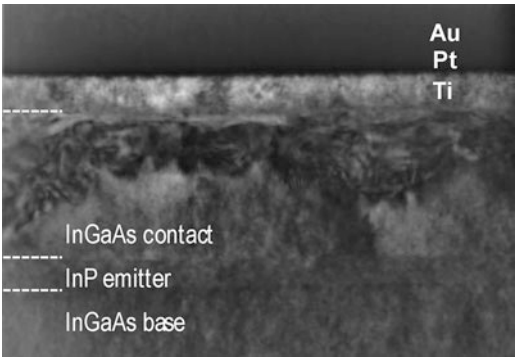



**Fig. 18.11** TEM image after the HT-storage test at 230°C for 2,000 h

crystalline damage is clearly observed in the upper side of the emitter-contact layer. This result indicates that only thermal stress can induce the metal diffusion from the electrode into the semiconductor layer. Incidentally, the dark region was not observed in the HBT tested.

### *18.5.3 EDS Analysis*

In order to understand the mechanism of the degradation in the contact layer, a microscopic analysis was performed by EDS to determine the atomic composition. Figure 18.12 shows the EDS spectra for the HBT in Fig. 18.8a. Notice that the measured points are marked on the TEM images. On the basis of these results, atomic composition was estimated for each point, as shown in Fig. 18.13. Here, Fig. 18.13a shows the atomic composition at the points where electrode

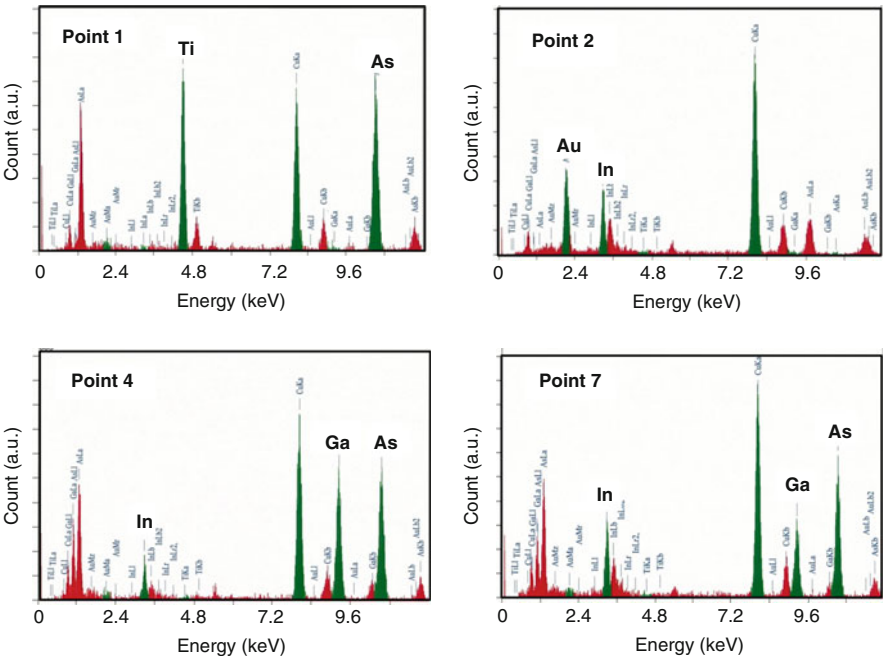



**Fig. 18.12** EDS spectra measured at points 1, 2, 4, and 7 marked in the TEM image of Fig. 18.8a

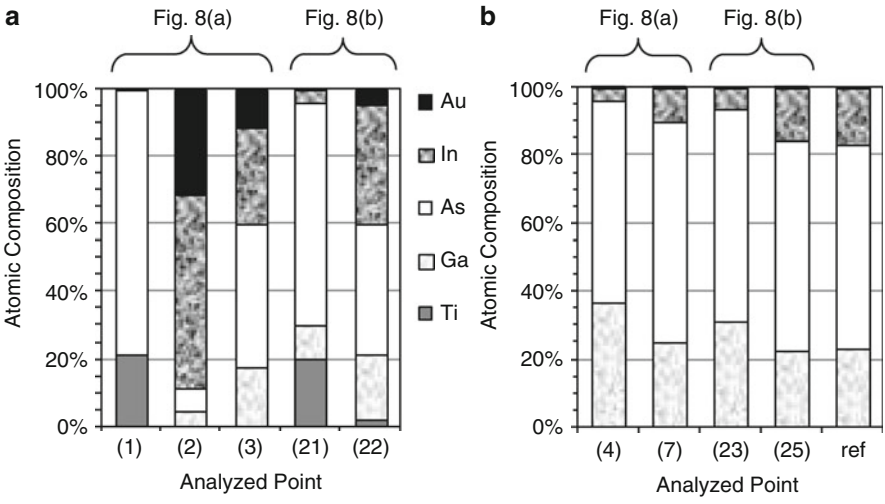



**Fig. 18.13** Atomic composition at several points in the emitter-contact layer of the HBTs with a conventional Ti/Pt/Au electrode. The column numbers represent the measured points in Fig. 18.8. For reference, normal composition for InGaAs is also shown in the column denoted by "*ref*"

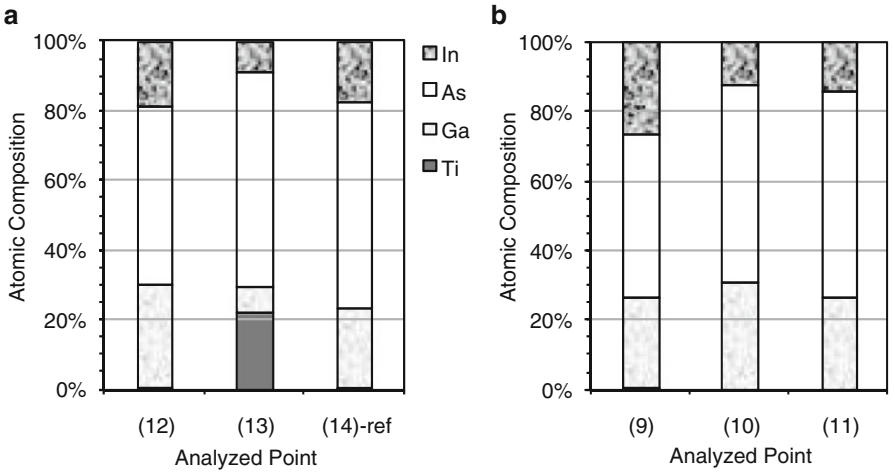



**Fig. 18.14** Atomic composition at several points in the emitter-contact layer of the HBTs with (**a**) a Ti/Mo-Ti/Pt/Au electrode and (**b**) a Ti/W-Ti/Pt/Au electrode. The column numbers represent the measured points in Figs. 18.9b and 18.10a

metal (Ti or Au) was detected. On the other hand, Fig. 18.13b shows the atomic composition at the points where electrode metal was not detected, but severe crystalline damage was confirmed from TEM observation. For reference, the normal atomic composition of lattice-matched InGaAs is also shown in the column denoted by "ref."

From Figs. 18.8 and 18.13, Ti diffusion is clearly confirmed in the original InGaAs region near the metal/semiconductor interface (points 1 and 21). In the Ti-diffusion region, As content significantly increases from the normal value, while In and Ga contents become very small. On the other hand, it is found that the dark region (points 2 and 3) is related to Au diffusion from the electrode. This suggests that Au atoms penetrate the Pt barrier metal via some weak spots. In the dark region, the atomic composition is rich in In and poor in Ga and As.

The diffusion of Ti and Au into the semiconductor layer causes a significant shift in the atomic composition, which strongly suggests the generation of interstitial atoms that then migrate in the emitter-contact layer. Thus, crystalline damage observed near the metal-diffusion region should also be associated with the shift in the atomic composition. Actually, the results of EDS measurements show clear difference in the atomic composition from the normal value. In the region (points 4 and 23), the Ga content increases and In content decreases, while the As content is not changed very much.

Figure 18.14 shows the EDS results for the HBTs with a refractory metal (Mo or W). Here, Fig. 18.14a and b shows the atomic composition at the points marked in Figs. 18.9b and 18.10a, respectively. In the semiconductor region adjacent to the Ti/Mo electrode (point 13), Ti is detected and Ga and In contents decrease, as in the case of the HBT with the conventional electrode. In the damaged region (point 12), the atomic composition of InGaAs is somewhat changed, while electrode metal is not detected. For the HBT with the Ti/W electrode, neither Ti nor Au is detected near

the metal/semiconductor interface (points 9 and 10), while the atomic composition is somewhat changed. In any case, however, the shift in the composition is not so large compared with the case for the conventional electrode.

The degradation in the emitter-contact layer is triggered by the diffusion of electrode metals. With refractory metal inserted between the thin Ti and the Ti/Pt/Au electrode, Au diffusion is suppressed even under high-current-density injection. In addition, Ti diffusion into the contact layer is not significant since the Ti is sufficiently thin. These are the reasons the newly developed electrodes suppress the degradation in the contact layer.

### *18.5.4 TED Analysis*

Au diffusion is observed in the HBTs stressed at the junction temperature of around 250°C, while Pt should act as a good barrier to the thermal diffusion of Au at such a relatively low temperature [26]. This result strongly suggests that the Au diffusion is due to the electromigration induced by high-current-density injection. The Au atoms in the electrode move along the direction of strong electric current and probably penetrate through weak spots in the Pt layer. The growth directions of the Au-diffusion region are along the crystallographic orientations, and a lattice-like periodic structure is confirmed from TEM observation, as shown in Fig. 18.15.

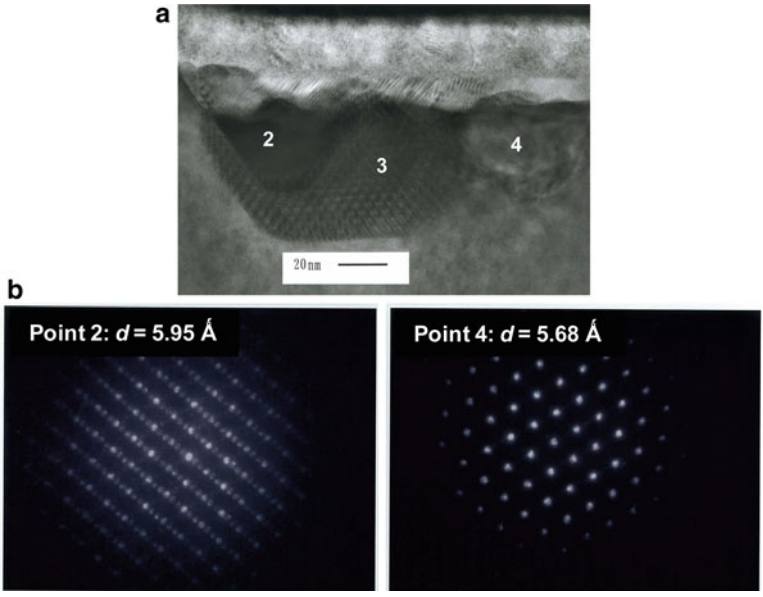



**Fig. 18.15** (**a**) Enlargement of the TEM image in Fig. 18.8a, showing a lattice-like periodic structure. (**b**) TED patterns measured at points 2 and 4 in the TEM image

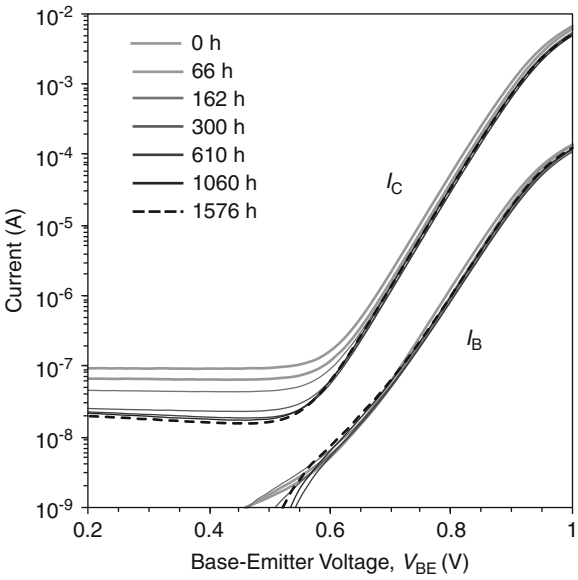



**Fig. 18.16** *I–V* characteristics (Gummel plots) during the BT-stress test for the HBT shown in Fig. 18.8a. The tested HBT exhibits good stability in the electrical characteristics, although there is significant degradation of crystal quality in the emitter-contact layer

These results suggest that the Au migration is limited by the local reaction with substitutional atoms of host species.

In order to get more insight into the Au diffusion, TED analyses were performed to determine the lattice constant in the degraded region (points 2 and 4). The diffraction patterns are also shown in Fig. 18.15. The lattice constant in the Au-diffusion region (point 2) is estimated to be 5.95 Å, which is larger than the value for latticed-matched InGaAs (5.87 Å). This result indicates that some mixed compound is generated in the region due to the Au invasion. From the EDS results, the creation of AuIn compound is strongly suggested; however, the lattice constant estimated by TED is larger than the value (5.83 Å) reported for AuIn [27]. The composition of the lattice-like structure observed in the Au-diffusion region is still unclear. Incidentally, the lattice constant in the region adjacent to the Au diffusion (point 4) is estimated to be 5.68 Å, which is very close to that of GaAs (5.65 Å). This result agrees well with the EDS analysis and indicates that a GaAs-like compound is generated near the Au-diffusion region.

### *18.5.5 Evaluation of Emitter Series Resistance*

The electrode-metal diffusion into the InGaAs contact layer during BT-stress tests triggers a local shift in the atomic composition, and an inhomogeneous compound is eventually generated. While the baseline HBTs exhibit little degradation of the *I–V* characteristics after the stress tests (as shown in Fig. 18.16), the local shift in the

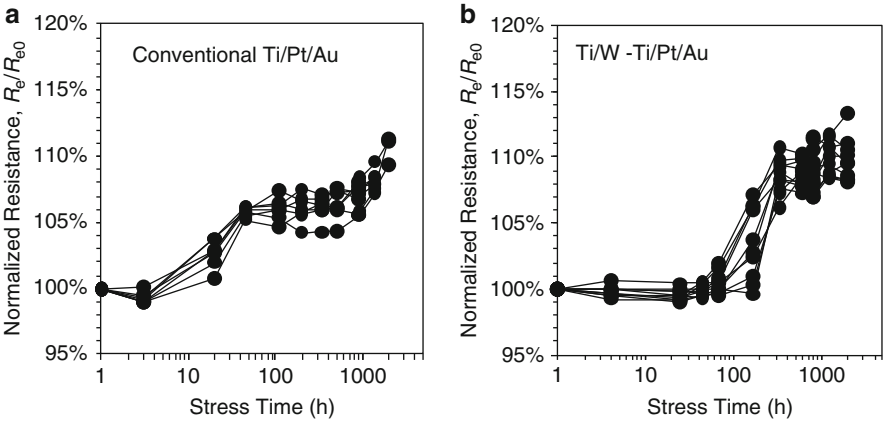



**Fig. 18.17** Time-wise change in normalized emitter resistance, $R_e/R_{e0}$, for the HBTs with (**a**) a Ti/Pt/Au electrode and (**b**) a Ti/W-Ti/Pt/Au electrode. The HBTs were stressed at (**a**) $J_C = 5$ mA/μm$^2$ and $T_j \sim 260$°C and (**b**) $J_C = 5$ mA/μm$^2$ and $T_j \sim 280$°C. The emitter resistance, $R_e$, was determined from the slope of $I_C$–$V_{BE}$ curves at $J_C = 2$ mA/μm$^2$

composition near the metal/semiconductor interface would cause some degradation of the emitter-contact characteristics.

The emitter resistance, $R_e$, was determined from the slope of $I_C$–$V_{BE}$ curves at $J_C = 2$ mA/μm$^2$, which is approximately equal to the sum of the dynamic emitter resistance, $r_e$ ($= nk_BT/qI_E$), and the emitter series resistance, $R_{ee}$. Figure 18.17 shows time-wise change in the emitter resistance normalized by the initial value, $R_e/R_{e0}$. Here, Fig. 18.17a and b, respectively, plots the normalized emitter resistance for the conventional electrode and for the refractory electrode with Ti/W. The stress current density used for the tests was 5 mA/μm$^2$. The HBTs with the refractory electrode exhibit an increase in the emitter resistance after 100 h of stress. On the other hand, for the HBTs with the conventional electrode, the emitter resistance begins to increase soon after 10 h and then saturates at about a 5% increase for several hundred hours.

Figure 18.18 shows the temperature dependence of MTTF for the HBTs with a Ti/Pt/Au, Ti/Mo-Ti/Pt/Au, or Ti/W-Ti/Pt/Au electrode. Here, critical time is defined as the time when $R_e$ increases by 3% of the initial value in order to avoid the influence of the $R_e$ saturation. In the figure, the same symbol denotes the results for the HBTs fabricated on the same wafer and stressed at the same current density. The solid lines indicate the temperature dependence for the HBTs (with Ti/Pt/Au or Ti/Mo-Ti/Pt/Au) stressed at $J_C = 2$ mA/μm$^2$. The dotted lines indicate the temperature dependence for the HBTs (with Ti/Pt/Au or Ti/W-Ti/Pt/Au) stressed at $J_C = 5$ mA/μm$^2$. As seen in the figure, the activation energy for the $R_e$ increase is nearly independent of the electrode configuration. This suggests that the Ti diffusion triggers the increase in the emitter resistance at the initial stage of the degradation. The activation energies are estimated to be 2.0 eV at $J_C = 2$ mA/μm$^2$ and 1.65 eV at $J_C = 5$ mA/μm$^2$.

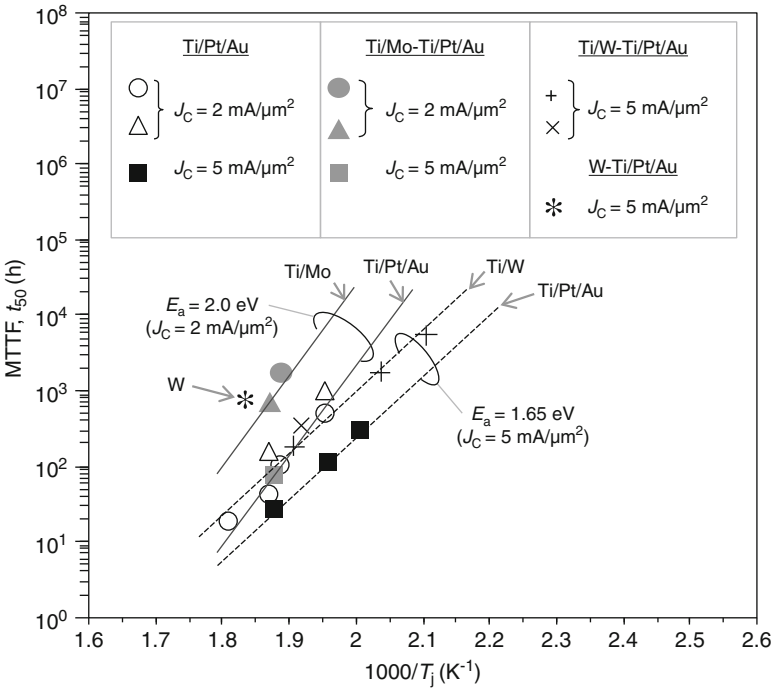



**Fig. 18.18** Arrhenius plot of MTTF for emitter resistance, $R_e$. Failure criterion is defined as a 3% increase in $R_e$ from the initial value. The activation energies, $E_a$, are estimated to be 2.0 eV at $J_C = 2$ mA/μm$^2$ and 1.65 eV at $J_C = 5$ mA/μm$^2$. In the figure, the same symbol denotes the results for the HBTs fabricated on the same wafer and stressed at the same current density

When the stress current density is 2 mA/μm$^2$, the MTTF for the Ti/Mo-Ti/Pt/Au electrode is about ten times larger than that for the conventional one. When it is 5 mA/μm$^2$, the MTTFs for the Ti/Mo- and Ti/W-Ti/Pt/Au electrodes are more than five times larger than that for the conventional one. These results are consistent with the microscopic analyses that earlier confirmed effective suppression of the Ti diffusion for the refractory electrodes. For the W-Ti/Pt/Au electrode, only one result is shown in Fig. 18.18. However, it is clear that this configuration provides the largest MTTF among the electrodes examined. When the Ti layer is used to achieve lower contact resistance, it should be designed to be as thin as possible in order to delay the onset of the increase in the emitter series resistance.

### *18.5.6 Discussion*

Here, we summarize the degradation behavior observed for HBTs with the conventional Ti/Pt/Au electrode (Fig. 18.19). The degradation starts with the Ti diffusion from the electrode into the semiconductor layer. Considering the fact that the

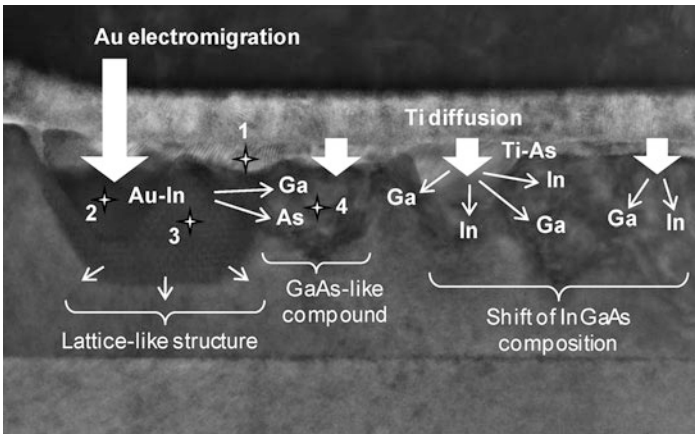



**Fig. 18.19** TEM image of the emitter-contact layer of the HBT with a conventional Ti/Pt/Au electrode. Behavior of defects is illustrated on the basis of the results of EDS and TED analyses. The Ti diffusion generates group-III interstitials. High-current-density injection induces electromigration of Au, which generates Ga and As interstitials

Ti-diffusion region (e.g., points 1 and 21 in Fig. 18.8) is rich in As and poor in In and Ga, the Ti atoms drive the group-III atoms out of substitutional sites. They also extract As atoms [26, 28, 29], and a TiAs compound is possibly generated. The resultant group-III interstitials then migrate through the crystal, and sometimes they may expel substitutional atoms to occupy their sites (the "kick-out" process [30]). Such a defect reaction causes local crystalline damage and shifts the atomic composition of InGaAs from the normal value. Presumably, the increase in the emitter resistance, $R_e$, at the initial stage reflects a change in the material properties at the metal/semiconductor interface (e.g., barrier height) due to the defect reaction. The rapid increase in $R_e$ observed for HBTs with the conventional electrode is simply due to a large supply of Ti from the thick (30 nm) layer.

Under high-current-density injection, Au atoms penetrate the Pt barrier metal and diffuse into the semiconductor layer. The Au migration is likely limited by the local reaction with the substitutional atoms of host species. Since Au and In have strong binding force, the migrating Au atoms extract In and drive Ga and As out of substitutional sites. As a result, the Au-diffusion region (points 2 and 3) becomes very rich in In and very poor in Ga and As. The resultant large amount of Ga and As interstitials then move to the adjacent region (point 4), where they form a GaAs-like compound. This reasonably explains the results of the EDS and TED analyses.

The crystalline damage associated with the shift in the atomic composition spreads over the InGaAs contact layer with stress time. This is simply because the interstitials generated in the contact layer migrate over a longer distance as the stress time elapses. As seen from Fig. 18.8b, however, crystalline damage is not observed in the InP contact or emitter layers even after long-term stressing. This strongly suggests that the InP layer can act as a barrier to the migration of defects generated in the InGaAs layer. For this reason, the contact-layer damage has little

impact on the current-gain characteristics for the HBT examined (Fig. 18.16). However, one has to note that devices with a degraded contact layer have a potential risk of fatal failure, which is described in the next section.

## 18.6 Fatal Degradation Observed in the Vicinity of Emitter-Base Junction

It has been reported that in GaAs-based HBTs, defect formation at the emitter–base junction results in fatal failure of electrical characteristics [31, 32]. The defect formation increases recombination current in the junction vicinity and causes a severe reduction of current gain. Such recombination current is often characterized by a large ideality factor of base current. In bipolar devices, the recombination process due to minority-electron injection is likely to enhance defect formation [33]. Thus, InP HBTs operating at high current density potentially suffer from degradation at the emitter–base junction. This section describes such degradation behavior observed in the study.

### *18.6.1 Degradation Caused by Defects from the Contact Layer*

First, it is shown that defects generated in the emitter-contact layer due to electrode-metal diffusion can induce severe degradation of crystal quality at the emitter–base junction. This degradation was observed for the HBTs with a relatively thin emitter layer.

Figure 18.20 shows the change in the electrical characteristics for the 1-μm-emitter HBTs with a conventional Ti/Pt/Au electrode stressed at $J_C = 5$ mA/μm$^2$ and $T_j = 260$°C. The tested HBTs have a 20-nm-thick InP emitter layer and do not have a heavily Si-doped InP contact layer. The other layer-structure parameters are exactly the same as those for the baseline HBTs described in Sect. 18.3. As seen in Fig. 18.20, the thin-emitter HBTs exhibit a sudden increase in leakage current and resultant drop in current gain within 1,000 h of stress. The ohmic-like *I–V* characteristics after the test suggest the fatal degradation at the emitter–base junction. These results are in contrast to those for the baseline HBT with a 90-nm-thick InP layer between the InGaAs contact and base layers (Fig. 18.16). Figure 18.21 shows cross-sectional TEM images for the region under the Ti/Pt/Au electrode. As expected, the dark region and crystalline damage are observed in the emitter-contact layer, indicating the generation of a large number of defects. Furthermore, significant damage is also confirmed at the emitter–base junction and in the base layer, which is not seen in the baseline HBT (Fig. 18.8). Also noteworthy is that the degradation in the emitter or collector layer is not so significant, however.

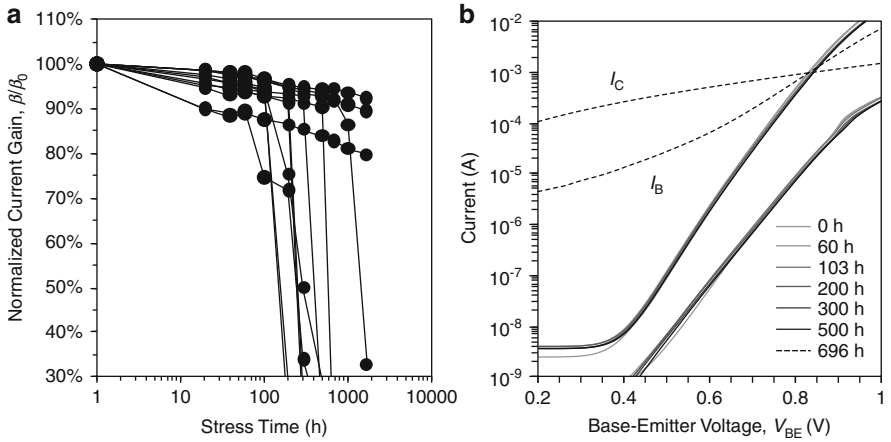



**Fig. 18.20** Results of the BT-stress test on the 1-μm-emitter HBTs with a conventional Ti/Pt/Au electrode. (**a**) Time-wise change in normalized current gain, $\beta/\beta_0$. (**b**) Typical *I–V* characteristics (Gummel plots) during the BT-stress tests. The $\beta$ was measured at $J_C = 1$ mA/μm$^2$. The HBTs were stressed at $J_C = 5$ mA/μm$^2$ and $T_j \sim 260°$C. The tested HBTs have a thin (20 nm) InP emitter layer and do not have a heavily Si-doped InP contact layer. Sudden drops of $\beta$ correspond to the fatal device failure characterized by the ohmic-like *I–V* characteristics

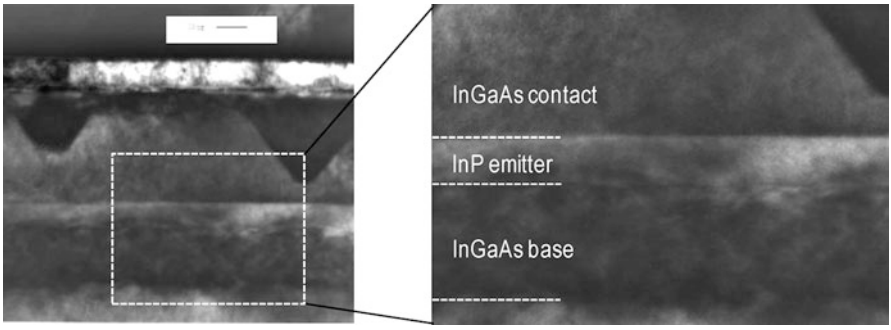



**Fig. 18.21** TEM image of the 1-μm-emitter HBT with a conventional Ti/Pt/Au electrode, whose *I–V* characteristics are shown in Fig. 18.20b. Significant degradation of crystal quality is confirmed in the entire base layer. Also noteworthy is that the degradation in the emitter or collector is not so significant, however

The sudden degradation of the *I–V* characteristics is most likely due to the damage in the vicinity of the emitter–base junction caused by the defects from the emitter-contact layer. One possible explanation is as follows: When the InP emitter layer is sufficiently thin, some defects continuously penetrate the InP emitter during the stress test and accumulate in the p-type InGaAs base. After long-term stressing, the number of the defects in the base becomes large enough to rapidly degrade the crystal quality at the junction.

### 18.6.2 Intrinsic Degradation Due to High-Current-Density Injection

The degradation mentioned above is based on the arrival of defects from the outside of the device intrinsic region. Such degradation can be avoided by employing a refractory electrode to suppress the generation of defects. Even for such devices, however, fatal degradation has been found to occur when they are subjected to very high current density (>5 mA/μm$^2$).

Figure 18.22 shows the change in the *I–V* characteristics for the 0.5-μm-emitter HBTs with a Ti/W-Ti/Pt/Au electrode stressed by 5- or 10-mA/μm$^2$ current. The junction temperature, $T_j$, was around 230°C during the tests. For the HBTs stressed at $J_C = 5$ mA/μm$^2$, good stability in the *I–V* characteristics is confirmed even after 10,000 h of stress. In contrast, the HBTs stressed at $J_C = 10$ mA/μm$^2$ exhibit a gradual increase in the base current at low $V_{BE}$ after 1,000 h of stress. The base current is characterized by a large ideality factor. Figure 18.23 shows the temperature dependence of MTTF for the increase in the base current when the stress current density is 7 or 10 mA/μm$^2$. Here, the failure criterion is defined as a 0.5-μA increase in the base current at $V_{BE} = 0.74$ V. The activation energy is estimated to be around 1.1 eV, which appears to be nearly independent of the current density. The relatively low value of the activation energy means that this degradation mode plays an important role in determining the device failure when very high current density (>5 mA/μm$^2$) is used for HBT operation.

Figure 18.24 plots the shift of current turn-on voltage, $\Delta V_o$, and normalized current gain, $\beta/\beta_0$, as a function of stress time. The turn-on voltage, $V_o$, is defined as $V_{BE}$ at $I_C = 10$ μA. For the devices stressed at $J_C = 10$ mA/μm$^2$, significant reductions in the $V_o$ and $\beta$ are confirmed after 1,000 h of stress. Such degradation behavior suggests that the generation of defects at the emitter–base junction enhances tunneling recombination in the junction vicinity. The diffusion of carbon from the base into the emitter is also suggested, since carbon tends to be a donor in InP and thereby enhances the tunneling current from the emitter by reducing the width of the emitter spike barrier.

Figure 18.25 shows a cross-sectional TEM image of the degraded device. The BT-stress conditions for the HBT were $J_C = 7$ mA/μm$^2$ and $T_j = 322$°C, and the stress time was about 2,000 h. While the degradation in the emitter-contact layer is not significant owing to the use of the refractory electrode, local crystalline damage is clearly observed at the emitter–base junction. Notice that the damage concentrates at the InP/InGaAs heterointerface. These results imply that some "native" defects, which have been introduced during the epitaxial growth or the fabrication process, are involved in the degradation. Considering that the population of minority electrons in the base is larger in the heterointerface vicinity, the degradation is possibly enhanced by recombination centers near the heterointerface. Consequently, in order to further improve the reliability of InP HBTs, the defects have to be reduced by optimizing the conditions for epitaxial growth and device fabrication.

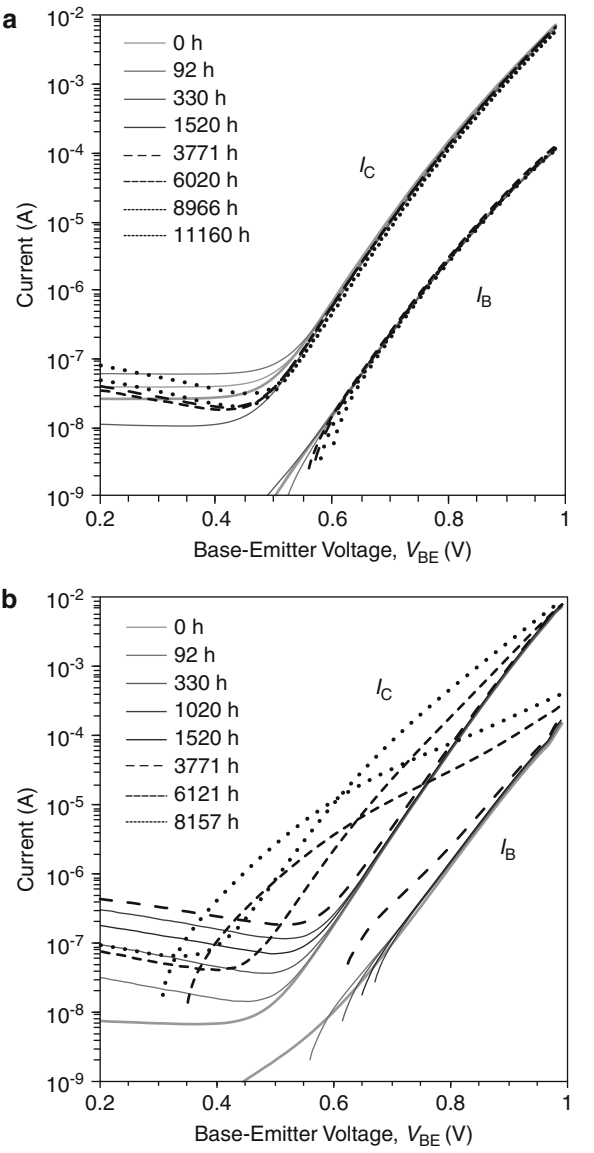



**Fig. 18.22** Typical *I–V* characteristics (Gummel plots) for the 0.5-μm-emitter HBTs with a Ti/W-Ti/Pt/Au electrode during BT-stress tests. The HBTs were stressed at (**a**) $J_C = 5$ mA/μm$^2$ and $T_j \sim 225$°C and (**b**) $J_C = 10$ mA/μm$^2$ and $T_j \sim 235$°C. For the HBT stressed at 5 mA/μm$^2$, good stability is confirmed even after 10,000 h of stress. In contrast, the HBT stressed at 10 mA/μm$^2$ exhibits a gradual increase in the base current at low $V_{BE}$ after 1,000 h of stress

## 18.7 Summary and Prospects

The degradation behavior of InP HBTs operating at current densities of up to 10 mA/μm$^2$ has been investigated in order to apply these HBTs to practical ICs. The rapid degradation of current gain, $\beta$, was observed for the HBTs without ledge passivation, which is associated with surface recombination in the external-base region. Such degradation can be suppressed by introducing InP-ledge structures

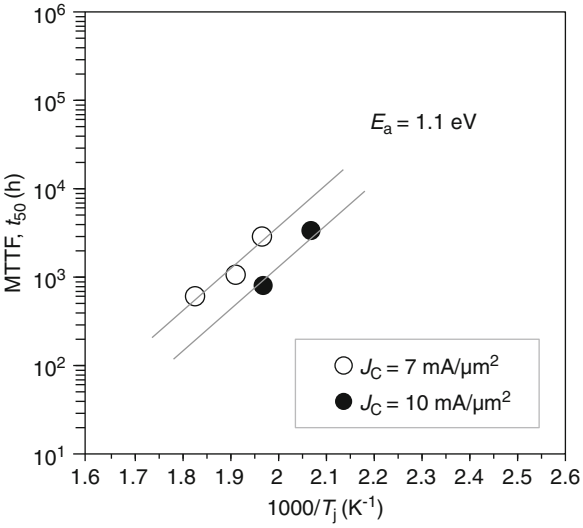



**Fig. 18.23** Arrhenius plot of MTTF for base leakage current. Failure criterion is defined as a 0.5-μA increase in $I_B$ at $V_{BE} = 0.74$ V. The relatively low activation energy ($E_a = 1.1$ eV) indicates the importance of this degradation mode for HBTs operating at very high current density

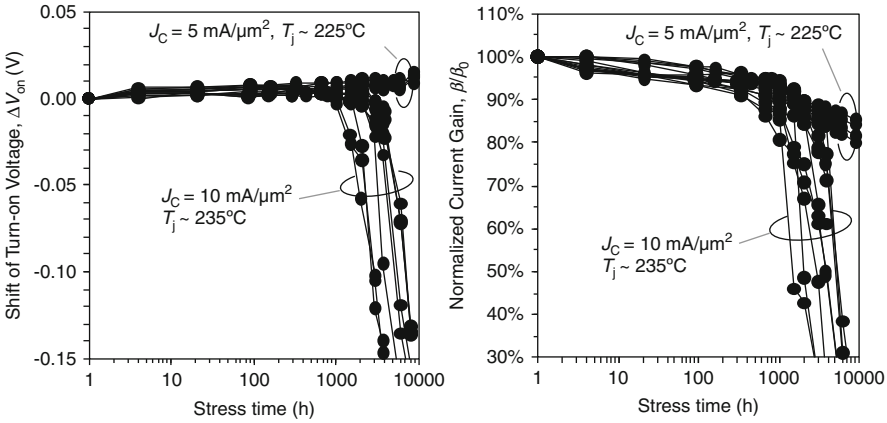



**Fig. 18.24** Time-wise change in current turn-on voltage, $V_o$, and normalized current gain, $\beta/\beta_0$, for the 0.5-μm-emitter HBTs with a Ti/W-Ti/Pt/Au electrode. The HBTs were stressed at $J_C = 10$ mA/μm$^2$ ($T_j$ ~ 235°C) or $J_C = 5$ mA/μm$^2$ ($T_j$ ~ 225°C)

with a SiN passivation film, and long-term stability in current gain can be achieved. For these HBTs, the activation energy (for a 15% reduction in $\beta$) is 1.7 eV, and it is nearly independent of current density (at least) up to 5 mA/μm$^2$. The increase in the recombination current is simply caused by the thermal stress that affects the material properties at the ledge surface.

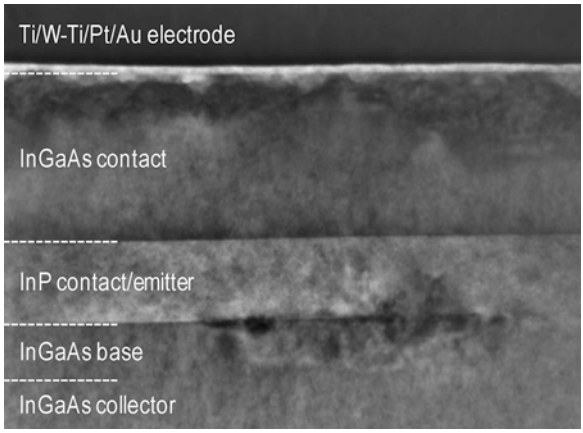



**Fig. 18.25** TEM image of the 0.5-μm-emitter HBT with a Ti/W-Ti/Pt/Au electrode. The HBT was stressed at $J_C = 7$ mA/μm$^2$ and $T_j \sim 322$°C. The stress time was about 2,000 h. Local crystalline damage is clearly observed at the emitter–base junction

TEM observations have revealed significant degradation of the crystal quality in the emitter-contact layer for HBTs with a conventional Ti/Pt/Au emitter electrode. Microscopic analyses by EDS and TED were conducted to investigate the underlying mechanism of the degradation. The results show that the degradation is triggered by metal diffusion from the electrode caused by the thermal effect and high densities of electron flow. The Ti diffusion generates interstitial atoms of host species, which then shift local atomic composition in the contact layer. High-current-density injection also induces electromigration of Au, which also generates a large number of interstitials in the contact layer. In order to suppress such degradation, inserting a very thin Ti layer and refractory metal, such as Mo and W, under the Ti/Pt/Au electrode is very useful. The very thin Ti layer keeps the contact resistance sufficiently low and suppresses the amount of Ti diffusion. The refractory metal plays an important role in blocking the Au migration into the semiconductor layer.

The emitter resistance, $R_e$, for the stressed HBTs was evaluated from *I–V* characteristics in order to elucidate the impact of the contact-layer degradation on emitter series resistance. The activation energy (for a 3% increase in $R_e$) is 2.0 eV at the current density of 2 mA/μm$^2$, while it is 1.65 eV at 5 mA/μm$^2$. The activation energy is independent of the electrode-metal configurations examined, which indicates that the increase in $R_e$ at the initial stage is simply due to the diffusion of Ti. For the refractory electrodes, critical time can be improved by thinning the Ti contact layer, and long-term stability of the emitter series resistance is expected even for operation at higher current densities.

Finally, the degradation in the vicinity of the emitter–base junction was investigated. For the HBTs with a thin InP emitter, defects generated in the contact layer can induce fatal damage at the emitter–base junction. Since device downscaling requires thinner InP-ledge structures to suppress the current-gain reduction (or, emitter-size effect), the refractory electrode is also a prerequisite to achieving small, high-speed HBTs. Even for such advanced devices, however, intrinsic degradation can take place at the emitter–base junction when they are subjected

to very high current density (>5 mA/μm$^2$). Further improvement of the crystal quality at the junction is important in order to increase the current density available for practical device operation.

**Acknowledgments** The authors thank Minoru Ida and Norihide Kashio for fabrication of HBTs and for discussions. They also thank Takatomo Enoki, Shoji Yamahata, Masami Tokumitsu, and Tomoyuki Akeyoshi for their encouragement throughout this work. Thanks are also due to Mayumi Mitsuhashi for her assistance.

## References

1. K. Sano, K. Ishii, K. Murata, K. Kurishima, M. Ida, T. Shibata, T. Enoki, H. Sugahara. Over-100-Gbit/s multiplexing operations of InP DHBT selector IC designed with high collector-current density. *Extended Abstracts of International Conference on Solid State Devices and Materials (SSDM)*, 2004, pp. 312–313. (JSAP, Tokyo)
2. S. Tsunashima, K. Murata, M. Ida, K. Kurishima, T. Kosugi, T. Enoki, H. Sugahara, A 150-GHz dynamic frequency divider using InP/InGaAs DHBTs. *Technical Digests of GaAs IC Symposium*, 2003, pp. 284–287. (IEEE, Piscataway, NJ)
3. M.J. Rodwell, B. Brar, InP bipolar ICs: scaling roadmaps, frequency limits, manufacturable technologies. Proc. IEEE **96**(2), 271–286 (2008)
4. H. Kamitsuna, K. Kurishima, M. Ida, K. Sano, 100 Gbit/s clock recovery OEIC using InP/InGaAs HPT. Electron. Lett. **40**(12), 764–766 (2004)
5. N. Kashio, K. Kurishima, K. Sano, M. Ida, N. Watanabe, H. Fukuyama, Monolithic integration of InP HBTs and uni-traveling-carrier photodiodes using nonselective regrowth. IEEE Trans. Electron Devices **54**(7), 1651–1657 (2007)
6. Y.K. Fukai, K. Kurishima, M. Ida, S. Yamahata, T. Enoki, Highly reliable InP-based HBTs with a ledge structure operating at high current density. Electron. Commun. Jpn, Pt 2 **90**(4), 1–8 (2007)
7. N. Kashio, K. Kurishima, Y.K. Fukai, S. Yamahata, Highly reliable submicron InP-based HBTs with over 300-GHz $f_T$. IEICE Trans. Electron. **E91-C**(7), 1084–1090 (2008)
8. N. Kashio, K. Kurishima, Y.K. Fukai, M. Ida, High-speed and high-reliability InP-based HBTs with a novel emitter. IEEE Trans. Electron Devices **57**(2), 373–379 (2010)
9. Y.K. Fukai, K. Kurishima, N. Kashio, M. Ida, S. Yamahata, T. Enoki, Emitter-metal-related degradation in InP-based HBTs operating at high current density and its suppression by refractory metal. Microelectron. Reliab. **49**(4), 357–364 (2009)
10. Y.K. Fukai, K. Kurishima, N. Kashio, M. Ida, S. Yamahata, Reliability study on InP/InGaAs emitter–base junction for high-speed and low-power InP HBT. *Proceedings of International Conference on Indium Phosphide and Related Materials (IPRM)*, 2010, pp. 119–122. (IEEE, Piscataway, NJ)
11. M. Hafizi, R.A. Metzger, W.E. Stanchina, D.B. Rensch, J.F. Jensen, W.W. Hooper, The effects of base dopant diffusion on DC and RF characteristics of InGaAs/InAlAs heterojunction bipolar transistors. IEEE Electron Device Lett. **13**(3), 140–142 (1992)
12. F.M. Yamada, A.K. Oki, D.C. Streit, Y. Saito, D.K. Umemoto, L.T. Tran, S. Bui, J.R. Velebir, G.W. Mclver, Reliability analysis of microwave GaAs/AlGaAs HBTs with beryllium and carbon doped base. *International Microwave Symposium Digests of IEEE MTT-S*, 1992, pp. 739–742. (IEEE, Piscataway, NJ)
13. S.A. Stockman, A.W. Hanson, G.E. Stillman, Growth of carbon-doped *p*-type $In_xGa_{1-x}As$ ($0 < x < 0.53$) by metalorganic chemical vapor deposition. Appl. Phys. Lett. **60**(23), 2903–2905 (1992)

14. H. Ito, S. Yamahata, N. Shigekawa, K. Kurishima, Y. Matsuoka, Growth and characterization of high-speed carbon-doped-base InP/InGaAs heterojunction bipolar transistors by metalorganic chemical vapor deposition. Jpn. J. Appl. Phys. Pt1 **35**(6A), 3343–3349 (1996)
15. K. Kurishima, S. Yamahata, H. Nakajima, H. Ito, Y. Ishii, Performance and stability of MOVPE-grown carbon-doped InP/InGaAs HBT's dehydrogenated by an anneal after emitter mesa formation. Jpn. J. Appl. Phys. Pt1 **37**(3B), 1353–1358 (1998)
16. M. Borgarino, R. Menozzi, D. Diecci, L. Cattani, F. Fantini, Reliability physics of compound semiconductor transistors for microwave applications. Microelectron. Reliab. **41**(1), 21–30 (2001)
17. S.R. Bahl, N. Moll, V.M. Robbins, H.-C. Kuo, B.G. Moser, G.E. Stillman, Be diffusion in InGaAs/InP heterojunction bipolar transistors. IEEE Electron Device Lett. **21**(7), 332–334 (2000)
18. K. Kiziloglu, S. Thomas III, F. Williams, B.M. Paine, Reliability and failure criteria for InAlAs/GaInAs/InP HBTs. *Proceedings of International Conference on Indium Phosphide and Related Materials (IPRM)*, 2000, pp. 294–297. (IEEE, Piscataway, NJ)
19. S. Thomas III, M. Chen, R. Bowen, Reliability of $1 \times 5\ \mu m^2$ emitter InAlAs/InGaAs HBTs under bias and thermal stress. *Proceedings of GaAs Reliability Workshop*, 2002, pp. 137–140. (JEDEC, Arlington, VA)
20. R. Lee, Accelerated life test results of InGaAs/InP heterojunction bipolar transistors. *Technical Digests of GaAs IC Symposium*, 2003, pp. 74–77. (IEEE, Piscataway, NJ)
21. K.T. Feng, N.X. Nguyen, C. Nguyen, Investigation of reliability for C-doped InP/InGaAs/InP HBTs under high current density operation. *Proceedings of GaAs Reliability Workshop*, 2003, pp. 117–120. (JEDEC, Arlington, VA)
22. C.-Y. Chen, S.-I. Fu, C.Y. Chang, C.-H. Tsai, C.-H. Yen, S.-F. Tsai, R.-C. Liu, W.-C. Liu, Influence of surface sulfur treatments on the temperature-dependent characteristics of HBTs. IEEE Trans. Electron Devices **51**(12), 1963–1971 (2004)
23. B.P. Yan, Y.F. Yang, C.C. Hsu, H.B. Lo, E.S. Yang, A reliability comparison of InGaP/GaAs HBTs with and without passivation ledge. Microelectron. Reliab. **41**(12), 1959–1963 (2001)
24. R. Yamabi, K. Kotani, K. Kawasaki, M. Yaegashi, H. Yano, Reliability of InGaAs/InP HBTs with InP passivation structure. *Proceedings of International Conference on Indium Phosphide and Related Materials (IPRM)*, 2003, pp. 122–125. (IEEE, Piscataway, NJ)
25. E.F. Chor, R.J. Malik, R.A. Hamm, R. Ryan, Metallurgical stability of ohmic contacts on thin base InP/InGaAs/InP HBT's. IEEE Electron Device Lett. **17**(2), 62–64 (1996)
26. Y.T. Lyu, K.L. Jaw, C.T. Lee, C.D. Tsui, Y.J. Lin, T.T. Cheng, Ohmic performance comparison for Ti/Ni/Au and Ti/Pt/Au on InAs/graded InGaAs/GaAs layers. Mater. Chem. Phys. **63**(2), 122–126 (2000)
27. A. Piotrowska, P. Auvray, A. Guivarc'h, G. Pelous, On the formation of binary compounds in Au/InP system. J. Appl. Phys. **52**(8), 5112–5117 (1981)
28. D.G. Ivey, S. Ingtrey, J.-P. Noel, W.M. Lau, Microstructural study of Ti/Pt/Au contacts to *p*-InGaAs. Mater. Sci. Eng. **B49**(1), 66–73 (1997)
29. J.S. Huang, C.B. Vartuli, Scanning transmission electron microscopy study of Au/Zn/Au/Cr/Au and Au/Ti/Pt/Au/Cr/Au contacts to *p*-type InGaAs/InP. J. Appl. Phys. **93**(9), 5196–5200 (2003)
30. U. Gösele, F. Morehead, Diffusion of zinc in gallium arsenide: a new model. J. Appl. Phys. **52**(7), 4617–4619 (1981)
31. R.E. Welser, M. Chaplin, C.R. Lutz, N. Pan, A. Gupta, B. Veasel, A. Ezis, Base current investigation of the long term reliability of GaAs-based HBTs. *Technical Digests of International Conference on GaAs Manufacturing Technology*, 2000, pp. 145–148. (CS MANTECH, Beaverton, OR)
32. R.E. Welser, P.M. Deluca, Exploring physical mechanisms for sudden beta degradation in GaAs-based HBTs. *Proceedings of GaAs Reliability Workshop*, 2001, pp. 135–138. (JEDEC, Arlington, VA)
33. M. Uematsu, K. Wada, Recombination-enhanced impurity diffusion in Be-doped GaAs. Appl. Phys. Lett. **58**(18), 2015–2017 (1991)

# Index


**A**
Ablation, 82
Acceleration
  excessive, 97, 189, 548
  factor, 156, 162, 168, 175–177, 179, 180, 320, 323–325, 327, 333, 343, 349, 353, 374
  matrix, 163, 168, 175, 179
  model, 163–169, 175–177
Activation energy, 7, 68, 97, 98, 105, 106, 164, 166, 168, 171, 175, 176, 179, 181, 182, 254, 264–266, 268–272, 275, 279, 303, 307, 310–312, 328, 331, 346, 347, 350, 351, 432, 433, 440, 442, 449, 457, 466, 480, 537, 585, 588, 590, 591, 600, 601, 605, 607, 608
Adiabatic instability, 285, 299
Adiabatic potential, 286, 287, 299–301, 304–306, 311, 313
Aging, 7, 22, 55, 96, 139, 154, 216, 251, 329, 455, 485, 578
AlGaAs, 3, 9, 27, 32, 36, 37, 89–102, 106, 109–119, 152, 182, 292, 296, 456, 465, 467, 470, 546
AlGaN, 220, 248, 381, 431, 456, 476, 515, 546
Aluminum-free epitaxial material, 130, 141
Antiphase defect, 276
APDs. *See* Avalanche photodiodes (APDs)
Arrhenius plot, 104, 168, 176, 182, 590, 591, 601, 607
Atomic composition, 34, 595–599, 602, 608
Auger recombination, 257, 314
Avalanche photodiodes (APDs), 55–83
Average conductivity effective mass, 384

**B**
Bandgap energy, 265, 266, 269, 272
Band-splitting, 396–397, 401
Band warping, 384, 396, 397, 399, 401
BCB. *See* Benzocyclobutene (BCB)
Benzocyclobutene (BCB), 587–589
Bias-temperature (BT) stress, 587, 590, 593, 599, 604–606
Biaxial, 384–385, 389, 396–401, 422–424, 552–555, 558–568, 571, 572
Bipolar device, 264, 279, 603
Bourgoin-Corbett (B-C) mechanism, 309
Broad area laser, 123–142
Built-in stress, 384–385, 560, 562
Bulk defect-assisted tunneling, 403, 415
Burn-in
  VCSEL, 153, 157–161, 185

**C**
Catastrophic optical damage (COD), 23, 60–64, 66, 67, 83, 89, 123–142, 238–239, 250
CCD. *See* Configuration coordinate diagram (CCD)
Channel electron mobility, 393, 396–401
Climb dislocation, 29, 67, 89, 91, 148, 152, 153, 159, 182–185, 191, 192, 194–200, 219, 264, 277–279
COD. *See* Catastrophic optical damage (COD)
Confidence limits, 171, 178–180
Configuration coordinate diagram (CCD), 31, 32, 193, 199, 285, 299–307, 310, 313, 315



O. Ueda and S.J. Pearton (eds.), *Materials and Reliability Handbook for Semiconductor Optical and Electron Devices*, DOI 10.1007/978-1-4614-4337-7,


© Springer Science+Business Media New York 2013

Contact, 20, 68, 89, 124, 156, 209, 249, 322, 385, 432, 456, 482, 516, 547, 584
degradation, 39, 249, 259, 432–434, 461, 465, 470
Pt/Ti/Pt/Au for base contact, 587
Ti/Pt/Au for emitter and collector contacts, 587
Copper contamination, 155, 188
Corrosion, 163, 322, 323, 328, 331–332, 341, 363, 456, 459, 471
Corrosion failures, 332, 341
Critical angle, 126, 134, 141
Crystalline damage, 593–595, 597, 602, 603, 605, 608
Current filamentation
VCSEL, 188–189
Current gain, 459, 466, 487, 584, 585, 588–591, 603–608
Current turn-on voltage, 584, 585, 607

**D**
Dangling bonds (DBs), 92, 152, 153, 268, 276, 296, 297, 499
Dark contrast, 108, 268
Dark current, 13, 14, 16, 75–81, 83
Dark line defects (DLDs), 20, 22, 25, 28, 47, 88, 90, 91, 94, 95, 97, 99, 106, 130, 131, 134, 140, 141, 148, 159, 160, 182–184, 189, 192, 194, 196
DBs. *See* Dangling bonds (DBs)
DC stress, 433, 434, 437–439, 446, 459, 460, 462–464
Deep-level (DL), 75, 89, 92, 96–102, 117–119, 240, 257–258, 285, 294, 296, 304–306, 480
Defects, 6, 21, 62, 89, 124, 152, 207, 248, 264, 283, 321, 382, 432, 488, 526, 547, 548, 584
band induced leakage, 403, 415, 421–422, 425
reaction, 67, 89, 102, 103, 116, 118, 239, 269, 283–315, 602
reliability, 13, 14, 49, 157, 201, 208, 218, 248, 284, 321, 332, 333, 340, 354, 359, 360, 363–367, 369–373, 376, 382, 548, 572, 573, 605
Degradation, 4, 23, 55, 88, 150, 208, 247, 264, 284, 320, 382, 432, 457, 485, 516, 548, 584
2DEG sheet carrier density, 382, 389, 393–396, 401, 408, 411, 424
DFB lasers. *See* Distributed feedback (DFB) lasers
Diffusion, 28, 58, 63, 69, 72, 91, 96, 104–106, 109, 116, 119, 173, 174, 221, 239–241, 253, 254, 259, 260, 268, 278, 307, 309, 314, 328, 363, 439, 449, 456, 461, 464–466, 468, 517, 528, 529, 532, 533, 537–543, 584, 585, 591–603, 605, 608
Direct tunneling, 403, 415, 417–420, 422, 425
Dislocation, 6, 24, 66, 88, 130, 148, 208, 248, 264, 284, 384, 439, 456, 488, 538, 552
climb, 29, 67, 89, 91, 148, 152, 153, 159, 182–185, 191, 192, 194–200, 219, 264, 277, 278
core, 91–93, 240, 267, 268, 277
glide, 67, 89, 93–96, 219, 258, 264, 269, 272, 275, 278
-related electronic levels, 267, 268
30° Dislocation, 266
90° Dislocation, 266, 271
Distributed feedback (DFB) lasers, 9–11, 68, 71, 72
DLDs. *See* Dark line defects (DLDs)
Drain current, 389–402, 424, 432, 437, 446, 447, 450, 459–461, 464, 472, 476, 477, 481, 482, 507, 508, 510, 537
Driving force, 219, 267, 272, 275–279
DX center, 118, 292, 296, 297

**E**
EDS. *See* Energy dispersive X-ray spectroscopy (EDS)
E2 facet passivation, 128, 139, 142
EL. *See* Electroluminescence (EL)
EL2, 118, 292, 296
Electrical overstress (EOS)
field applications, 337
in lab testing, 186
Electric field model, 407–412
Electroluminescence (EL), 20, 22, 24–27, 31, 90, 94, 95, 99–103, 106, 133, 134, 137, 171, 185, 189, 191–193, 195–196, 433, 446–448
Electromigration, 322, 323, 328, 330–331, 351, 363, 456, 461, 472, 516, 598, 602, 608
Electronic excitation, 219, 220, 283, 285, 296, 307–309
Electronic levels, 267, 268

Electron-lattice interactions, 285–298, 301, 304, 314
Electrostatic discharge (ESD), 6, 20–22, 24, 152, 161, 184–186, 194, 197, 199, 322, 323, 332–333, 337–339, 341, 345, 352, 371, 372, 456, 457, 471
Emitter-base junction, 584, 603–606, 608
Emitter resistance, 600–602, 608
  dynamic emitter resistance, 600
  emitter series resistance, 592, 599–601, 608
Energy dispersive X-ray spectroscopy (EDS), 36–38, 228, 460, 464, 587, 591, 595–599, 602, 608
EOS. *See* Electrical overstress (EOS)
Epitaxial defects, 187, 195
ESD. *See* Electrostatic discharge (ESD)
External base, 584, 585, 587–589, 606
External stress, 119, 394–396, 424
Extrinsic self-trapping, 285, 291–292, 314

**F**
Facet degradation, 60–64, 66, 83, 150, 239, 244, 250–251
Failure, 4, 19, 57, 88, 125, 148, 236, 254, 320, 431, 455, 515, 548, 585
  criteria, 8, 161–162
  mechanisms, 20–22, 24, 25, 27, 51, 148, 150, 164, 168–171, 181–191, 196, 199, 200, 323, 324, 327–346, 350–353, 359, 360, 363, 365, 367, 373, 374, 431, 432, 455, 456, 459, 548, 578
  modes, 4, 12–13, 15–16, 23, 28, 62, 64, 123, 138, 141, 171, 181, 335, 351, 457, 548, 585
Failure analysis, 19–51, 55–83, 148, 156, 166, 167, 169–171, 184, 186–199, 332–334, 336, 341, 342, 587
  VCSEL, 21, 25, 30–32, 34, 42, 47–49, 148, 184, 190–199
Failure modes and effects analysis (FMEA), 157, 169
Field crowding, 433
Field plate, 432, 436, 439, 440, 547, 552, 573, 578–579
Filaments, 133, 188–189
First-principles calculations, 275, 527, 532, 533, 535–537
FIT rate, 180–181
FMEA. *See* Failure modes and effects analysis (FMEA)
Forward biasing, 4–13, 15, 188, 195, 264, 273, 279, 446, 448, 467, 497
Four-point bending, 386, 387, 389, 564

**G**
GaAs, 3, 26, 59, 89, 124, 150, 208, 250, 264, 287, 321, 381, 431, 456, 476, 546, 583
Gallium nitride (GaN), 4, 109, 200, 207–241, 247–260, 327, 355, 376, 381–425, 431–450, 456, 475–510, 515, 518–523, 537–543, 546–548, 551–579
GaN. *See* Gallium nitride (GaN)
Gate current, 382, 383, 403–423, 433, 434, 437–442, 444, 446, 459, 462–464, 471, 487, 537, 541
Gate lag, 329, 382, 432, 459, 472, 481, 483–486, 488, 489
Gate leakage, 403, 405, 415–425, 437, 440, 442, 446, 448–450, 487, 498, 504
Gate sinking degradation, 351, 433, 459, 461, 464
Gauge factor, 388–393, 401–402, 424
Glide set, 266, 267
Glide velocity, 264, 265, 272, 273, 275, 279
Gradual degradation, 70, 87–120, 150, 238–241, 259

**H**
HBTs. *See* Heterojunction bipolar transistors (HBTs)
HEMTs. *See* High electron mobility transistors (HEMTs)
Heterojunction bipolar transistors (HBTs), 331, 355, 455–472, 583–609
High electron mobility transistors (HEMTs), 154, 381–425, 431–450, 455–472, 476, 477, 479–497, 509, 510, 515, 519–525, 537–539, 541–543, 546–548, 551–553, 568–576, 579
High power lasers, 24, 251
High-temperature (HT) storage test, 459, 590, 594–595
Hot electrons, 329, 389, 432–434, 482, 548, 573
$H_2$ profile, 254
4H-SiC. *See* 4H-silicon carbide (4H-SiC)

4H-silicon carbide (4H-SiC), 266, 268, 271–273, 275, 278, 279, 435, 556, 560, 572, 576
Huang-Rhys factor, 302
Hydrogen ($H_2$), 33, 173, 190, 224, 230, 241, 310, 322, 330, 340, 456, 457, 465–469, 477, 480, 482, 483, 486, 487, 503, 526–528, 535, 543, 584

**I**
IC modes. *See* Internally circulating (IC) modes
IID. *See* Impurity induced disordering (IID)
III-V compounds, 92–94, 140
Impurity induced disordering (IID), 125, 128, 139, 140, 142, 236
InP-based HBTs, 583–609
Intensity dependence, 265, 268, 271, 276, 279
Internally circulating (IC) modes, 124–129, 132, 134, 135, 137, 138, 141, 142
Interstitial, 45, 90–93, 100–104, 107, 109, 114–116, 118, 119, 241, 277, 278, 288, 294, 295, 309, 314, 465, 466, 527, 597, 602, 608
  atoms, 91, 92, 101–104, 107, 114, 116, 288, 597, 608
  dislocation loops, 277, 278
Inverse-piezoelectric effects, 382, 385, 403, 422, 424, 432, 434–437, 440, 523, 524, 537, 540, 548, 574
Inverse-piezoelectric stress, 385, 552, 553, 572
Inversion symmetry, 295

**J**
Jahn-Teller (J-T) effect, 285, 292, 294, 315
J-T effect. *See* Jahn-Teller (J-T) effect
Junction temperature, 98, 100, 105, 152, 159, 163, 164, 176, 433–435, 466, 548, 573, 578, 588–590, 593, 598, 605

**K**
Kink-collision case, 270
Kink-collisionless case, 270, 271
Kink migration (KM), 269–271, 275, 276, 278, 279
Kink-pair formation (KPF), 269–271, 275–279
KM. *See* Kink migration (KM)
KM barrier, 269
KPF. *See* Kink-pair formation (KPF)
KPF energy, 269

**L**
Laser diodes (LDs), 9, 29, 57, 106, 115, 123–142, 207–241, 247–260, 264, 267, 284
Laser scanning microscopy (LSM), 27, 30–31, 58, 195–197
Lasing threshold, 10, 191
Lasing wavelengths, 5, 9–12, 114
Lattice mismatch stress, 384–385, 424
Lattice relaxation, 285, 290, 294, 296, 299–303, 306, 307, 315, 437
LDs. *See* Laser diodes (LDs)
LEDs. *See* Light emitting diodes (LEDs)
Light emitting diodes (LEDs), 3–11, 13, 25, 32, 87, 88, 97–100, 104, 105, 119, 199, 207, 208, 222, 231, 247, 248, 253–255, 264, 277, 321
Lognormal plot, 179
LSM. *See* Laser scanning microscopy (LSM)

**M**
Magnesium doping, 239, 240
Materials, 4, 32, 87, 123, 150, 208, 251, 264, 284, 321, 381, 432, 456, 476, 518, 546, 583
Maverick failures, 158, 159, 169–171, 181–187, 200
MBE. *See* Molecular beam epitaxy (MBE)
Mean time between failures (MTBF), 320
Mean time to failure (MTTF), 162, 177, 220, 221, 432, 449, 458, 459, 471, 590, 591, 600, 601, 605, 607
Mechanical stress, 6, 10, 153, 272, 273, 275, 277, 330–332, 382–385, 387, 389, 395, 397, 399, 400, 403, 404, 406, 422–425, 435, 459, 516, 522, 523, 537, 539, 548, 551, 552, 554, 577
Median time to failure, 12, 16, 178
Metal-semiconductor interface, 461, 597–598, 600, 602
MgO, 476–478, 481, 483–486, 495–496, 499, 500, 502, 506–510
Minority-carrier injection, 264, 279
Molecular beam epitaxy (MBE), 118, 124, 248, 258, 476, 477, 481, 483, 495, 496, 498, 499, 506, 509, 547, 564, 584–586

MTBF. *See* Mean time between failures (MTBF)
MTTF. *See* Mean time to failure (MTTF)
Multiplication, 208, 218–221, 239, 258, 264, 267, 279, 314

**N**
Negative photoplastic effects, 266
Nonradiative multiphonon recombination, 285, 307

**O**
OBIC. *See* Optical-beam-induced current (OBIC)
Ohmic contacts, 119, 322, 413, 432–434, 436, 442, 459, 461, 465, 470, 503, 504, 548, 568, 573, 577, 584, 585
Optical absorption, 15, 123, 128, 131, 138–142, 173, 231, 238–240, 250, 253, 302, 303
Optical-beam-induced current (OBIC), 55–83, 196, 199
Optical beams, 75
Optical gain, 7, 12, 126, 255–256
Oxides, 26, 96, 150, 323, 445, 472, 477, 516, 546

**P**
Parametric fitting of degradation, 174, 175
Passivation, 230, 385, 414, 432–436, 456, 465, 466, 469, 475–510, 547, 552, 568, 570, 572, 584, 585, 587–589, 606, 607
  ledge passivation, ledge structure, 584, 588–590, 606
  SiN passivation film, 584, 588, 607
Peierls potential, 93, 269, 270, 272
PFE. *See* Poole-Frenkel emission (PFE)
Phonon-kick mechanism, 269–272, 275, 278, 308, 310–312, 315
Photodiodes, 4, 13–16, 47, 149, 162
Photoemission microscopy (EMMI), 22, 32, 199
Photoluminescence (PL), 56, 70, 82, 110, 111, 115, 210, 241, 268, 272, 274–276, 439, 448, 548, 550–552, 554, 555, 559, 560, 566–572
Piezoelectric, 330, 332, 382, 394–396, 402, 403, 413, 422, 424, 432, 434, 435, 523, 524, 537–541, 546, 551–552
Point defects, 13, 47, 89, 92, 93, 96, 99–104, 107, 109, 110, 114–118, 130, 134, 152, 171, 173, 198, 201, 221, 239–241, 260, 268, 269, 277, 284, 285, 291–293, 528, 529, 552
Polaron, 289–291
Poole-Frenkel emission (PFE), 403, 415, 420
Propagation velocity, 465

**Q**
Quantum-well lasers, 115, 134

**R**
Radiation-enhanced dislocation climb (REDC), 277–279
Radiation-enhanced dislocation glide (REDG), 263–279
Radiative recombination, 47, 257, 264
Raman scattering, 548–550, 553, 573, 574
Raman spectroscopy, 543–579
Random failure rate calculations, 180–181
Rapid degradation, 6, 86, 89–96, 103, 107, 120, 152, 154, 208, 238, 239, 267, 606
Recombination-enhanced defect reaction (REDR), 67, 104, 268–269
Recombination rates, 77, 92, 256, 257, 269, 270
Reconstruction, 268, 275, 276, 314, 445, 490
REDC. *See* Radiation-enhanced dislocation climb (REDC)
REDG. *See* Radiation-enhanced dislocation glide (REDG)
REDR. *See* Recombination-enhanced defect reaction (REDR)
Reduced activation energy, 264, 279
Refractory metal, 584, 591–603
  Mo, 591, 592, 608
  W, 591, 592, 608
Relaxation oscillation frequency, 9, 10, 129
Reliability, 4, 19, 57, 87, 124, 147, 208, 247, 284, 319, 382, 431, 455, 486, 515, 547, 584
Reliability calculations
  use of spreadsheet, 177–178
Reliability projections (Possible sources of error), 181–182

Reliability testing
VCSEL, 148, 161, 165, 166, 171
RF stress, 446–448, 450, 457
Ring-cavity modes, 128, 141

**S**
Scanning electron microscopy in cathodoluminescence mode (SEM-CL), 268
$Sc_2O_3$, 476–478, 481, 483, 485, 486, 494–495, 499–502, 509, 510
Self-trapping, 285, 289–292, 314
SEM-CL. *See* Scanning electron microscopy in cathodoluminescence mode (SEM-CL)
Semiconductor field effect transistors,
Semiconductor lasers, 3–13, 20, 28, 30, 32, 33, 40, 50, 51, 87, 88, 104, 123, 125, 140, 147, 148, 154, 162, 172, 177, 186, 187, 201, 249, 284
Semiconductor optical amplifiers, 4–5, 7, 11–12
Semiconductor optical modulators, 4, 13–15
Shockley partial dislocation, 213, 266, 267, 271, 272
Shockley stacking fault (SSF), 266, 267, 273–275, 278
Sinking gates, 322, 328–329, 351, 459
Solder voids, 125, 138
Soliton, 276, 277
SSF. *See* Shockley stacking fault (SSF)
Stabilized measurement, 390–391
Stacking faults, 22, 28, 31, 32, 36, 37, 44, 45, 95, 156, 186, 213, 266, 272, 273, 552
energy, 272, 273
Strain, 7, 59, 109, 152, 218, 273, 382, 432, 516, 546
Stress, 6, 44, 67, 93, 141, 153, 213, 254, 272, 308, 320, 383, 431, 455, 516, 548, 584
Stress-strength model, 172
Sudden failure, 6, 59–68, 83, 236–238
Surface-recombination current, 588–589
Symmetry breaking, 285, 292–294, 307, 312, 314

**T**
TED. *See* Transmission electron diffraction (TED)
TEM. *See* Transmission electron microscopy (TEM)
Temperature, junction, 98, 100, 105, 159, 163, 164, 176, 177, 433–435, 466, 548, 573, 578, 588–590, 593, 598, 605
Thermally induced voltage alteration (TIVA), 22, 27, 30–31, 184, 187, 195–199
TIVA. *See* Thermally induced voltage alteration (TIVA)
Transmission electron diffraction (TED), 587, 591, 598–599, 602, 608
Transmission electron microscopy (TEM), 20–22, 25, 32, 40–48, 60, 61, 63–66, 90, 94–95, 100–101, 103, 106, 108, 110–112, 114–116, 130, 131, 139, 140, 183, 184, 186, 188, 189, 197–199, 210–229, 232, 233, 235, 258, 265, 434–435, 439, 450, 460, 461, 463–465, 490, 493, 498, 504, 538, 587, 591, 593–598, 602–605, 608
Trap generation, 433, 457
Traps, 24, 96, 154, 258, 269, 283, 329, 382, 432, 456, 476, 520, 548

**U**
Uniaxial, 384, 386, 389, 392, 393, 396–399, 401, 404, 405, 422–424, 553, 555, 565, 571

**V**
VCSEL. *See* Vertical-cavity surface-emitting laser (VCSEL)
Vertical-cavity surface-emitting laser (VCSEL)
aging, normal, 175, 189
introductory information, 147–148

**W**
Wafer bending, 383, 385–389, 403, 404, 422–424
Wear-out
failures, 31, 57, 67–75, 83, 169–170, 187–190, 196, 239, 345, 363
source of premature failure, 188
Wear-out lifetime
due to epitaxial fluctuations, 605
sources of variation, 172–173
Window laser, 142