Politecnico di Torino

Electronic Engineering
(program in Electronic Devices)
–
29th Cycle

Ph.D. Thesis

Advances in Quantum Tunneling Models for Semiconductor Optoelectronic Device Simulation

Marco Mandurrino

Supervisors
Michele Goano
Francesco Bertazzi

Doctorate Coordinator
Marco G. Ajmone Marsan

Examination Board
Åsa Haglund
Aldo Di Carlo
Sahar Sharifzadeh
Gianluca Fiori
Enrico Bellotti
Pierluigi Debernardi

March 2017

Examination Board

1. First Member and Thesis Reviewer

Åsa Haglund
Ph.D., Associate Professor at Chalmers University of Technology, Gothenburg (Sweden)

2. Second Member and Thesis Reviewer

Aldo Di Carlo
Ph.D., Full Professor at University of Rome "Tor Vergata" (Italy)

3. Third Member

Gianluca Fiori
Ph.D., Associate Professor at University of Pisa (Italy)

4. Fourth Member

Sahar Sharifzadeh
Ph.D., Assistant Professor at Boston University (Massachusetts, USA)

5. Fifth Member

Enrico Bellotti
Ph.D., Full Professor at Boston University (Massachusetts, USA)

6. Sixth Member

Pierluigi Debernardi
Ph.D., Senior Researcher at Politecnico di Torino, CNR and IEIIT (Italy)

Preface

The core of my Ph.D. activity has been spent in the study and modeling of quantum transport effects in wide- and narrow-gap semiconductor devices. The choice of such topic came rather straightforwardly since my first degree in physics and the ensuing specialization in physics engineering in the field of nanotechnologies and device modeling. So, the interest in tunneling mechanisms arose very naturally already as undergraduate student. Anyway, I would not to dedicate more words on this topic than what I've done in the almost 2-hundred pages following this introduction. Here I wish commenting my work as Ph.D. candidate in a wider sense, by re-browsing the years passed within the Electronics and Telecommunications department.

As my first occupation, in the early 2014 I had the possibility to more deeply develop as main investigator the study started with my Master's thesis, namely the physics-based numerical simulation of direct tunneling mechanisms in reverse-biased HgCdTe p-i-n structures for infrared detection. As undergraduate student this activity dealt with purely theoretical investigations, while the first months of 2014 have been dedicated to adapt my results to real case-study devices thanks to a collaboration, in which I had the pleasure to be partially involved, between my department and the German company AIM Infrarot Module.

Very soon the efforts put into this area gave their tangible results (by the scientific standpoint) and the simulations coming also from my previous work provided several results suitable for the first publications (from here one journal paper and a number of conference contributions arose).

The first year also represented, more than the others, the year dedicated to the academic training. Both as student and as tutor. From April to June 2014, indeed, I was allowed to give my contribution to the laboratory classes about devices numerical simulation intended for undergraduate students afferent to the electronics and physics courses held by professors Simona Donati and Francesco Bertazzi. This represented one of the most enriching experiences I had during the Ph.D. because nothing gives the feeling of what it has really been learned and understood as the teaching practice.

Contemporarily and immediately after the activity on photodetectors I started my research on LEDs, again, as main investigator searching for tunneling signatures. Thanks to a collaboration with colleagues from Padua University and to the data on the first (grown-on-SiC) structure fabricated by OSRAM they provided us, I've been able to enter

into this new framework. After studying a bit of related literature and performing some conjectures and modeling activities, in the 2014 summer I yet had the opportunity to give my first two talks, at NUSOD international conference (Palma de Mallorca) and also at the annual meeting of the Italian Physical Society (SIF) held in Pisa, Italy. The NUSOD experience and the poster presented at the International Workshop on Nitrides (IWN) held in Wrocław (Poland) were fundamental since they represented the *casus* fostering the papers I wrote the following year.

The 2014, moreover, saw a big bunch of efforts spent in setting up a self-consistent library of GaN-based material properties suitable for LED simulations. The other relevant activity in which I've been involved was the supervision of a graduating student working on a thesis focusing on LED efficiency-related issues.

In 2015 the proposals to write a journal paper coming both from the NUSOD president J. Piprek and from IWN experience became realities: in the first months of the second year I wrote a paper for a special issue on nitride semiconductors of *Physica Status Solidi A* and also for another special issue, of the *Journal of Computational Electronics*, dealing with numerical simulations of GaN-based LEDs. The first work was more engineering-oriented, being related to some modeling technicalities and methodologies. Instead, the second one was especially intended for a physicist-like audience since there I presented a big section concerning the results of my theoretical studies on phonon emission effects on trapping processes related to defect-assisted tunneling transitions.

During the 2015 I enriched my formative career thanks to some courses and seminars offered by the Politecnico, as the excellence course held by Goeffrey Grossman from MIT (Boston) on nanoscale modeling, or the summer school for researchers "SCS2015" focusing on dissemination and scientific communication for which I ranked among the 40 people elected to participate after a national-based selection. At the same time I attended also external courses, as the one about defects in GaN and GaN for power electronics held in Pisa, Italy, on June 2015 and organized by Infineon Technologies and University of Pisa, or the school for researchers about detectors for high-energy physics experiments organized by the Italian Institute of Nuclear Physics (INFN). In the second year I was also one of the nine people selected by our Doctoral School among all the Ph.D. students in Politecnico to carry out a part-time collaboration activity for an overall amount of 100 hours which was dedicated to assist the management of doctoral didactics.

By the research standpoint, 2015 saw again a little contribution of mine on infrared photodetectors with the analysis of impact ionization modeling strategies. But the most intensive efforts have been directed towards the comprehension of trap-assisted tunneling by heavy holes in LEDs grown-on-Si (again, provided by OSRAM and characterized in Padua). Here the most advanced data analysis I conducted on LEDs measurements saw the light, is appropriate to say. Then, in this second year I worked also within a certain number of collateral research areas: for instance, this is the period in which I gave my contribution to a collaborative study with colleagues from Padua about defects implications on LED reliability and robustness that resulted in a couple of publications totally uncorrelated to the topic of tunneling: I collaborate with M. Meneghini and his group for a 2015 paper published in the *Microelectronics Reliability* journal and I was also involved in a conference contribution presented at the 26th Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF2015). Also uncorrelated to any tunneling phenomenon is the contribution I gave within the area of LED efficiency droop. Here the publications were numerous: as an example I report here the paper written in 2015 and published the

following year in the *Journal of Applied Physics*, as reported in the list of my publications.

Also in 2016 I was selected (resulting the first ranked) to carry out the same part-time collaboration activity I gave the year before, again, for a total budget of 100 hours. By the scientific standpoint, the last year has been connoted by a change of perspective in my research on tunneling. Moreover, having the unique opportunity to meet and talk with the Nobel Laureate in Physics, Professor H. Amano, at the GaN Marathon held in Padua on April 2016 has been illuminating for my consideration about the future of GaN in solid-state electronics.

After being involved also in several methodological studies where different commercial simulation tools have been compared for what concerns classical and quantum physical models (some results can be found in the contributions presented on February 2016 at the 20th SPIE International Conference on LEDs in San Francisco and at other minor events), already since the last months of 2015 my research focus – albeit still tunneling-related – moved from semiclassical to full-quantum modeling approach. In this context I started to be increasingly involved in the development of an in-house code for NEGF simulations which has led me to explore the world of advanced programming: so, I attended a course on parallel computing organized by the CINECA university consortium. To this purpose, my third year of Ph.D. was definitely oriented towards the genuine quantum simulations, both via NEGF method and also by developing hand-made codes for the solution of confinement problems in nanostructures, as explained in the very last part of this thesis.

Marco Mandurrino,
Pinerolo, October 2016

Declaration

This dissertation is presented in partial fulfillment of the requirements for the Ph.D. degree in Electronic Engineering, program in Electronic Devices, at the doctoral school "ScuDo" (Scuola di Dottorato) of Politecnico di Torino.

I declare that all the content and its organization within the present dissertation constitutes the result of my own original work and, for this reason, it does not compromise in any way and form the rights of third parties, including those relating to the security of personal data.

It is also worth pointing out that all figures – as well as data published in tables – have been originally developed by the author or, otherwise, freely inspired by their source and then originally re-interpreted and re-edited for personal purposes, as always specified in the related captions.

Marco Mandurrino,
Pinerolo, October 2016

Contents

Abstract

The undiscussed role of solid-state optoelectronics covers nowadays a wide range of applications. Within this scenario, infrared (IR) detection is becoming crucial by the technological point of view, as well as for scientific purposes, from biology to aerospace. Its commercial and strategic role, however, is confirmed by its spreading use for surveillance, clinical diagnostics, environmental analysis, national/private security, military purposes or quality control as in food industry. At the same time solid-state lighting is emerging among the most efficient electronic applications of the modern era, with a billion-dollar business which is just destined to increase in the next decades.

The ongoing development of such technologies must be accompanied by a sufficiently fast scientific progress, which is able to meet the growing demand of high-quality production standards and, as immediate but not obvious consequence, the need of performances which would be the highest possible. One issue affecting both kinds of applications we mentioned is the quantum efficiency, no matter the signal they produce is coming from absorbed or emitted photons. At any rate, the balance between the stimulus coming from the surrounding environment is and the generated electrical current is absolutely crucial in each modern optoelectronic device.

More in depth, since IR detectors are asked to convert photons into electrons, device designers must ensure that mechanisms concurring to this conversion should be dominant with respect to any opponent phenomenon. Symmetrically, light-emitting diodes should realize the inverse process, where electrons are converted into photons. In real life this mechanism never take place in a one-to-one electron-photon correspondence. Indeed tunneling, a quantum effect related to the probabilistic nature of particles and, thus, also of charges, contributes to unbalance this correspondence by degrading the signal produced within the device active region. In IR photodetectors this translates into of a current even in absence of light (and, by virtue of this fact, this current is known as "dark current") while in light-emitters tunneling is responsible for leakages that may undermine the quantum efficiency and the power consumption also below the optical turn-on.

The present dissertation is part of such framework being the result of studying and modeling different tunneling mechanisms occurring in narrow-gap infrared photodetectors (IRPDs) for mid-wavelength IR (MWIR) applications (3 to 5 μm) and in wide-gap blue LEDs (around 450 nm) based on nitride material system. This study has been possible

thanks to the collaboration with several academic institutions (Boston University, Padua and Modena e Reggio Emilia Universities) and two important German industries, AIM Infrarot Module and OSRAM Opto Semiconductors, which provided the case-study devices here analyzed. After reviewing basic concepts of solid-state physics, the first part of this work deals with the description of the above cited optoelectronic devices, along with their constituent materials: the HgCdTe alloy, in the case of photodetectors, and GaN and its ternary alloys with In and Al, for what concerns blue LEDs.

Since the literature focusing on this research area is still not mature enough, in the second part different tunneling mechanisms and models are proposed, described in detail and then tested for the first time, as in the case of a novel formulation intended for direct tunneling in IRPDs or the description of defect-assisted tunneling in LEDs which also includes elements coming from the microscopic theory of multiphonon emission (MPE) in solids. Simulations are carried out by means of several numerical simulation approaches, using either commercial TCAD (Technology Computer Aided Design) tools and codes developed *ad hoc* for this purpose.

The encouraging and fully satisfying results of numerical modeling here proposed confirm, on the one hand, the widely accepted relevance of tunneling in modern electronics and, on the other hand, also propose a new perspective about possible tunneling mechanism in optoelectronic devices and their appropriate physical, mathematical and numerical investigation tools. Furthermore, the role of device modeling does not end here because many physical details and technological information can be inferred from simulations, with enormous beneficial effects for the electronic industry and the quality improvement of its fabrication processes such those invoked above.

Acknowledgments

I wish to primarily thank all my family and the people who supported me in the last, unrepeatable, three years of my life: Aurora, Matteo, my brother, and – above all – my parents. They have always endorsed me... and also endured me. Either in more and less positive days (and nights). Without their help I would not be the person I am, even professionally. To them I dedicate this work.

Strictly concerning the work, a thanks must inevitably go to the Compagnia di San Paolo, without whose financial support I could not carry out my Ph.D. program. I would also remember that my activity was in part supported by the U.S. Army Research Laboratory (ARL) through the Collaborative Research Alliance (CRA) for MultiScale multidisciplinary Modeling of Electronic materials (MSME). Immediately after, I'm grateful to my Department of Electronics and Telecommunications (DET), in Politecnico di Torino, for all logistic, financial and formative support, then also to my Ph.D. supervisors and colleagues.

Finally, my gratitude also goes to all the external colleagues: Giovanni Verzellesi, to whom goes the merit of initiating me into LEDs simulation and with whom I had interesting discussions about tunneling in these devices; then also to our collaborators from Boston University – in the person of Enrico Bellotti – for part of the financial support and from Padua University, Matteo Meneghini and his coworkers, for their precious experimental data and for prompt feedback on technological issues and lastly, but not by importance, to the German companies AIM Infrarot Module GmbH and OSRAM Opto Semiconductors GmbH, that provided us the structures and the authorization to publish upon them, and their representatives who have been our co-authors in several papers.

List of Publications

Journal Papers

[J1] C. De Santi, M. Meneghini, M. La Grassa, B. Galler, R. Zeisel, M. Goano, S. Dominici, M. Mandurrino, F. Bertazzi, D. Robidas, G. Meneghesso, and E. Zanoni, "Role of defects in the thermal droop of InGaN-based Light Emitting Diodes," *J. Appl. Phys.*, vol. 119, pp. 094501, 2016.

[J2] M. La Grassa, M. Meneghini, C. De Santi, M. Mandurrino, M. Goano, F. Bertazzi, R. Zeisel, B. Galler, G. Meneghesso, and E. Zanoni, "Ageing of InGaN-based LEDs: effects on internal quantum efficiency and role of defects," *Microelectron. Reliability*, vol. 55, no. 9–10, pp. 1775–1778, 2015.

[J3] M. Vallone, M. Mandurrino, M. Goano, F. Bertazzi, G. Ghione, W. Schirmacher, S. Hanna, and H. Figgemeier, "Numerical Modeling of SRH and Tunneling Mechanisms in High Operating Temperature MWIR HgCdTe Photodetectors," *J. Electron. Mater.*, vol. 44, no. 9, pp. 3056–3063, 2015.

[J4] M. Mandurrino, M. Goano, M. Vallone, F. Bertazzi, G. Ghione, G. Verzellesi, M. Meneghini, G. Meneghesso, and E. Zanoni, "Semiclassical simulation of trap-assisted tunneling in GaN-based light-emitting diodes," *J. Comput. Electron.*, vol. 14, no. 2, pp. 444–455, 2015 [Special Issue Invited Paper].

[J5] M. Mandurrino, G. Verzellesi, M. Goano, M. Vallone, F. Bertazzi, G. Ghione, M. Meneghini, G. Meneghesso, and E. Zanoni, "Physics-based modeling and experimental implications of trap-assisted tunneling in InGaN/GaN light-emitting diodes," *Phys. Status Solidi A*, vol. 212, no. 5, pp. 947–953, 2015 [Special Issue].

Conference Papers (*speaker)

[C1] C. De Santi, M. Meneghini, M. La Grassa, B. Galler, R. Zeisel, M. Goano, S. Dominici, M. Mandurrino, F. Bertazzi, G. Meneghesso, and E. Zanoni, "Investigation of the thermal droop in InGaN-layers and blue LEDs," in *International Workshop on Nitride Semiconductors (IWN 2016)*, Oct 2016, Orlando (FL, USA).

[C2] M. Goano, F. Bertazzi, X. Zhou, M. Mandurrino, S. Dominici, M. Vallone, G. Ghione, A. Tibaldi, M. Calciati, P. Debernardi, F. Dolcini, F. Rossi, G. Verzellesi, M. Meneghini, N. Trivellin, C. De Santi, E. Zanoni, and E. Bellotti, "Simulation of GaN-based active optoelectronic devices: semiclassical models and beyond," in *Gallium Nitride Technology in Europe*, GaN Marathon, Apr 2016, Padova (Italy) [Invited Paper].

[C3] C. De Santi, M. Meneghini, M. La Grassa, N. Trivellin, B. Galler, R. Zeisel, B. Hahn, M. Goano, S. Dominici, M. Mandurrino, F. Bertazzi, G. Meneghesso, and E. Zanoni, "Thermal droop in InGaN-based LEDs: physical origin and dependence on material properties," in *Proc. SPIE* 9768, Light-Emitting Diodes: Materials, Devices, and Applications for Solid State Lighting XX, OPTO Photonics West, Feb 2016, San Francisco (CA, USA), pp. 97680D–97680D-8.

[C4] M. Goano, F. Bertazzi, X. Zhou, M. Mandurrino, S. Dominici, M. Vallone, G. Ghione, F. Dolcini, F. Rossi, G. Verzellesi, M. Meneghini, E. Zanoni, and E. Bellotti, "Challenges towards the simulation of GaN-based LEDs beyond the semiclassical framework," in *Proc. SPIE* 9742, Physics and Simulation of Optoelectronic Devices XXIV, OPTO Photonics West, Feb 2016, San Francisco (CA, USA), pp. 974202–974202-11 [Invited Paper].

[C5] G. Verzellesi, M. Bernabei, L. Rovati, M. Mandurrino, F. Bertazzi, M. Goano, M. Meneghini, G. Meneghesso, and E. Zanoni, "Efficiency droop in Gallium Nitride based light-emitting diodes: physical mechanisms and technological remedies," in *EMN Meeting*, Dec 2015, Guangzhou (China) [Invited Paper].

[C6] M. La Grassa, M. Meneghini, C. De Santi, M. Mandurrino, M. Goano, F. Bertazzi, R. Zeisel, B. Galler, G. Meneghesso, and E. Zanoni, "Ageing of InGaN-based LEDs: effects on internal quantum efficiency and role of defects," in *26th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF)*, Oct 2015, Toulouse (France).

[C7] F. Bertazzi, S. Dominici, M. Mandurrino, D. Robidas, X. Zhou, M. Vallone, M. Calciati, P. Debernardi, G. Verzellesi, M. Meneghini, E. Bellotti, G. Ghione, and M. Goano, "Modeling challenges for high-efficiency visible light-emitting diodes," in *Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), 2015 IEEE 1st International Forum on*, Sept 2015, Turin (Italy), pp. 157–160.

[C8] M. Mandurrino, M. Goano, S. Dominici, M. Vallone, F. Bertazzi, G. Ghione, M. Bernabei, L. Rovati, G. Verzellesi, M. Meneghini, G. Meneghesso, and E. Zanoni, "Trap-assisted tunneling contributions to subthreshold forward current in InGaN/GaN LEDs," in *Proc. SPIE* 9571, Solid State Lighting and LED-based Illumination Systems XIV, Optical Engineering + Applications, Aug 2015, San Diego (CA, USA), pp. 95710U–95710U-6.

[C9] F. Bertazzi, S. Dominici, M. Mandurrino, D. Robidas, X. Zhou, M. Vallone, G. Verzellesi, M. Meneghini, G. Meneghesso, E. Zanoni, E. Bellotti, G. Ghione, and M. Goano, "Transport modeling challenges for GaN-based light-emitting diodes," in *GE 2015 Annual Meeting*, Jun 2015, Siena (Italy).

*[C10] M. Mandurrino, G. Verzellesi, M. Goano, S. Dominici, F. Bertazzi, G. Ghione, M. Meneghini, G. Meneghesso, and E. Zanoni, "Defect-related tunneling contributions to subthreshold forward current in GaN-based LEDs," in *Fotonica AEIT Italian Conference on Photonics Technologies, 2015*, May 2015, Torino (Italy), pp. 1–4.

[C11] M. Vallone, M. Mandurrino, M. Goano, F. Bertazzi, G. Ghione, W. Schirmacher, S. Hanna, and H. Figgemeier, "Contributions to the dark current in HOT MWIR HgCdTe photodetectors: experiments and numerical modeling," in *42nd Freiburg Infrared Colloquium*, Mar 2015, Freiburg (Germany).

[C12] M. Vallone, M. Mandurrino, M. Goano, F. Bertazzi, G. Ghione, W. Schirmacher, S. Hanna, and H. Figgemeier, "Numerical modeling of SRH and tunneling mechanisms in HOT MWIR HgCdTe photodetectors," in *II-VI US Workshop, on the Physics and Chemistry of II-VI materials*, Oct 2014, Baltimore (MD, USA).

*[C13] M. Mandurrino, G. Verzellesi, M. Goano, M. Vallone, F. Bertazzi, G. Ghione, M. Meneghini, G. Meneghesso, and E. Zanoni, "Tunneling elettronico in strutture LED in InGaN/GaN: analisi sperimentale e simulazioni numeriche," in *100° Congresso Nazionale Società Italiana di Fisica (SIF 2014)*, Sept 2014, Pisa (Italy).

*[C14] M. Mandurrino, G. Verzellesi, M. Goano, M. Vallone, F. Bertazzi, G. Ghione, M. Meneghini, G. Meneghesso, and E. Zanoni, "Trap-assisted tunneling in InGaN/GaN LEDs: experiments and physics-based simulation," in *Numerical Simulation of Optoelectronic Devices (NUSOD), 2014 14th International Conference on*, Sept 2014, Palma de Mallorca (Mallorca, Spain), pp. 13–14.

[C15] M. Mandurrino, G. Verzellesi, M. Goano, M. Vallone, F. Bertazzi, G. Ghione, M. Meneghini, G. Meneghesso, and E. Zanoni, "Trap-assisted forward tunneling current in InGaN/GaN LEDs: experiments and physics-based simulation," in *International Workshop on Nitride Semiconductors (IWN 2014)*, Aug 2014, Wrocław (Poland).

Bibliometric Data
Citations: **44**
H-index: **3**
i10-index: **2**
RG Score: **13.61***
Most cited paper: J5 (**15** citations)
Highest journal IF (2016): **2.04**, *Journal of Applied Physics* (J1)
Sources: https://scholar.google.it and *https://www.researchgate.net (Mar 2017)

≪Nature, as we understand it today, behaves in such a way that it is fundamentally impossible to make a precise prediction of exactly what will happen in a given experiment≫

Richard P. Feynman — "Six Easy Pieces" (1964)

General Introduction

For many decades electronics considered carriers as classical particles. The real breakthrough came, probably, in the early 90s with the work done by C. Zener, who realized that electronic devices could no longer be considered as classical systems due to the presence of excess currents unpredicted by the standard transport models: for the first time the technology and a newborn quantum mechanics came into contact, and real-life objects experienced effects that only laboratories did before. This dissertation deals with quantum phenomena in solid matter that are at the origin of such kind of currents and, in particular, with tunneling in electronic nano-systems for optoelectronic applications.

In the next paragraphs we shall introduce the framework in which the present work is collocated and then a brief perspective about the organization of the presented topics will be also provided.

I Motivation and Research Focus

The rapidly growing field represented by lighting and light detection is known to be affected by the quantum probabilistic nature of carriers in solids, especially when devices dimensions tend to be reduced down to the nanoscale. Moreover, photon emission or absorption, more than other areas of interest as power electronics or digital ones, requires a certain signal quality, besides standard noise controlling. Let us mention, for instance, night vision for security purposes or the use of particular emission wavelengths to reproduce specific natural habitats essential for the artificial breeding of certain fauna.

Quantum tunneling represents, seen from the optoelectronics applications perspective, a cause of current leakage. Thus it may heavily contribute to the signal degradation, no matter whether is coming from emitted or received photons. Nonetheless, the rigorous description of these effects is not yet mature like for silicon technology, where tunneling is under study since several decades: think, for instance, to all the developments made about direct or Fowler-Nordheim tunneling

control in MOSFETs. In some other applications, moreover, tunneling became a resource, an underlying operation principle, like for resonant tunneling diodes (RTDs) or tunneling-field-effect transistors (TFETs).

The present dissertation has been developed with the aim of trying to fill, in part, this gap through the study and modeling of different tunneling-related phenomena occurring in different kinds of optoelectronic devices, by means of different theoretical and analytical approaches offered by the computational physics applied to the simulation of electronic devices. To do so, the author focused both on photodetectors (in our case intended to detect IR light) and light-emitters (here GaN-based LEDs have been studied). Since the materials constituting our devices highly differ from the well-known silicon and its properties this research activity has been a bit challenging and – for such reason – extremely exciting and mentally stimulating.

In order to make this work as clear as possible, we next present the thesis plan, discussing how it articulates throughout the chapters sequence.

II About this Work

This work is divided into two main parts, plus a section which includes three important appendices that will be described in the following. Let us start from Part I, concerning "*Fundamentals of Solid-State Optoelectronic Devices*".

After a very short, but necessary, introduction about some basic concepts concerning semiconductors, theory of bands and carrier statistics/dynamics, Chapter 1 also provides several important information about one of the most relevant category of physical processes occurring in semiconductor devices: generation/recombination (GR) mechanisms. This is absolutely propaedeutical to the continuation of all the thesis since tunneling mechanisms could involve a GR event or, they could be somehow assimilated to a GR process as net result of a carrier transition across the band gap. The last part of this chapter also deals with a brief recall about one of the most used model among classical theories about carrier transport in semiconductors: the drift-diffusion (DD) formalism and its computational/mathematical properties, as boundary conditions and geometrical discretization of space-time domain. This formalism will be used in the following chapters when semiclassical tunneling models will be explained and applied.

Chapter 2 is intended to introduce the theory of solid-state infrared (IR) photon detection along with a complete description of all the physical properties of the elective material system used in today IR industry: the narrow- and tunable-gap semiconducting ternary alloy HgCdTe, also known as "Mer-Ca-Tel", or MCT. These first two topics correspond to as many subsection of the chapter and, in part, have their origin in the research work done by the author during the Master's thesis development, which was focusing on a theoretical investigation of direct tunneling within (unrealistic) MCT-based *p-i-n* structures. The chapter concludes with a description of modern devices and of the most relevant transport properties of which they are characterized. Knowing the standard behavior of these objects is crucial in order to recognize the famous excess current we mentioned at the beginning

which is referred, in our case, to particular (direct) tunneling transitions occurring in HgCdTe-based devices.

Chapter 3 opens retracing the history of solid-state lighting, from first attempts with macroscopic samples made of SiC in the first years of the 20th century to the modern era, dominated by light-emitting diodes (LEDs) and laser diodes. Then we immediately focus on our area of interest, represented by GaN-based blue LEDs, by describing the properties of narrow-gap III-nitride semiconductors (essentially GaN and its ternary alloys made with Al and In used as Ga-substitutional) and also by analyzing some important LED operation principles. These topics deal with a new research area in which the author worked during its Ph.D. activities: in particular, setting up and calibrating a material library (here reported) has represented one of the main prerequisite to be accomplished before entering more deeply in tunneling analysis and modeling. The chapter concludes by treating some advanced descriptions of quantum confinement in LED active regions and also about light emission and efficiency-related issues, of which also tunneling makes part.

The first three chapters complete Part I of the present dissertation. Part II, or "*Tunneling in Direct Band Gap Semiconductor Devices*", represents the core of the author's scientific production, meant as relevant theoretical results and their practical impact on published studies over real devices. However, as pointed out in the Preface, we are not talking about *all* the scientific production, but only of that one concerning the main topic, which is tunneling.

Chapter 4 is surely the most important of all the dissertation. Its articulation is quite complex and its purpose is to bring the reader from pure theoretical principles of quantum mechanics to the art of modeling them with a calculator. The passage is conceptually rather crucial since the physics must necessarily be simplified in order to be modeled and the way of doing this finally determines the accuracy level of the adopted strategy and, substantially, even the outcoming results of the overall modeling procedure. For these simple but important reasons it has been decided to organize the treatment starting from one of the most used and, at the same time, also the most effective approximations that can be adopted, known as the semiclassical approximation of quantum mechanics or "WKB approximation", from the names of the three scientists, G. Wentzel, H. A. Kramers and L. N. Brillouin, who separately came to the same formalism, in 1926. From here we shall progressively move towards more accurate descriptions, by shrinking the observation scale and decreasing the number of approximations. Being the WKB formalism at the basis of a wide part of scientific literature dealing with the modeling of quantum particles within confining potentials in solids, its analysis becomes even more useful for us. So, after the initial theoretical section, shall be able to present through the WBK perspective two important tunneling effects, i.e. the direct (BTBT) and trap-assisted tunneling (TAT). Then some of the most relevant models for both processes are browsed and reviewed: we are talking about the rigorous theory of BTBT by E. O. Kane and the widely-used TAT formalism developed by G. A. M. Hurkx. Since the target here is to deviate from standard framework towards *ad hoc* applications conceived for not fully conventional materials as MCT or III-N, the treatment continues presenting a novel BTBT formulation intended for MCT-based *p*-*i*-*n* structures to which the au-

thor began working in 2013 thanks to the above mentioned Master's thesis and here fully developed. The model takes origin from Kane's studies and introduces some new elements allowing its application to realistic mid-wavelength-infrared (MWIR) HgCdTe-based photodetectors under opportune conditions about the electric field across the diode junction. For what concerns TAT, a third section focuses on a microscopic general theory here supposed to be valid to describe also defect-assisted tunneling in wide-gap light-emitters: this theory is referred to as the multiphonon-emission (MPE) theory. The reader can find its derivation as a theoretical introduction of the 2015 paper published in the *Journal of Computational Electronics*, vol. 14 no. 2. On the basis of such rigorous approach, some new features have been described and then added – as in the case of a WKB-related probabilistic term or as for nonlocal-effects terms – in order to obtain a formulation of trapping rates suitable for semiclassical simulations of TAT in GaN-based blue LEDs. Still moving towards the microscopic world, the chapter closes with a brief but intense description of two of the most advanced full-quantum modeling approaches for tunneling mechanisms in semiconductor devices: the first class is represented by the so-called density-matrix approach (here only described) while the second one concerns the promising Non-Equilibrium Green's Functions (NEGF) method, which has been described both by the theoretical and computational standpoint. This method will be used in the following when a code developed by our group will be applied to a realistic GaN-based structure.

After the digression about tunneling mechanisms and models, Chapter 5 begins the series of applicative chapters dedicated to the results of simulations obtained including all the previously exposed theories in real optoelectronic structures. Here the case-study involves two HgCdTe IR photodetectors (IRPDs) manufactured by AIM Infrarot Module GmbH which are nominally identical except for their p-type doping technology. After a parametric calibration, the measured dark characteristics of both devices have been successfully reproduced with semiclassical simulations which included the author's novel BTBT formulation and also suitable impact ionization models. The second part of the chapter is somehow a review of the publications in this field and, in particular, of one journal paper (*Journal of Electronic Materials*, vol. 44 no. 9, 2015) and two conference papers (presented at the 2014 US Workshop on II-VI materials, Baltimora, and at the 2015 Infrared Colloquium, Freiburg). Besides the review, even some unpublished results are included. The author would thank the collaboration with AIM Infrarot Module and its researchers (among all, W. Schirmacher, S. Hanna and H. Figgemeier) and also with Boston University (especially Prof. E. Bellotti, from the Electrical & Computer Engineering group), without which this study could not have been developed.

Another collaboration with a well-known German company, OSRAM Opto Semiconductors GmbH, is at the center of my second research area. Thanks to two InGaN/GaN single quantum well (SQW) blue LEDs fabricated by such company, Chapter 6 analyzes the structures searching for tunneling signatures and then describes the results of semiclassical TAT simulations performed accounting for the MPE model previously discussed which demonstrate, for the first time in the literature, its efficacy in reproducing forward electrical characteristics of such GaN-based

light-emitters. Besides the defect-related transport across the band gap, the present thesis proposes the existence of a further tunneling mechanism occurring between quasi confined states within the quantum well (QW) and the surrounding barriers: the intra-band-to-band tunneling (BTBT), a direct transition involving only one band, conduction for electrons or valence for holes. The formalism chosen in order to investigate this hypothesis is the full-quantum NEGF method since, in the author's opinion, only a genuine microscopic description could capture the quantum essence of a mechanism involving quasi-confined fermions. After choosing a further (realistic) case-study structure, the NEGF code developed by our Computational Electronics group is tested, confirming – at least via simulations – the presence of the inferred intra-BTBT mechanism. In summary, this chapter essentially proposes, in a completely rearranged form, all the studies published about tunneling in light-emitters, also including some unpublished analysis and results: in particular, at least two special-issue journal papers (*Physica Status Solidi A*, vol. 212 no. 5, and *Journal of Computational Electronics*, vol. 14 no. 2, both dated 2015) and as many conference papers (presented at the 2014 International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD), Palma de Mallorca, and at the 14th SPIE International Conference on Solid-State Lighting and LED-based Illumination Systems, San Diego) plus other less prestigious but even important national/international conference papers. Besides OSRAM Opto Semiconductors, this research was possible also thanks to our collaboration with Padua University, Italy (and particularly with Prof. E. Zanoni, G. Meneghesso and Dr. M. Meneghini) and Prof. G. Verzellesi from Modena e Reggio Emilia University, Italy.

Chapter 7, finally, reports some useful global comments and considerations about this work. Then a section dedicated to three Appendices has been included. In Appendices A and B we described the way of implementing the most relevant material parameters of MCT and III-nitrides, respectively, into the simulator used for semiclassical numerical modeling. Lastly, Appendix C presents the C++ code used to perform simulations via the author's BTBT formulation: being a novel model, the commercial simulator employed for semiclassical analysis needs to receive the routine as external file. So, for the sake of completeness (and also to give reproducibility to the present work) we reported it for the reader benefit.

Part I

Fundamentals of Solid-State Optoelectronic Devices

Chapter 1

General Concepts concerning Semiconductors

Before introducing the two main categories of optoelectronic devices presented in this thesis, that is the narrow gap photodiodes for infrared light detection (see Chapter 2) and the wide gap, GaN-based, light-emitting diodes (see Chapter 3), a general introduction about classical solid-state physics in semiconductors must be provided. To this purpose we present here a propaedeutic description organized as follows: Section 1.1 is devoted to recalling the origins of energetic band structure, the fundamental ingredient for each electronic transition, as tunneling mechanisms are. A simple derivation of the $\mathbf{k} \cdot \mathbf{p}$ method, one of the most exploited in order to predict the energy band structure in semiconductors, is also included. Then, among these transitions, Shockley-Read-Hall, radiative and Auger processes, which are the most important occurring in both devices categories, will be the topic of Section 1.4. As we shall see, both mechanisms dynamics and the equations commonly used to represent them in the classical framework will be described.

1.1 From Orbitals to the Theory of Bands

According to the Bloch theorem, electrons in solid matter experiencing a periodic lattice potential $U(\mathbf{r})$ has a wavefunction of the form

$$\Psi_{n,\mathbf{k}}(\mathbf{r}) = u_{n,\mathbf{k}}(\mathbf{r})\, \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}} \tag{1.1}$$

(see panel (a) in Figure 1.1), where $\mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}$ is a propagating plane wave and the term $u_{\mathbf{k}}(\mathbf{r})$ has the same period of $U(\mathbf{r})$: if $\mathbf{R}$ is the unitary vector in the real space, $\mathbf{K}$ its reciprocal (i.e., referred to the lattice in the $\mathbf{k}$-space) and $U(\mathbf{r}+\mathbf{R}) = U(\mathbf{r})$ by construction, than also $u_{\mathbf{k}}(\mathbf{r}+\mathbf{R}) = u_{\mathbf{k}}(\mathbf{r})$ and $\Psi_{\mathbf{k}+\mathbf{K}}(\mathbf{r}) = \Psi_{\mathbf{k}}(\mathbf{r})$ are automatically verified.

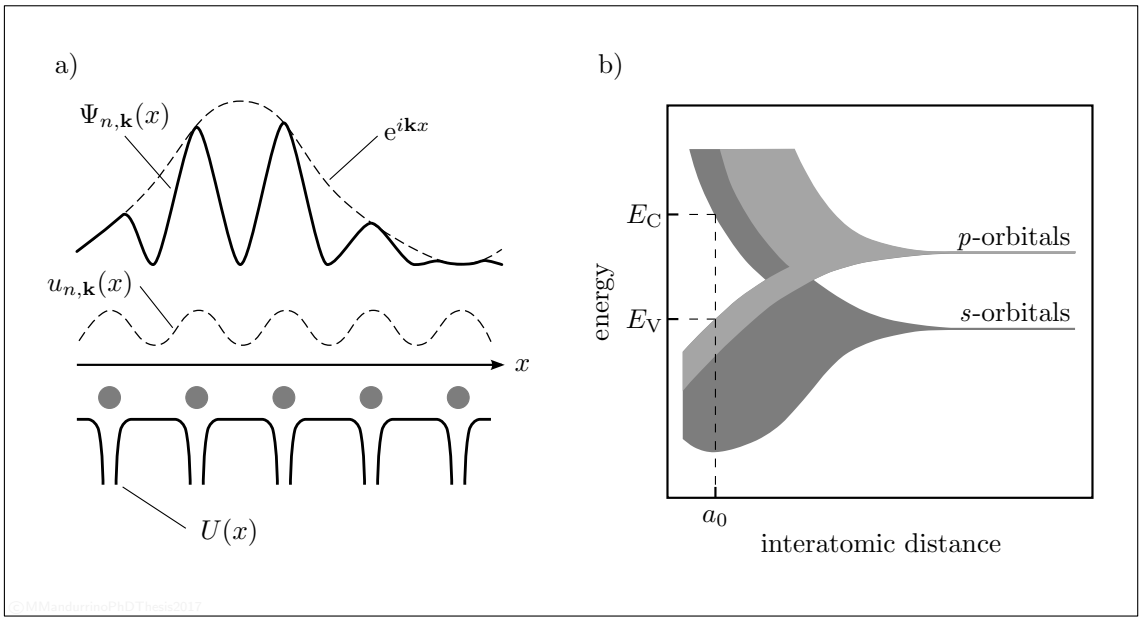

FIGURE 1.1. (a) Graphical representation of a generic monodimensional Bloch-wavefunction $\Psi_{n,\mathbf{k}}(x) = u_{n,\mathbf{k}}(x)\, e^{i\mathbf{k}\cdot x}$ and its components: the slowly oscillating plane wave $e^{i\mathbf{k}\cdot x}$ and the lattice-period function $u_{n,\mathbf{k}}(x)$. Grey dots are crystal nuclei and $U(x)$ is the relative 1D lattice periodic potential (Kronig-Penney). (b) Energy bands formation in a tetrahedral crystal cell (Si or Ge): as one can see from the plot representing the transformation of electronic energy levels as a function of the interatomic distance, orbitals may overlap and degenerate into bands which open at the equilibrium radius a_0 determining a typical forbidden gap $E_g = E_C - E_V$.

The first observation about Eq. (1.1) is that the wavefunction $\Psi_{n,\mathbf{k}}(\mathbf{r})$ cannot be any possible function but it must be eigensolution of the (time-independent) Schrödinger equation

$$-\frac{\hbar^2}{2m^*}\Delta_{\mathbf{r}}\Psi_{n,\mathbf{k}}(\mathbf{r}) + U(\mathbf{r})\Psi_{n,\mathbf{k}}(\mathbf{r}) = E_n(\mathbf{k})\Psi_{n,\mathbf{k}}(\mathbf{r})\,, \tag{1.2}$$

where $\hbar$ is the reduced Plank constant, m^* the carrier effective mass,

$$E_n(\mathbf{k}) \simeq E_n + \frac{\hbar^2\mathbf{k}^2}{2m^*} \quad \text{with } n = 1, 2, 3, \ldots \tag{1.3}$$

(n is the principal quantum number of the state/sub-band) and

$$H = -\frac{\hbar^2}{2m^*}\Delta_{\mathbf{r}}(\mathbf{r}) + U(\mathbf{r}) \tag{1.4}$$

is the Hamiltonian operator of the system. The other important consideration has to do with the presence of the subscript $\mathbf{k}$, and consists in the fact that we can find a completely different solution of Eq. (1.1) by changing the wavevector $\mathbf{k}$: the wavevector here becomes a true quantum number, labeling in Eq. (1.2) the complete basis of Bloch-wavefunctions.

As a consequence of the Bloch theorem, also energy eigenvalues are periodic as $\Psi_{n,\mathbf{k}}(x)$ in the reciprocal lattice, so:

$$E_n(\mathbf{k} + \mathbf{K}) = E_n(\mathbf{k})\,. \tag{1.5}$$

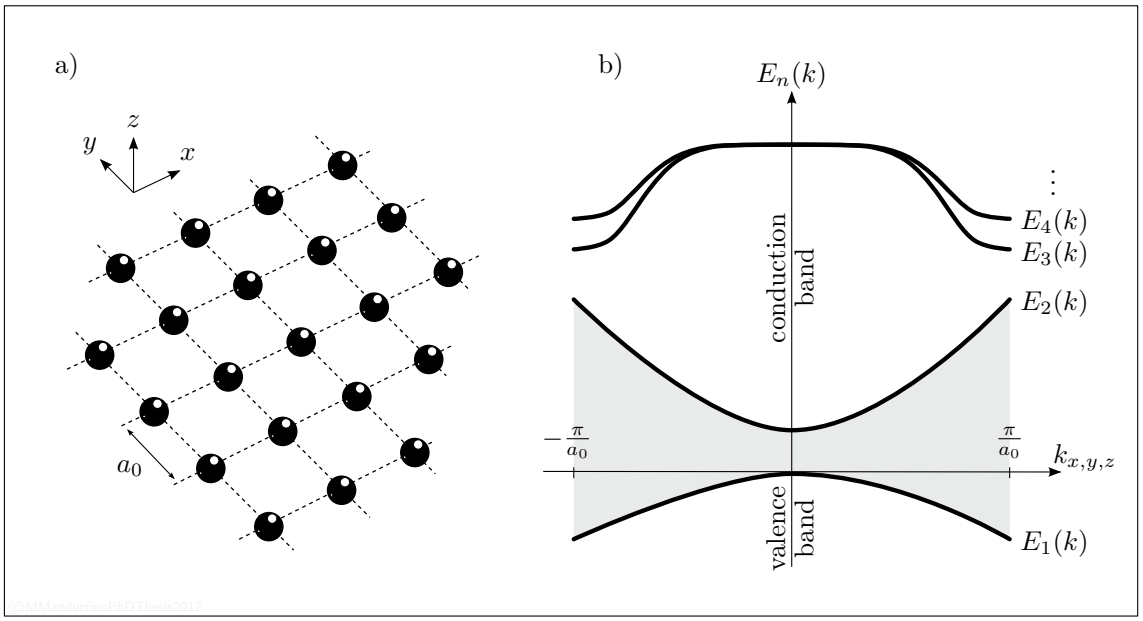

FIGURE 1.2. (a) Simplified 2D representation of a periodic cubic crystal, where the lattice parameter a_0 is highlighted. (b) Corresponding band structure of a direct band gap material in the first Brillouin zone (FBZ), i.e. for $|k| \leq \frac{\pi}{a_0}$, where k is the crystal wavevector $\mathbf{k}$ projected along one of the three axes of the real space. First conduction and valence band states $E_n(k)$ around the Γ symmetry point ($\mathbf{k} = 0$) are also represented, while the grey area is the band gap.

These results represent a remarkable difference from what one can observe in free atoms, where the electronic potential of nuclei determine a well-defined spectrum of energy levels and wavefunctions, whatever the value $\mathbf{k}$. In case of solids, valence electrons are seen as an ensemble of fermions whose wavefunction is represented by a term $u_{n,\mathbf{k}}(\mathbf{r})$ with the same period of lattice potential (Kronig-Penney potential [1]) modulated by a slowly oscillating plane wave $e^{i\mathbf{k}\cdot\mathbf{r}}$. The energy spectrum now depends on the lattice momentum $\hbar\mathbf{k}$ and is no more a set of discrete spectrum of possible states but the orbital levels result to be degenerated into a system of allowed (or forbidden) "bands", where states are energetically dense. When $\mathbf{k}$ assumes real values the band is allowed, otherwise we are in presence of forbidden states, since there the eigenfunctions are exponentially decaying. The framework we developed is at the basis of the *energy band theory*, a formalism merging quantum mechanics and solid-state physics in a unique model which is able to predict the behavior of electrons in a periodic crystal.

Now we have the fundamental elements to investigate the energy structure of each solid. What makes the difference between a descriptive model and another one is the way of approximating the lattice potential (see, for instance, the *virtual-crystal approximation* in Section 2.2.2). Rigorous approaches consist in the *tight-binding approximation* and the $\mathbf{k}\cdot\mathbf{p}$ *perturbation method*: the first one describes the system accounting for atomic orbitals of only nearby atoms, and the second one is based on an Hamiltonian with perturbation terms in the momentum space. Before entering in the details of this last method, let us start from the bands formation (see panel

(b) in Figure 1.1). Typically, in direct band gap semiconductors, atoms are connected through σ-bonds, consisting in a superposition of p-orbitals in a three-fold sp^3 hybridization (p-lobes are lying on all the three directions of the real space: p_x, p_y, p_z). These bonds give the major contribution to the shape of $u_{n,\mathbf{k}}(\mathbf{r})$ and the related probability density $|u_{n,\mathbf{k}}(\mathbf{r})^2|$ while, at the same time, they are responsible for the differentiation of valence electrons in at least three sub-bands. Moreover, if we include also spin-orbit interactions, that is the overlapping term $(\mathbf{s} \times \mathbf{L})$ of interaction between the intrinsic electron angular momentum (i.e. the spin $\mathbf{s}$) and the magnetic field generated by the nuclei (or their momentum $\mathbf{L}$, almost equal to the electron orbit momentum), the energy splitting becomes even more evident.

Similarly, for what concerns conduction states, the s-orbitals and their quasi-spherical symmetry dominate the shape of the rapidly oscillating part of the wavefuction. Thanks to this symmetry, the formation of a conduction band could be explained solely through the overlap/degeneration of s-orbitals, without accounting for any sub-band splitting effect, as we did instead for valence band. Nevertheless, the lack of orbital momentum $\mathbf{L}$ in s-states does not allow spin-orbit effects. The only possible degeneration, at this stage of approximation, could be a slight contribution of the p-orbital to the conduction band at high energies (see, again, panel (b) in Figure 1.1). But as long as we assume $\mathbf{L} \simeq 0$ the sp two-fold splitting in this case will be considered negligible.

As we may appreciate from Figure 1.2, panel (b), in direct band gap semiconductors like $Hg_{1-x}Cd_xTe$, GaN, $In_xGa_{1-x}N$ or $Al_xGa_{1-x}N$ the conduction band (CB) minimum and the top of the valence band (VB) are located at the critical point Γ of the first Brillouin zone (FBZ), i.e. at $\mathbf{k} = 0$. This represents an advantage for some kinds of applications, especially in optoelectronics, where the most probable part of direct transitions between CB and VB can be established without interactions in the momentum space ($\Delta\mathbf{k} = 0$), as instead occurs in phonon-assisted processes. Being the conduction and valence band trend almost parabolic in $\mathbf{k}$ (see again panel (b) in Figure 1.2 and also Eq. (1.3)), if we neglect the tensorial notation for carrier effective masses around Γ, then one can see that m^* can be considered proportional to the curvature of the function $E(\mathbf{k})$. In fact:

$$\frac{1}{m^*} \simeq \frac{1}{\hbar^2} \frac{\mathrm{d}^2 E_n(\mathbf{k})}{\mathrm{d}\mathbf{k}^2} \,. \tag{1.6}$$

The lower the concavity, the higher m^*.

Now we briefly derive the $\mathbf{k} \cdot \mathbf{p}$ formalism starting from the substitution of a generic Bloch-wavefunction into the Schrödinger equation:

$$\left(-\frac{\hbar^2}{2m^*}\Delta_{\mathbf{r}} + \frac{\hbar}{m^*}\mathbf{k} \cdot \mathbf{p} + \frac{\hbar^2\mathbf{k}^2}{2m^*} + U(\mathbf{r}) \right) u_{n,\mathbf{k}}(\mathbf{r}) = E_n(\mathbf{k})\, u_{n,\mathbf{k}}(\mathbf{r}) \tag{1.7}$$

where $\mathbf{p} = -i\hbar\Delta$ is the momentum operator. Supposing to know the Hamiltonian and the wavefunction at a given $\mathbf{k}_0$

$$H_{\mathbf{k}_0} = -\frac{\hbar^2}{2m^*}\nabla^2_{\mathbf{r}} + \frac{\hbar}{m^*}\mathbf{k}_0 \cdot \mathbf{p} + \frac{\hbar^2\mathbf{k}_0^2}{2m^*} + U(\mathbf{r}) \tag{1.8}$$

and

$$\Psi_{n,\mathbf{k}_0}(\mathbf{r}) = u_{n,\mathbf{k}_0}(\mathbf{r})\, \mathrm{e}^{i\mathbf{k}_0\cdot\mathbf{r}}\,, \tag{1.9}$$

then Eq. (1.2) becomes

$$\left(H_{\mathbf{k}_0} + \frac{\hbar}{m^*}(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{p} + \frac{\hbar^2}{2m^*}(\mathbf{k}^2-\mathbf{k}_0^2)\right) u_{n,\mathbf{k}_0}(\mathbf{r}) = E_n(\mathbf{k})\, u_{n,\mathbf{k}_0}(\mathbf{r})\,, \tag{1.10}$$

where

$$u_{n,\mathbf{k}_0}(\mathbf{r}) = \sum_{n'} A_{n',n}\left(|\mathbf{k}-\mathbf{k}_0|\right) u_{n',\mathbf{k}_0}(\mathbf{r})\,, \tag{1.11}$$

If we plug Eq. (1.11) into Eq. (1.10), after multiplying both sides by $u^*_{n,\mathbf{k}_0}(\mathbf{r})$ and integrating over the volume of a unit cell, we obtain the eigenvalue *secular equation*:

$$\sum_{n'}\left[\left(E_n(\mathbf{k}_0) + \frac{\hbar^2}{2m^*}(\mathbf{k}^2-\mathbf{k}_0^2)\right)\delta_{n',n} + \frac{\hbar}{m^*}(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{p}_{n',n}\right] A_{n',n} = E_n(\mathbf{k})\, A_{n,n} \tag{1.12}$$

in which $\frac{\hbar}{m^*}(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{p}_{n',n}$ is our perturbation term and where

$$\mathbf{p}_{n',n} = \int u^*_{n,\mathbf{k}_0}(\mathbf{r})\,\mathbf{p}\,u^*_{n',\mathbf{k}_0}(\mathbf{r})\,\mathrm{d}\mathbf{r}\,. \tag{1.13}$$

Now one can proceed to solve the $\mathbf{k}\cdot\mathbf{p}$ secular equation written for $\mathbf{k}\simeq\mathbf{k}_0$, determining the coefficients $A_{n',n}$ and then also the eigenvalues $E_n(\mathbf{k})$ of the system in the perturbation framework. As we shall see in next chapters, the most important transitions in direct band gap semiconductors involve the main conduction band (CB) and at least three valence bands, depending on the effective hole mass m^*_{h}: so we have the heavy hole (HH) band, the light hole (LH) band and the split-off (SO) band. The presence of three different valence bands is due to the state separation produced, in turn, by spin-orbit interaction effects, since otherwise they would be degenerate.

1.2 Carrier Statistics

The classical picture describing a charge population in solids and its related density is usually based on the Fermi-Dirac statistical distributions

$$f_{\mathrm{FD}}(E) = \frac{1}{\mathrm{e}^{\frac{E-E_{\mathrm{F}}}{k_{\mathrm{B}}T}}+1}\,, \tag{1.14}$$

with k_{B} the Boltzmann constant and E_{F} the Fermi energy, which represents the highest possible energy of carriers when $T=0$. This formalism is straightforward for electrons, which are fermions (1/2-spin particles obeying to the Pauli exclusion principle) and, by symmetry, is also applied to holes.

Now, differential carrier densities are given by

$$\begin{cases}\mathrm{d}n = \rho_{\mathrm{C}}(E)\, f_{\mathrm{FD}}(E)\,\mathrm{d}E \\ \mathrm{d}p = \rho_{\mathrm{V}}(E)\,\left(1-f_{\mathrm{FD}}(E)\right)\,\mathrm{d}E\end{cases}, \tag{1.15}$$

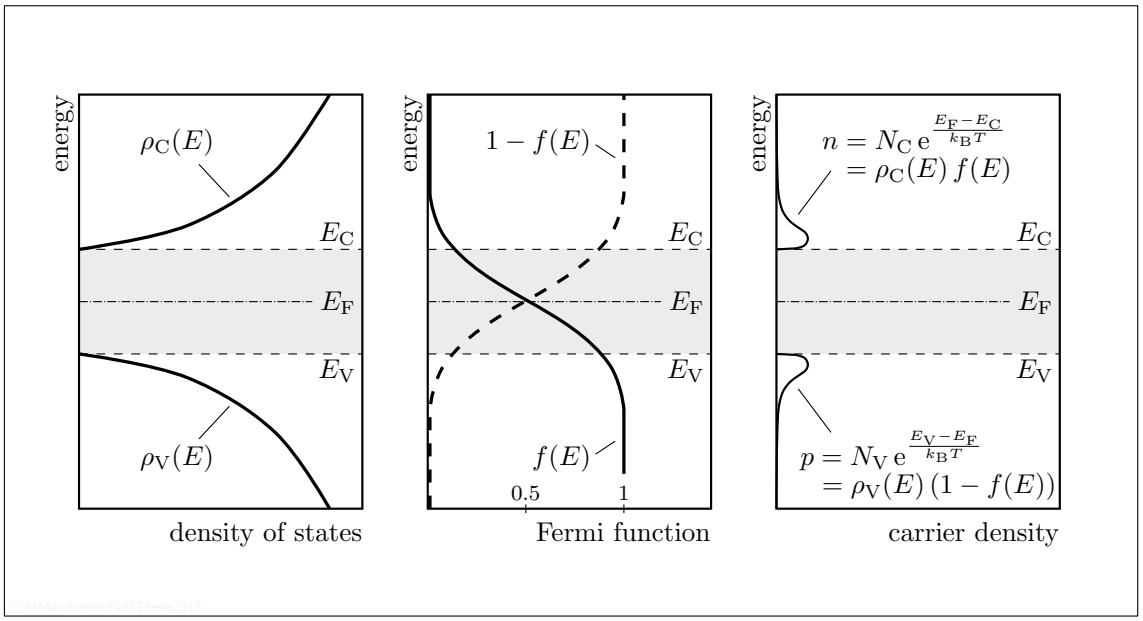

FIGURE 1.3. Density of states (DOS), Fermi function and carrier density represented as a function of energy in intrinsic semiconductors ($E_F \simeq (E_V + E_C)/2$ and $n = p = n_i$ at equilibrium) for $T > 0$.

where $\rho_C(E)$ and $\rho_V(E)$ are the so-called density of states (DOS)

$$\begin{cases} \rho_C(E) = \dfrac{dN_C(E)}{dE} = \dfrac{4\pi\sqrt{8m_e^*}}{h}\sqrt{E - E_C}\,, & \text{for} \quad E \geq E_C \\ \rho_V(E) = \dfrac{dN_V(E)}{dE} = \dfrac{4\pi\sqrt{8m_h^*}}{h}\sqrt{E_V - E}\,, & \text{for} \quad E \leq E_V \end{cases}, \tag{1.16}$$

which indicate how many states (occupied or not) can be found for a given energy E, and where $N_C(E)$ and $N_V(E)$ are the effective densities in conduction and valence band

$$N_{C,V} = 2\left(\frac{m_{e,h}^* k_B T}{2\pi\hbar^2}\right)^{\frac{3}{2}}, \tag{1.17}$$

where $m_{e,h}^*$ is the electron/hole effective mass.

At $T = 0$ the function $f_{FD}(E)$ is a step-function such that

$$f_{FD} = \begin{cases} 1 & \text{if } E \leq E_F \\ 0 & \text{if } E > E_F \end{cases} \tag{1.18}$$

while, as represented in Figure 1.3 for a generic intrinsic semiconductor, when T is greather than zero the Fermi-Dirac distribution function becomes smoothed in a region approximately wide $\sim k_B T$. This implies that only at thermal equilibrium ($T = 0$) E_F represents the energy which corresponds to completely filled VB and empty CB. At higher temperatures we have that $f(E_F) = 1/2$. So, if E_F stays within the band gap (i.e. for low doping densities) and the energy gap is sufficiently greater than $k_B T$, then the semiconductor can be still considered somehow at thermodynamic equilibrium. If the gap is narrower, instead, the finite probability $f_{FD}(E)$ to

have an electron in CB (and, at the same time, the probability $(1 - f(E))$ to have a hole in VB) can induce "spontaneous" band-to-band transitions typical of these materials, as we will see talking about tunneling phenomena.

The Fermi-Dirac distribution is not only valid for intrinsic semiconductors, i.e. the one in which the Fermi level equals the intrinsic one

$$E_{\mathrm{F}} \equiv E_{\mathrm{F}_i} = \frac{1}{2}\left((E_{\mathrm{C}} + E_{\mathrm{V}}) + k_{\mathrm{B}}T \log\left[\frac{N_{\mathrm{V}}}{N_{\mathrm{C}}}\right]\right) , \tag{1.19}$$

but it can be also applied to other materials like degenerate semiconductors, where doping levels are very high. Nonetheless, the choice of an approximated description is common when we study non-degenerate materials. This approximation consists in using the Boltzmann distribution function

$$f_{\mathrm{B}}(E) = \mathrm{e}^{-\frac{E-E_{\mathrm{F}}}{k_{\mathrm{B}}T}} \tag{1.20}$$

in place of f_{FD}, where now the carrier population no longer obeys to the Pauli exclusion principle. Comparing the two formulations one can notice that

$$f_{\mathrm{FD}}(E) \simeq f_{\mathrm{B}}(E) \quad \Leftrightarrow \quad \exp\left[\frac{E - E_{\mathrm{F}}}{k_{\mathrm{B}}T}\right] \gg 1 , \tag{1.21}$$

i.e. for $E \gg k_{\mathrm{B}}T$. At this limit the classical Boltzmann treatment in place of the semiclassical Fermi-Dirac one is rather well acceptable.

Finally, integrating the differential density for electrons and holes in Eq. (1.15) with respect to the bands we obtain their concentrations at equilibrium:

$$\begin{cases} n = N_{\mathrm{C}}\, \mathrm{e}^{\frac{E_{\mathrm{F}}-E_{\mathrm{C}}}{k_{\mathrm{B}}T}} \\ p = N_{\mathrm{V}}\, \mathrm{e}^{\frac{E_{\mathrm{V}}-E_{\mathrm{F}}}{k_{\mathrm{B}}T}} \end{cases} , \tag{1.22}$$

where

$$n\,p = \sqrt{N_{\mathrm{C}}\,N_{\mathrm{V}}}\, \mathrm{e}^{-\frac{E_{\mathrm{g}}}{2k_{\mathrm{B}}T}} \equiv n_i . \tag{1.23}$$

Alternatively, one can also write

$$\begin{cases} n = n_i\, e^{\frac{E_{\mathrm{F}}-E_{\mathrm{F}_i}}{k_{\mathrm{B}}T}} \\ p = n_i\, e^{\frac{E_{\mathrm{F}_i}-E_{\mathrm{F}}}{k_{\mathrm{B}}T}} \end{cases} . \tag{1.24}$$

1.3 Doping and Carrier Concentrations

A building-block of each solid-state system is constituted by the feature of artificially enhance the native carrier concentrations through the introduction of impurity atoms. These dopants elements may act as donors or acceptors and their energy E_{D} or E_{A} tipically lies just below the CB minimum or above the top of the VB,

respectively. The difference between this energy and the corresponding band edge is called ionization energy E_{ion}:

$$E_{\text{D,ion}} = E_{\text{C}} - E_{\text{D}} \quad \text{or} \quad E_{\text{A,ion}} = E_{\text{A}} - E_{\text{V}} \,. \tag{1.25}$$

In neutral conditions the new Fermi level due to doping is determined by the balance equation

$$n_{\text{tot}} = n + n_{\text{D}} = p + p_{\text{A}} = p_{\text{tot}} \tag{1.26}$$

where the ionized carrier densities, still obeying to the Fermi statistics, are

$$\begin{cases} n_{\text{D}} = N_{\text{D}} \dfrac{1}{1 + g_{\text{D}}\, \text{e}^{\left(\frac{E_{\text{F}} - E_{\text{D}}}{k_{\text{B}} T}\right)}} \\ p_{\text{A}} = N_{\text{A}} \dfrac{1}{1 + g_{\text{A}}\, \text{e}^{\left(\frac{E_{\text{A}} - E_{\text{F}}}{k_{\text{B}} T}\right)}} \end{cases} , \tag{1.27}$$

being g_{D} and g_{A} the degeneracy factors (usually $g_{\text{D}} = 2$ and $g_{\text{A}} = 4$). Eq. (1.27) is known as the *incomplete ionization* law, since n_{D} and p_{A} determine the fraction of thermally ionized dopant, once their nominal values N_{D} and N_{A} are given. In other words, effective doping can reach the concentration of implanted chemical dopants only in the limit of high T, so:

$$\begin{cases} n_{\text{D}} \overset{T \to \infty}{=} N_{\text{D}} \\ p_{\text{A}} \overset{T \to \infty}{=} N_{\text{A}} \end{cases} . \tag{1.28}$$

When a doped semiconductor is out of equilibrium (by carrier injection or other external stimuli), the Fermi level changes its position going close to the CB in n-type materials and to the VB in p-type ones:

$$\begin{cases} E_{\text{F}_n} = E_{\text{F}_i} + k_{\text{B}} T \, \ln\left(n_{\text{tot}} / n_i\right) \\ E_{\text{F}_p} = E_{\text{F}_i} - k_{\text{B}} T \, \ln\left(p_{\text{tot}} / n_i\right) \end{cases} . \tag{1.29}$$

These two new energies are called quasi-Fermi levels.

1.4 Generation/Recombination Mechanisms

The usual operation state of semiconductor-based electronic devices consists in a non-equilibrium condition. In this regime the external energy given to the system is partially converted into a signal (optical or electrical) through specific transitions which can take place in the energy domain, as well as in real or momentum space. In the case of direct band gap semiconductors, most of the common transitions between bands occur at the Γ symmetry point, where the gap reaches its minimum energy, E_{g}. This is the condition in which electron/hole pairs are *generated* or *recombined* (annihilated), depending on whether the energy is absorbed or delivered, respectively. This energy can involve a photon, i.e. a quantum of light, as in case of

optical processes. But, in general, interband transitions may occur also in the non-radiative regime. Generation/recombination mechanisms change the availability of carriers in CB and VB, both majority and minority, determining *relaxation* processes towards equilibrium which can affect and limit the device performance by means of carriers lifetime enhancement.

So, GR processes can be divided into two main categories, non-radiative and radiative mechanisms. In both, generation consists of an electron coming from CB that occupies an empty state (hole) in VB, while recombination is its inverse, and these two competitive processes are perfectly balanced at equilibrium. Electron/hole pairs are generated or recombined through a characteristic lifetime, depending on the material and on the particular GR mechanism involved. Out of equilibrium the algebraic sum of the generation G and recombination R rate is the so-called *net generation/recombination rate* $U = G - R$, i.e. the number of pairs recombined per volume in the unit time ($\mathrm{cm^{-3}s^{-1}}$), which is a function of the carrier densities at equilibrium n_0, p_0 and out of equilibrium n', p' such that $U \neq 0$ if and only if $\Delta n \neq 0$ and $\Delta p \neq 0$, where $\Delta n = n' - n_0$ and $\Delta p = p' - p_0$. This set of properties goes under the name of *relaxation time approximation*, according to which we have:

$$\begin{cases} U_n \simeq \dfrac{n' - n_0}{\tau_n} = \dfrac{\Delta n}{\tau_n} \\ U_p \simeq \dfrac{p' - p_0}{\tau_p} = \dfrac{\Delta p}{\tau_p} \end{cases}, \tag{1.30}$$

where τ_n and τ_p are the electron and hole lifetimes, respectively. For small excess carrier concentrations Δn, Δp and within the Boltzmann statistics framework:

$$\begin{cases} \dfrac{\partial \Delta n}{\partial t} \simeq -\dfrac{\Delta n}{\tau_n} \\ \dfrac{\partial \Delta p}{\partial t} \simeq -\dfrac{\Delta p}{\tau_p} \end{cases} \tag{1.31}$$

that yields:

$$\begin{cases} n(t) = \Delta n\, e^{-t/\tau_n} + n_0 \\ p(t) = \Delta p\, e^{-t/\tau_p} + p_0 \end{cases}, \tag{1.32}$$

from which it is even more clear why τ are called relaxation times, being the time constant needed to restore the original equilibrium concentrations n_0, p_0.

Under the same assumptions given before, the lifetime contribution coming from each GR process (non-radiative, radiative and others) can be treated as independent. For this reason:

$$\frac{1}{\tau_{\mathrm{tot}}} = \sum_j \frac{1}{\tau_j} \tag{1.33}$$

where index j stands for the contribution from a j-th GR process.

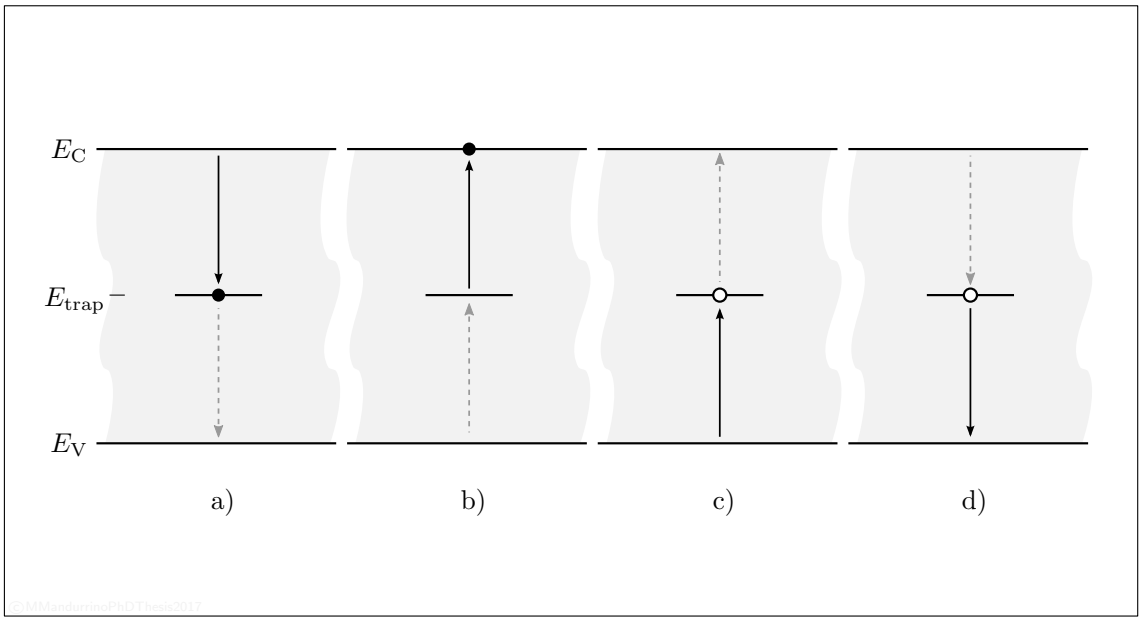

FIGURE 1.4. Single-level model of Shockley-Read-Hall (SRH) generation and recombination processes in a direct band gap semiconductor. Solid arrows represent: (a) electron capture. (b) electron emission. (c) hole capture. (d) hole emission. Completing the processes with dashed arrows we have recombination in (a) and (c) and generation in (b) and (d).

1.4.1 Non-radiative mechanisms

1.4.1.1 SRH generation/recombination

A first class of non-radiative processes involves a certain density N_{trap} of energy levels within the semiconductor band gap, like dopants, impurities or defects. These states, also called Shockley-Read-Hall (SRH) centers, can assist interband transitions which are typically stimulated by energies lower than E_g. As depicted in Figure 1.4 for a single-level model, these transitions may follow four different channels. In panel (a) an empty SRH trap (with and energy E_{trap}) is occupied by a conduction electron with capture cross-section σ_n and capture rate $c_n = \sigma_n \, v_{\text{th}_n}$, where v_{th_n} is the thermal speed of the electron. In (b) an electron escapes from a trap with emission rate e_n and is promoted into CB. Panels (c) and (d), instead, respectively represent the capture and emission of a valence hole with rates $c_p = \sigma_p \, v_{\text{th}_p}$ and e_p. Likewise, v_{th_p} is the thermal speed of the hole and:

$$v_{\text{th}_{n,p}} = \sqrt{\frac{3k_B T}{m^*_{\text{e,h}}}} \, . \tag{1.34}$$

Deriving Eqs. (1.32) with respect to time and recalling Eq. (1.30), which is valid for small Δn and Δp, one can write the time-evolution of carriers in SRH processes

for direct band gap semiconductors as [2]

$$\begin{cases} U_{\mathrm{SRH}_n} = -\dfrac{\mathrm{d}n(t)}{\mathrm{d}t} = -c_n\, n(t)\, N_{\mathrm{trap}}\,(1 - f_{\mathrm{trap}}(E_{\mathrm{trap}})) + e_n\, N_{\mathrm{trap}}\, f_{\mathrm{trap}}(E_{\mathrm{trap}}) \\ U_{\mathrm{SRH}_p} = -\dfrac{\mathrm{d}p(t)}{\mathrm{d}t} = -c_p\, p(t)\, N_{\mathrm{trap}}\, f_{\mathrm{trap}}(E_{\mathrm{trap}}) + e_p\, N_{\mathrm{trap}}\,(1 - f_{\mathrm{trap}}(E_{\mathrm{trap}})) \end{cases} \tag{1.35}$$

that, thanks to the charge neutrality law, are linked via

$$n(t) + N_{\mathrm{trap}}\, f_{\mathrm{trap}}(E_{\mathrm{trap}}) = p(t)\,, \tag{1.36}$$

where $f_{\mathrm{trap}}(E_{\mathrm{trap}})$ is the occupation function of the trap level. At steady-state equilibrium (i.e. when $\frac{\mathrm{d}n}{\mathrm{d}t} = 0$, $\frac{\mathrm{d}p}{\mathrm{d}t} = 0$, $n \equiv n_0$ and $p \equiv p_0$):

$$\begin{cases} \dfrac{e_{n_0}}{c_{n_0}} = n_i\, \mathrm{e}^{\frac{E_{\mathrm{trap}} - E_{\mathrm{F}_i}}{k_{\mathrm{B}} T}} \equiv n_1 \\ \dfrac{e_{p_0}}{c_{p_0}} = n_i\, \mathrm{e}^{\frac{E_{\mathrm{F}_i} - E_{\mathrm{trap}}}{k_{\mathrm{B}} T}} \equiv p_1 \end{cases} \tag{1.37}$$

and

$$\begin{cases} \tau_{n_0} = \dfrac{1}{c_n\, N_{\mathrm{trap}}} = \dfrac{1}{\sigma_n\, v_{\mathrm{th}_n}\, N_{\mathrm{trap}}} \\ \tau_{p_0} = \dfrac{1}{c_p\, N_{\mathrm{trap}}} = \dfrac{1}{\sigma_p\, v_{\mathrm{th}_p}\, N_{\mathrm{trap}}} \end{cases}, \tag{1.38}$$

where n_1 and p_1 are valid when $E_{\mathrm{trap}} \equiv E_{\mathrm{F}}$.

Now, assuming $N^{\pm}_{\mathrm{trap}} = N_{\mathrm{trap}} f_{\mathrm{trap}}(E_{\mathrm{trap}})$ the fraction of occupied (charged) traps and $N^{0}_{\mathrm{trap}} = N_{\mathrm{trap}}(1 - f_{\mathrm{trap}}(E_{\mathrm{trap}}))$ the same quantity referred to unoccupied (neutral) ones, we can rewrite Eq. (1.35) as

$$\begin{cases} \dfrac{\mathrm{d}n(t)}{\mathrm{d}t} = -c_n\,(n\, N^{0}_{\mathrm{trap}} - n_1\, N^{\pm}_{\mathrm{trap}}) \\ \dfrac{\mathrm{d}p(t)}{\mathrm{d}t} = -c_p\,(p\, N^{\pm}_{\mathrm{trap}} - p_1\, N^{0}_{\mathrm{trap}}) \end{cases}. \tag{1.39}$$

Then we have (still in steady-state conditions) that

$$\begin{aligned} \frac{\mathrm{d}N^{\pm}_{\mathrm{trap}}}{\mathrm{d}t} &= -\frac{\mathrm{d}n(t)}{\mathrm{d}t} + \frac{\mathrm{d}p(t)}{\mathrm{d}t} \\ &= c_n\,(n\, N^{0}_{\mathrm{trap}} - n_1\, N^{\pm}_{\mathrm{trap}}) - c_p\,(p\, N^{\pm}_{\mathrm{trap}} - p_1\, N^{0}_{\mathrm{trap}}) \end{aligned} \tag{1.40}$$

which must be null by definition (except for transient regimes). Thus:

$$\begin{cases} N^{0}_{\mathrm{trap}} = N_{\mathrm{trap}}\, \dfrac{c_n\, n_1 + c_p\, p}{c_n(n + n_1) + c_p(p + p_1)} \\ N^{\pm}_{\mathrm{trap}} = N_{\mathrm{trap}}\, \dfrac{c_n\, n + c_p\, p_1}{c_n(n + n_1) + c_p(p + p_1)} \end{cases}. \tag{1.41}$$

By using all the previous relations, the relaxation process involving excited carrier densities n', p' can be written as

$$\frac{\mathrm{d}n'(t)}{\mathrm{d}t} = \frac{\mathrm{d}p'(t)}{\mathrm{d}t} = \frac{(n_0 + \Delta n)(p_0 + \Delta p) - n_i^2}{\tau_{p_0}(n_0 + \Delta n + n_1) + \tau_{n_0}(p_0 + \Delta p + p_1)} = U_{\mathrm{SRH}}\,. \tag{1.42}$$

So, for electrons:

$$\tau_{\mathrm{SRH}_n} = -\left(\frac{\mathrm{d}n'}{\mathrm{d}t}\right)^{-1} \Delta n \tag{1.43}$$

$$= -\frac{\tau_{p_0}(n_0 + \Delta n + n_1) + \tau_{n_0}(p_0 + \Delta p + p_1)}{(n_0 + \Delta n)(p_0 + \Delta p) - n_i^2}\,\Delta n \tag{1.44}$$

$$= -\frac{\sigma_n\, v_{\mathrm{th}_n}(n_0 + \Delta n + n_1) + \sigma_p\, v_{\mathrm{th}_p}(p_0 + \Delta p + p_1)}{\sigma_n\, v_{\mathrm{th}_n}\, \sigma_p\, v_{\mathrm{th}_p}\, N_{\mathrm{trap}}^2\,[(n_0 + \Delta n)(p_0 + \Delta p) - n_i^2]}\,\Delta n \tag{1.45}$$

and similarly for the holes. Then, recovering our previous steady-state assumptions:

$$U_{\mathrm{SRH}} = \frac{np - \gamma_n\gamma_p\, n_i^2}{\tau_p\left(n + \gamma_n\, n_i\, \mathrm{e}^{\frac{E_{\mathrm{trap}} - E_{\mathrm{F}_i}}{k_{\mathrm{B}}T}}\right) + \tau_n\left(p + \gamma_p\, n_i\, \mathrm{e}^{\frac{E_{\mathrm{F}_i} - E_{\mathrm{trap}}}{k_{\mathrm{B}}T}}\right)} \tag{1.46}$$

(from which is evident that U_{SRH} is maximized when $E_{\mathrm{trap}} = E_{\mathrm{F}_i}$) and

$$\tau_{\mathrm{SRH}_{n,p}} = \frac{1}{\sigma_{n,p}\, v_{\mathrm{th}_{n,p}}\, N_{\mathrm{trap}}}\,, \tag{1.47}$$

where the functions $\gamma_{n,p}$ are equal to 1 in case of Boltzmann statistics (as supposed so far) otherwise, in the Fermi statistics picture

$$\begin{cases} \gamma_n = \dfrac{n}{N_{\mathrm{C}}}\,\mathrm{e}^{-\frac{E_{\mathrm{F}_n} - E_{\mathrm{C}}}{k_{\mathrm{B}}T}} \\ \gamma_p = \dfrac{p}{N_{\mathrm{V}}}\,\mathrm{e}^{-\frac{E_{\mathrm{V}} - E_{\mathrm{F}_p}}{k_{\mathrm{B}}T}} \end{cases} \tag{1.48}$$

and the carrier densities become

$$\begin{cases} n_{\mathrm{F}} = N_{\mathrm{C}}\,\mathcal{F}_{1/2}\left(\dfrac{E_{\mathrm{F}_n} - E_{\mathrm{C}}}{k_{\mathrm{B}}T}\right) \\ p_{\mathrm{F}} = N_{\mathrm{V}}\,\mathcal{F}_{1/2}\left(\dfrac{E_{\mathrm{V}} - E_{\mathrm{F}_p}}{k_{\mathrm{B}}T}\right) \end{cases}, \tag{1.49}$$

where $N_{\mathrm{C,V}}$ are the conduction/valence band effective densities, $\mathcal{F}_{1/2}(\epsilon)$ represents the 1/2-th order *Fermi integral*

$$\mathcal{F}_{1/2}(\epsilon) = \frac{2}{\sqrt{\pi}}\int_0^\infty \frac{\sqrt{E}}{1 + \mathrm{e}^{E+\epsilon}}\,\mathrm{d}E \tag{1.50}$$

and $E_{\mathrm{F}_{n,p}}$ are the quasi-Fermi levels for electrons and holes, which generate from the Fermi level when the material is out of equilibrium.

Now, since $\tau_n = \Delta n/U_{\mathrm{SRH}}$ and $\tau_p = \Delta p/U_{\mathrm{SRH}}$, as stated by Eq. (1.43), by using one of the possible representations of carrier densities (Boltzmann or Fermi) one can appreciate the T-dependence of the SRH rate, which is increasing with temperature.

1.4.1.2 Auger generation/recombination

Besides defect-assisted non-radiative GR processes, a very important family of processes is represented by Auger mechanisms, "three-body" interactions classified depending on which carriers or sub-bands are involved. So, in principle, different types of Auger GR can exist. The most important are depicted in Figure 1.5: in panel (a) we find the "eeh" transition, which is the recombination of a CB electron, where the excess energy is released to a second electron that can be promoted to a higher conduction level; then in (b) the "ehh" process is represented, where the energy is released to a hole, contrarily to the previous case; finally (c) represents the "hhe" generation due to a hole promoted into the CB, where the excess energy is transferred to another hole that is excited into a deeper valence state.

As just mentioned, processes (a) and (b) involve the recombination of an electron/hole pair before interacting with a third carrier. The case (c), instead, represents an example of Auger generation, a process which is commonly referred to as *impact ionization*. In this sense Auger and impact ionization are symmetric processes, whatever the kind of carriers involved. If Auger GR can be heavily present in forward-biased light-emitting diodes based on wide band gap materials, especially at high T or injection regimes, impact ionization can be more important in narrow band gap IR photodiodes, which operate in reverse bias conditions.

Being a three-body mechanism, the net Auger recombination rate is fundamentally different from traditional non-radiative rates. Moreover one has to observe that its rate roughly goes like $\sim n^2 p$ for "eeh" processes or $\sim np^2$ for "ehh" ones and, in general, it is proportional to the third power of the total carrier density or, in turn, to the doping level. Now we go through the derivation of a macroscopic description of the Auger recombination rate and lifetime, starting from quantum theory based on microscopic quantities towards the common expressions used in experimental or modeling studies. To do this, a formalism among all those present in the literature has been chosen, which is the one developed by P. T. Landsberg and well discussed in Ref. [3].

Let us start with considering as a template example the process CHCC. A new detailed scheme is proposed in Figure 1.6, where all the single steps constituting the whole process are highlighted and properly labeled: numbers 1 and $1'$ indicate respectively the initial and the final states of the first "direct transition", i.e. the band-to-band recombination of an electron/hole pair, while 2 and $2'$ are the initial and final state of the second process, which we shall call "exchange interaction". In the Landsberg formalism they are all described by Bloch states.

So, by using the Fermi golden rule in the first-order perturbation theory, the overall transition rate as a function of time t can be written as

$$|W_{1\to 2'}(t)| = \frac{2\pi t}{\hbar} \left| H^{(1)}_{2'\to 1} \right|^2 \delta\left(E^{(0)}_{2'} - E^{(0)}_{1} \right) , \tag{1.51}$$

where $H^{(1)}_{2'\to 1}$ is the first-order perturbed interaction Hamiltonian while energies at the initial and final state $E^{(0)}_1$, $E^{(0)}_{2'}$ are evaluated as zero-th order (non-perturbed) terms. Notice that $H^{(1)}$ is time-independent and also Hermitian, so $H^{(1)}_{2'\to 1} \equiv H^{(1)*}_{1\to 2'}$.

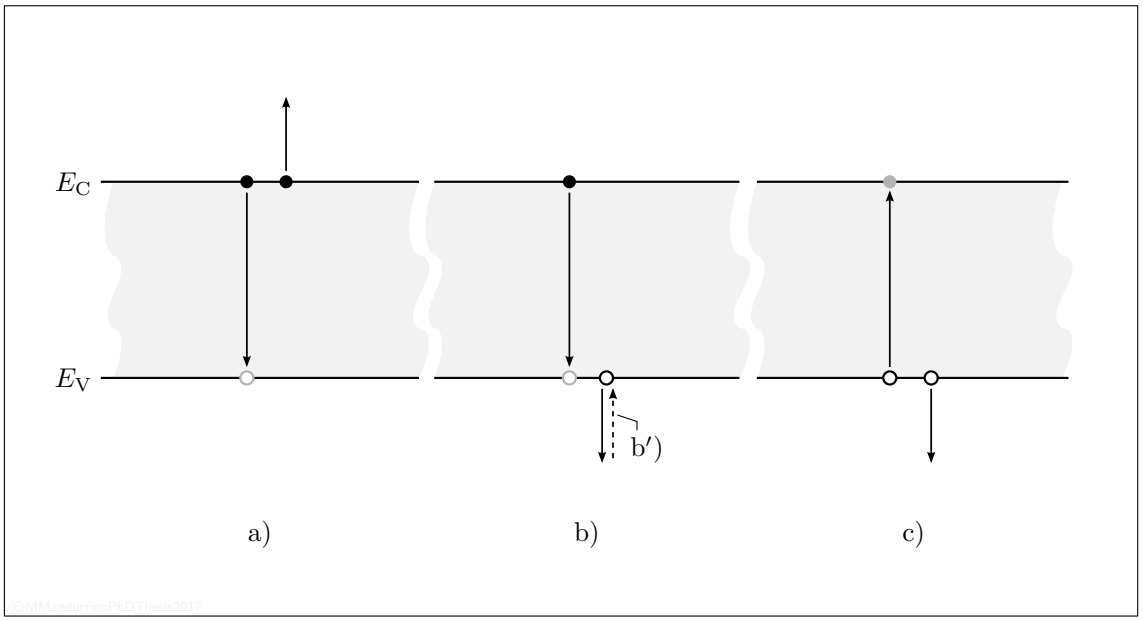

FIGURE 1.5. Most important Auger generation/recombination processes in a direct band gap semiconductor (without any phonon- or defect-assisted mechanism). (a) "eeh" Auger recombination also named CHCC where, after the electron/hole has recombined, a second electron is excited. (b) "ehh" Auger recombination, or CHHL, in which the same recombination is followed by a transition involving a hole. Process (b′) is only a slight variant of the previous one and is labeled as CHLH. (c) "hhe" Auger generation (or, better, impact ionization) where, after the electron/hole pair generation, a second hole is promoted to another valence energy state. Notice that this scheme also includes the third carrier participating to the GR transition, here represented by filled/empty grey dots, respectively for electrons and holes.

Moreover:

$$\left|H^{(1)}_{2'\to 1}\right|^2 = 2\,\beta\,\left(|M_\mathrm{d}|^2 + |M_\mathrm{e}|^2\right)\,, \tag{1.52}$$

where β is a spin-related term such that $1 < \beta < 2$ while M_d and M_e are the Bloch matrix elements for direct and exchange transition, respectively. It can be demonstrated that they can be approximated as

$$|M_\mathrm{d}| \simeq \frac{4\pi q^2}{\epsilon V}\frac{|F_{1,2}F_{1',2'}|}{\lambda^2 + |\mathbf{k}_1 - \mathbf{k}_{1'}|^2}\,\delta_{\mathbf{k}_1+\mathbf{k}_2,\,\mathbf{k}_{1'}+\mathbf{k}_{2'}} \tag{1.53}$$

and

$$|M_\mathrm{e}| \simeq \frac{4\pi q^2}{\epsilon V}\frac{|F_{1,2'}F_{1',2}|}{\lambda^2 + |\mathbf{k}_2 - \mathbf{k}_{1'}|^2}\,\delta_{\mathbf{k}_1+\mathbf{k}_2,\,\mathbf{k}_{1'}+\mathbf{k}_{2'}}\,, \tag{1.54}$$

in which q is the elementary charge, ϵ and V are the material dielectric constant and volume, respectively, λ^{-1} is the Coulomb screening radius, the four F represent all the possible Bloch wavefunction overlaps and $\mathbf{k}_1$, $\mathbf{k}_{1'}$, $\mathbf{k}_2$ and $\mathbf{k}_{2'}$ are the momenta corresponding to the states involved in the "eeh" process (see Figure 1.6).

Putting together the equations we reported, the interaction Hamiltonian can be

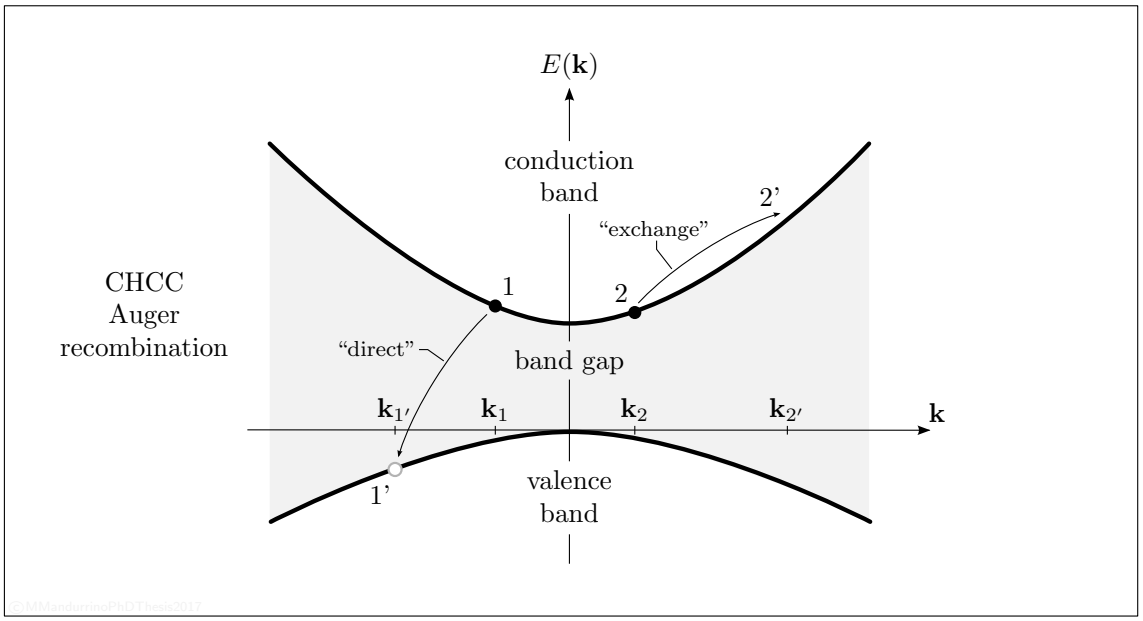

FIGURE 1.6. CHCC Auger ("eeh") recombination process, where the four Bloch states and their momenta have been labeled. The first "direct" transition corresponds to the electron/hole recombination while the subsequent "exchange" interaction due to momentum conservation consists of an excitation of a second electron within the conduction band (CB).

rewritten as follows:

$$\left|H^{(1)}_{2'\to 1}\right|^2 \simeq \frac{32\pi^2 q^4}{\epsilon^2 V^2}\left\{\frac{\left|F_{1,2}F_{1',2'}\right|^2}{\left(\lambda^2+\left|\mathbf{k}_1-\mathbf{k}_{1'}\right|^2\right)^2}+\frac{\left|F_{1,2'}F_{1',2}\right|^2}{\left(\lambda^2+\left|\mathbf{k}_2-\mathbf{k}_{1'}\right|^2\right)^2}\right.$$
$$\left.+\left|\frac{F_{1,2}F_{1',2'}}{\lambda^2+\left|\mathbf{k}_1-\mathbf{k}_{1'}\right|^2}-\frac{F_{1,2'}F_{1',2}}{\lambda^2+\left|\mathbf{k}_2-\mathbf{k}_{1'}\right|^2}\right|^2\right\}\delta_{\mathbf{k}_1+\mathbf{k}_2,\,\mathbf{k}_{1'}+\mathbf{k}_{2'}}\,. \quad (1.55)$$

Now the CHCC Auger recombination probability per unit volume and per unit time is given by:

$$\begin{aligned} P &\simeq \frac{2t}{\hbar^2 V}\int \Phi\left|H^{(1)}_{2'\to 1}\right|^2\frac{1-\cos x}{x^2}\mathrm{d}\mathbf{k}_1\,\mathrm{d}\mathbf{k}_2\,\mathrm{d}\mathbf{k}_{1'}\,\mathrm{d}\mathbf{k}_{2'} \\ &\simeq \frac{32\pi^2 q^4 t}{V^3\hbar^2\epsilon^3}\left(\frac{\Delta n}{n_0}+\frac{\Delta p}{p_0}\right)\left(\frac{V}{8\pi^3}\right)^3\int f(1)\,f(2)\,(1-f(1'))\,(1-f(2')) \\ &\quad\times\{\cdots\}\left(1-\mathrm{e}^{-\frac{E_{\mathrm{F}n}-E_{\mathrm{F}p}}{k_{\mathrm{B}}T}}\right)\frac{1-\cos x}{x^2}\,\mathrm{d}\mathbf{k}_1\,\mathrm{d}\mathbf{k}_2\,\mathrm{d}\mathbf{k}_{1'}\,\mathrm{d}\mathbf{k}_{2'} \end{aligned} \quad (1.56)$$

where Φ is computed at equilibrium as

$$\Phi = f(1)\,f(2)\,(1-f(1'))\,(1-f(2'))\left(1-\mathrm{e}^{-\frac{E_{\mathrm{F}n}-E_{\mathrm{F}p}}{k_{\mathrm{B}}T}}\right), \quad (1.57)$$

in which the f's are the Fermi-Dirac occupation functions, and where

$$x = \frac{t}{\hbar}\left(E^{(0)}_{2'}-E^{(0)}_{1}\right). \quad (1.58)$$

The term omitted for reasons of brevity in Eq. (1.56) is the same within braces in Eq. (1.55). If one notices that the total CHCC recombination probability is intrinsically dominated by the "direct" process, then the bracketed term can be reduced to the only first addend.

At this point, lifetimes and recombination rate of any Auger process can be computed from the CHCC probability P through

$$\begin{aligned}\frac{1}{\tau_{\mathrm{A}}} &= \frac{1}{\tau_{n,\mathrm{A}}} + \frac{1}{\tau_{p,\mathrm{A}}} \\ &= \frac{P}{n-n_0} + \frac{P}{p-p_0} \\ &= C_n \frac{n(np-n_0p_0)}{\Delta n} + C_p \frac{p(np-n_0p_0)}{\Delta p} \\ &\simeq \frac{C_n n}{\Delta n} + \frac{C_p p}{\Delta p}\left(np - n_i^2\right), \end{aligned} \tag{1.59}$$

in which the last equality becomes exact for intrinsic semiconductors. From lifetimes, as usual, we can derive the recombination rate in equilibrium conditions ($\Delta n = \Delta p$) remembering that $R = \Delta n/\tau$. So:

$$R_{\mathrm{A}} = \left(C_n n + C_p p\right)\left(np - n_i^2\right). \tag{1.60}$$

Coefficients $C_{n,p}$ represent a macroscopic (and somehow semi-empirical) description of the many-body theoretical probability integral P. They are material and process dependent and are expressed in $\mathrm{cm}^6\mathrm{s}^{-1}$ units.

1.4.2 Radiative mechanisms

The last important category of GR processes is represented by radiative mechanisms. They are essential in optoelectronics since (spontaneous) radiative recombination is the building-block physical mechanism in light-emitters and photogeneration, its symmetrical process, is at the basis of light detection. Discarding at this moment all tunneling or trap-assisted transitions, one can say that the simplest way to define radiative GR is that they are processes where the excess energy released (absorbed) during the band-to-band transition is given to (by) a photon, as schematically reproduced in Figure 1.7. Due to the finite probability of having optical transitions at $\mathbf{k} > 0$, usually the emission spectrum of a given semiconductor does not coincide with a δ-Dirac function but it corresponds to a broad curve, typically modeled by a Lorentian as

$$L(\nu) = \frac{\Gamma(\nu)}{2\pi} \frac{1}{\left(E_{\mathrm{g}} - 2\pi\hbar\nu + \hbar\nu\right)^2 + \left(\Gamma(\nu)/2\right)^2}, \tag{1.61}$$

where $\Gamma(\nu)$ is a Gamma function of the photon energy $E_{\mathrm{ph}} = h\nu$.

Now we re-propose here a very short derivation of the radiative recombination rate R_{rad} from quantum principles, which finds its origins in the $\mathbf{k}\cdot\mathbf{p}$ method.

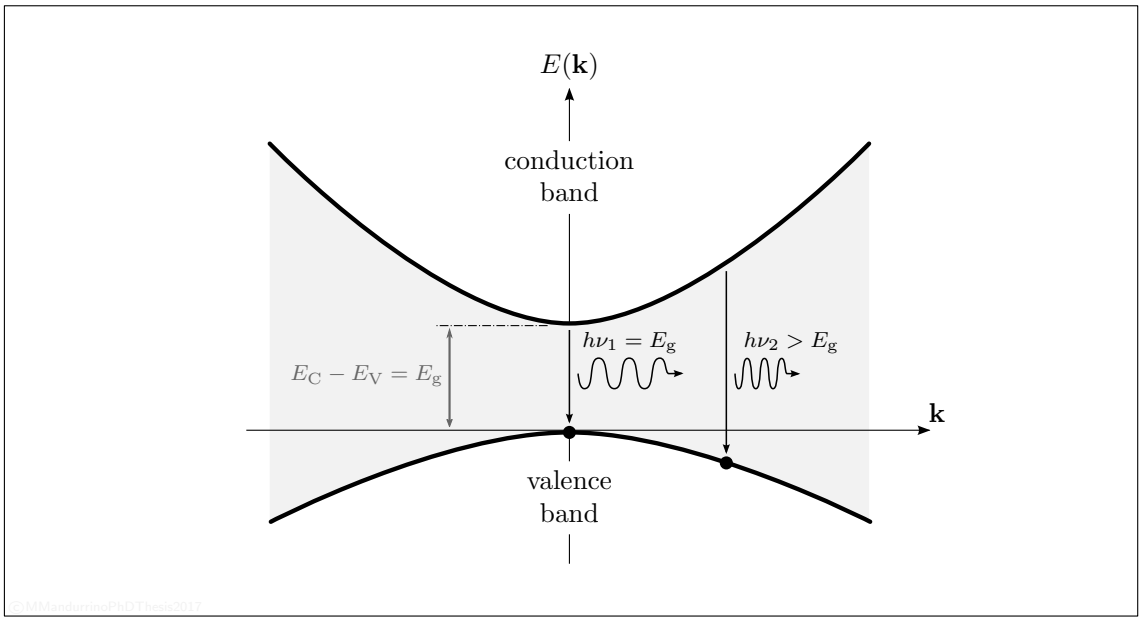

FIGURE 1.7. Simplified representation of spontaneous radiative recombination with photon emission at different energies and, thus, involving states with different wavevectors: for transitions at the symmetry point Γ (i.e. $\mathbf{k} = 0$) the photon energy $h\nu$ exactly corresponds to the energy gap E_g otherwise, at higher wavevectors, must be $h\nu_2 > E_g$.

Again we exploit the Fermi golden rule to write the recombination rate probability in multi-photon emission mode, which is [3]

$$P_{\text{rad}}(E, \nu_{\mathbf{k}}) = \sum_{\text{i,f}} \frac{4q^2 \hat{n} \nu_{\mathbf{k}}}{\bar{m}^2 V c^3 \hbar^2} \, |M_{\text{i,f}}(\mathbf{k})|^2 \, \delta\left(E_{\text{i}} - E_{\text{f}} - \hbar\nu_{\mathbf{k}}\right) \left(N_{\mathbf{k}} + 1\right) , \tag{1.62}$$

where "i" and "f" label the initial and final state in conduction and valence band, respectively, $\hbar\nu_{\mathbf{k}}$ is the $\mathbf{k}$-dependent photon energy, $N_{\mathbf{k}}$ the number of emitted photons per each $\mathbf{k}$ point, c is the speed of light in vacuum, V the material volume, $\hat{n}$ its refractive index, $\bar{m} = m^*_{\text{C}} \left|m^*_{\text{HH}}\right| / \left(m^*_{\text{C}} + \left|m^*_{\text{HH}}\right|\right)$ and $M_{\text{i,f}}$ is the transition matrix element

$$|M_{\text{i,f}}(\mathbf{k})|^2 = P_{\text{i,f}} \, |F_{\text{i,f}}(\mathbf{k})|^2 \left(\frac{m^*_{\text{C}}(\mathbf{k})}{m_0} - 1\right) \frac{m_0 E_g \left(E_g + \Delta_{\text{SO}}\right)}{12 \left(E_g + \frac{2}{3}\Delta_{\text{SO}}\right)} , \tag{1.63}$$

being $P_{\text{i,f}}$ a polarization term (due to spontaneous piezoelectric charges induced by lattice bond potentials), Δ_{SO} is the spin-orbit splitting energy responsible for the gap between HH/LH bands and the SO band in Γ and, finally, $F_{\text{i,f}}$ is the wavefunction overlap integral of conduction/valence band states involved in the radiative transition.

The expression which describes the probability rate is $\mathbf{k}$- and E-dependent, so (as we published in the paper of Ref. [4]) we can obtain the total net recombination

rate by integrating it over all the possible wavevectors and energies:

$$\begin{aligned} R_{\text{rad}} &= \iint P_{\text{rad}}(E, \nu_{\mathbf{k}})\, \mathrm{d}\mathbf{k}\, \mathrm{d}E \\ &= \sum_{\text{i,f}} \iint \frac{4q^2 \hat{n} \nu_{\mathbf{k}}}{m^2 V c^3 \hbar^2} \left|M_{\text{i,f}}(\mathbf{k})\right|^2 f(E_{\text{f}}) \left(1 - f(E_{\text{i}})\right)\, L(\nu_{\mathbf{k}})\, \mathbf{k}\, \mathrm{d}\mathbf{k}\, \mathrm{d}E \end{aligned} \tag{1.64}$$

where f are the Fermi-Dirac functions and $L(\nu_{\mathbf{k}})$ is the wavevector-dependent Lorentian broadening.

Again, as well as for Auger processes, even the rate in Eq. (1.64) can be modeled via an empirical point of view. The typical simplified description is

$$\begin{aligned} R_{\text{rad}} &\simeq B\,(n_0 + \Delta n)\,(p_0 + \Delta p) \\ &= Bnp\,, \end{aligned} \tag{1.65}$$

being B, also called 'bi-molecular' coefficient, a material dependent parameter (in $\mathrm{cm^3 s^{-1}}$). The second equality in Eq. (1.65) is valid for intrinsic semiconductors. In these conditions

$$\tau_{i,\text{rad}} = \frac{n_i}{2R_{\text{rad}}} = \frac{n_i}{2Bnp}\,, \tag{1.66}$$

from which the radiative lifetime at equilibrium

$$\tau_{\text{rad}} = \frac{2n_i \tau_{i,\text{rad}}}{n + p}\,. \tag{1.67}$$

Coefficients $A = \tau_{\text{SRH}}^{-1}$, B_{rad} and $C_{n,p}$ define the so-called "ABC model" describing GR processes in semiconductors.

1.5 Classical Transport in Semiconductors

In order to describe electrical/optical operation of semiconductor devices and to model them the knowledge about carriers dynamics is essential. In addition, device simulation isn't only a strategy to predict specific structure behaviors, but it also represents today's workbench for investigating advanced device performances. Although solving the electrical problem may constitute a formidable mathematical challenge – especially in three-dimensional (3D) simulations – it is possible to simplify the real physics by modeling only the dynamic description of locally averaged variables through approximated solutions of opportune partial differential equations (PDEs). Such average values are obtained thanks to a geometrical discretization method, like the finite element (FEM), finite boxes (FB) or finite volume (FVM). These spatial discretization schemes allow to reduce the PDE equations to a discrete system of ordinary differential equations (ODEs) that it is possible to linearize under specific theoretical hypotheses.

In the present chapter we analyzed some of the most important GR mechanisms. Moreover, in Chapter 4 direct (BTBT) and defect-assisted tunneling (TAT) models

designed for physics-based numerical simulation of optoelectronic devices will be treated and finally expressed through opportune rates (number of carriers which tunnel per volume per second). These rates may correspond to generation rates, in case of BTBT in narrow band gap reverse biased diodes, or recombination rates, as for TAT in wide band gap forward biased junctions. This section has the purpose to clarify the framework in which both standard GR and such tunneling rates play and how they can enter into device-level numerical simulation.

1.5.1 Short recall about drift-diffusion (DD) model

The simpler starting point in semiconductor transport theory is represented by the *Poisson equation*, given by the PDE

$$\begin{aligned}\nabla^2_{\mathbf{r}}\phi(\mathbf{r}) &= -\frac{q}{\epsilon}\Bigg(N_{\mathrm{D}}(\mathbf{r}) - N_{\mathrm{A}}(\mathbf{r}) + N_{\mathrm{V}}(\mathbf{r})\,\exp\left(\frac{q\phi(\mathbf{r}) - \chi(\mathbf{r}) - E_{\mathrm{g}}(\mathbf{r})}{k_{\mathrm{B}}T}\right) \\ &\qquad - N_{\mathrm{C}}(\mathbf{r})\,\exp\left(\frac{-q\phi(\mathbf{r}) - \chi(\mathbf{r})}{k_{\mathrm{B}}T}\right)\Bigg),\end{aligned} \tag{1.68}$$

where ϕ is the electrostatic potential, ϵ the dielectric constant, N_{D}, N_{A}, N_{V} and N_{C} are, respectively, the concentrations of donors and acceptors and the effective densities in valence and conduction band, χ is the electron affinity and q the elementary charge. Note that Eq. (1.68) is written in its most generic form since all the possible spatial dependencies have been made explicit. By introducing now the Boltzmann statistics we have carrier distribution functions $f(\mathbf{k},\mathbf{r},t)$ that must obey the so-called *Boltzmann transport equation* (BTE):

$$\frac{\partial f(\mathbf{k},\mathbf{r},t)}{\partial t} + \mathbf{v}(\mathbf{k})\cdot\nabla_{\mathbf{r}} f(\mathbf{k},\mathbf{r},t) + \frac{\mathbf{F}}{\hbar}\cdot\nabla_{\mathbf{k}} f(\mathbf{k},\mathbf{r},t) = \left.\frac{\mathrm{d}f(\mathbf{k},\mathbf{r},t)}{\mathrm{d}t}\right|_{\mathrm{coll}} \tag{1.69}$$

where $\mathbf{v}(\mathbf{k})$ is a group velocity and $\mathbf{F}$ is the specific force field acting on the particles constituting our system, such that $\mathbf{F} = \hbar(\mathrm{d}\mathbf{k}/\mathrm{d}t)$. The left-hand side expression represents a *drift term* while $(\mathrm{d}f/\mathrm{d}t)|_{\mathrm{coll}}$ is a *collision term*. Furthermore, under relaxation time approximation (see Section 1.4):

$$\left.\frac{\mathrm{d}f(\mathbf{k},\mathbf{r},t)}{\mathrm{d}t}\right|_{\mathrm{coll}} = -\frac{|f(\mathbf{k},\mathbf{r},t) - f_0(\mathbf{k})|}{\tau(\mathbf{k})}, \tag{1.70}$$

where $f_0(\mathbf{k})$ refers to thermodynamic equilibrium.

Once the acting field is set, there are several techniques to solve the BTE. One of these is the *Hydrodynamic* (HD) model, suitable for systems with many electron-electron scattering events, like in submicrometric and micrometric devices, heterostructures or quantum wells. Its unknowns are the carrier concentrations n and p, their energy E and the electrostatic potential ϕ. Besides HD we find the *Drift-Diffusion* (DD) model, used in the analysis of micrometric devices (down to

0.5 μm) in diffusive transport regime, by self-consistently solving Eq. (1.68) and the *continuity equations* for electron/hole current densities $J_{n,p}(\mathbf{r},t)$

$$\begin{cases} \dfrac{\partial n(\mathbf{r},t)}{\partial t} = +\dfrac{1}{q}\nabla_{\mathbf{r}} J_n(\mathbf{r},t) - U_n(\mathbf{r},t) \\ \dfrac{\partial p(\mathbf{r},t)}{\partial t} = -\dfrac{1}{q}\nabla_{\mathbf{r}} J_p(\mathbf{r},t) - U_p(\mathbf{r},t) \end{cases}, \tag{1.71}$$

where $U(\mathbf{r},t) = G(\mathbf{r},t) - R(\mathbf{r},t)$ is the net generation/recombination rate, i.e. the algebraic summation of terms G and R, and current densities are

$$\begin{cases} \dfrac{J_n(\mathbf{r},t)}{q} = D_n \nabla_{\mathbf{r}} n(\mathbf{r},t) - \mu_n \left(\nabla_{\mathbf{r}}\phi(\mathbf{r},t)\right) n(\mathbf{r},t) \\ \dfrac{J_p(\mathbf{r},t)}{q} = D_p \nabla_{\mathbf{r}} p(\mathbf{r},t) - \mu_p \left(\nabla_{\mathbf{r}}\phi(\mathbf{r},t)\right) p(\mathbf{r},t) \end{cases}, \tag{1.72}$$

where $D_{n,p}$ are the diffusion coefficients and

$$\mu_{n,p} = \frac{\mathbf{v}_{n,p}^{\text{drift}}}{\mathcal{E}} = -\frac{\mathbf{v}_{n,p}^{\text{drift}}}{\nabla_{\mathbf{r}}\phi(\mathbf{r},t)} \tag{1.73}$$

the electron/hole mobilities.

The DD model is valid under the assumption according to which field variations produce in all the unknowns – $n(\mathbf{r},t)$, $p(\mathbf{r},t)$ and $\phi(\mathbf{r},t)$ – slower changes with respect to the energy relaxation time of the system τ. A formal link between BTE and DD model can be established through the so-called *method of moments*, according to which the *generalized statistical moment* is computed by integrating in the $\mathbf{k}$-space Eq. (1.69) multiplied by λ, an opportune function of $\mathbf{k}$:

$$\frac{\partial}{\partial t}\int_{\mathbf{k}} \lambda\, f \,\mathrm{d}\mathbf{k} + \nabla_{\mathbf{r}} \int_{\mathbf{k}} \mathbf{v}(\mathbf{k})\,\lambda\, f \,\mathrm{d}\mathbf{k} - \frac{\mathbf{F}}{\hbar}\int_{\mathbf{k}} \nabla_{\mathbf{k}}\,\lambda\, f \,\mathrm{d}\mathbf{k} = \int_{\mathbf{k}} \lambda \left.\frac{\mathrm{d}f}{\mathrm{d}t}\right|_{\text{coll}} \mathrm{d}\mathbf{k}\,, \tag{1.74}$$

where $f = f(\mathbf{k},\mathbf{r},t)$ and assuming

$$\begin{aligned} \lambda = \lambda(\mathbf{k}) &= \sum_{j=0}^{N} a_j\, \mathbf{k}^j \\ &= a_0 + a_1\,\mathbf{k} + a_2\,\mathbf{k}^2 + \dots \end{aligned} \tag{1.75}$$

with

$$\begin{cases} a_0 = 1 \\ a_1 = \hbar\,\mathbf{k} \\ a_2 = \dfrac{\hbar^2\,\mathbf{k}^2}{2\,m_e^*} \\ \quad \vdots \end{cases}. \tag{1.76}$$

Now we are able to define

$$\mathsf{M}_j = \int_{\mathbf{k}} \lambda_j(\mathbf{k})\, f\,(\mathbf{k},\mathbf{r},t)\ \mathrm{d}\mathbf{k} \tag{1.77}$$

as the so-called j-th order moment of $f\,(\mathbf{k},\mathbf{r},t)$. So one can see that the first three moments correspond, respectively, to the carrier concentration, the carrier average momentum and the carrier average energy. In particular, by using

$$\mathsf{M}_0 = \int_{\mathbf{k}} f\,(\mathbf{k},\mathbf{r},t)\ \mathrm{d}\mathbf{k} = n(\mathbf{r},t) \tag{1.78}$$

into Eq. (1.69) one can obtain the charge conservation law for electrons

$$\frac{\partial n(\mathbf{r},t)}{\partial t} + \nabla_{\mathbf{r}}\left(\bar{v}_n\, n(\mathbf{r},t)\right) = \left.\frac{\mathrm{d}n(\mathbf{r},t)}{\mathrm{d}t}\right|_{\mathrm{coll}}, \tag{1.79}$$

that, being

$$J_n = -q\,\bar{v}_n\, n(\mathbf{r},t) \tag{1.80}$$

and

$$\left.\frac{\mathrm{d}n(\mathbf{r},t)}{\mathrm{d}t}\right|_{\mathrm{coll}} = -U_n(\mathbf{r},t)\,, \tag{1.81}$$

coincides with the electron continuity equation, where

$$\bar{v}_n = \frac{\displaystyle\int_{\mathbf{k}} \mathbf{v}(\mathbf{k})\, f\,(\mathbf{k},\mathbf{r},t)\ \mathrm{d}\mathbf{k}}{\displaystyle\int_{\mathbf{k}} f\,(\mathbf{k},\mathbf{r},t)\ \mathrm{d}\mathbf{k}} \tag{1.82}$$

is the average electron velocity.

Now notice that Eq. (1.71) states that each carrier is characterized by a conservation law where the outgoing charge in the transport term $\nabla_{\mathbf{r}} J_{n,p}$ is (fully or partially) balanced by the generation/recombination contribute $U_{n,p}$. Not only electrons and holes obey this conservation principle but also fixed charges – like defects or dopants – do so, even if they don't have a transport counterpart at play. Defining the effective concentration of these level k as

$$C_k^{\pm}(\mathbf{r},t) = \begin{cases} N_{\mathrm{D},k}^{+}(\mathbf{r},t) & \text{for donors} \\ -N_{\mathrm{A},k}^{-}(\mathbf{r},t) & \text{for acceptors} \end{cases} \tag{1.83}$$

we can write the corresponding continuity equation

$$\frac{\partial C_k^{\pm}(\mathbf{r},t)}{\partial t} + \sum_k \Big(U_{n,k}(\mathbf{r},t) - U_{p,k}(\mathbf{r},t)\Big) = 0 \tag{1.84}$$

The final set of DD equations, in a 3D time-variant framework, is

$$\begin{cases} \dfrac{\partial n(\mathbf{r},t)}{\partial t} = -\nabla\Big(n(\mathbf{r},t)\,\mu_n\nabla_{\mathbf{r}}\phi(\mathbf{r},t) - D_n\,\nabla_{\mathbf{r}}n(\mathbf{r},t)\Big) - U_n(\mathbf{r},t) \\ \dfrac{\partial p(\mathbf{r},t)}{\partial t} = \nabla\Big(p(\mathbf{r},t)\,\mu_p\nabla_{\mathbf{r}}\phi(\mathbf{r},t) + D_p\,\nabla_{\mathbf{r}}p(\mathbf{r},t)\Big) - U_p(\mathbf{r},t) \\ \dfrac{\partial C_k^{\pm}(\mathbf{r},t)}{\partial t} + \displaystyle\sum_k \Big(U_{n,k}(\mathbf{r},t) - U_{p,k}(\mathbf{r},t)\Big) = 0 \\ \nabla_{\mathbf{r}}^2\phi(\mathbf{r},t) = -\dfrac{\rho(\mathbf{r},t)}{\epsilon} \end{cases}, \tag{1.85}$$

where the overall charge density ρ has become

$$\rho(\mathbf{r},t) = q\left(p(\mathbf{r},t) - n(\mathbf{r},t) + \sum_k C_k^{\pm}(\mathbf{r},t)\right) \tag{1.86}$$

and where all generation/recombination (and tunneling) mechanisms we discussed in the present chapter can enter by means of their G or R rate.

1.5.2 Boundary conditions

In this subsection we deal with the solution of PDEs which constitute the system (1.85), starting with a suitable ensemble of initial conditions (IC) and boundary conditions (BC), then by applying a certain spatial discretization allowing to reduce the PDE complexity.

Typically, the IC stem from the specific physical problem one has to solve, whereas BC are quite standard in numerical simulations and deserve special attention as well as the GR models included into the continuity equation. These conditions can be of *Neumann*-type if applied to the unknowns derivatives or of *Dirichlet*-type when applied to the unknowns only. They obviously can be mixed or used together contemporarily in different part of the device, according to the energy exchange between the system and the external environment. Anyway, in the first kind of BC the (simplest) homogeneous condition is defined as

$$\frac{\partial n(\mathbf{r},t)}{\partial \hat{\mathbf{n}}} = 0\,, \quad \frac{\partial p(\mathbf{r},t)}{\partial \hat{\mathbf{n}}} = 0 \quad \text{and} \quad \frac{\partial \phi(\mathbf{r},t)}{\partial \hat{\mathbf{n}}} = 0 \tag{1.87}$$

where $\hat{\mathbf{n}}$ is the unit vector normal to a given boundary surface or interface. The conditions stated in Eq. (1.87) are often used for surrounding insulators since they are equivalent to impose the requirement of a zero normal component of the current density, that corresponds to the situation given by an ideal insulation. In presence of a dielectric medium in place of the insulator the Neumann conditions are no more

homogeneous and read

$$\begin{cases} n(\mathbf{r},t)\,\mu_n\,\dfrac{\partial\phi(\mathbf{r},t)}{\partial\hat{\mathbf{n}}} = D_n\,\dfrac{\partial n(\mathbf{r},t)}{\partial\hat{\mathbf{n}}} \\ p(\mathbf{r},t)\,\mu_p\,\dfrac{\partial\phi(\mathbf{r},t)}{\partial\hat{\mathbf{n}}} = -D_p\,\dfrac{\partial p(\mathbf{r},t)}{\partial\hat{\mathbf{n}}} \\ \epsilon_{\mathrm{s}}\ \left.\dfrac{\partial\phi(\mathbf{r},t)}{\partial\hat{\mathbf{n}}}\right|_{\mathrm{s}} = \epsilon_{\mathrm{diel}}\ \left.\dfrac{\partial\phi(\mathbf{r},t)}{\partial\hat{\mathbf{n}}}\right|_{\mathrm{diel}} \end{cases} \tag{1.88}$$

Often applied at contacts, we find Dirichlet BC. Supposing to have ideal metals, i.e. ohmic (behaving as a short-circuit) and equipotential (the electrostatic potential $\phi(\mathbf{r},t) \equiv \phi(t)$ is the same at each point of the metal), our BC are

$$\begin{cases} n(\mathbf{r},t) = \dfrac{1}{2}\left(\sqrt{\displaystyle\sum_k C_k^{\pm 2}(\mathbf{r},t) + 4n_i^2} + \displaystyle\sum_k C_k^{\pm}(\mathbf{r},t)\right) \\ p(\mathbf{r},t) = \dfrac{1}{2}\left(\sqrt{\displaystyle\sum_k C_k^{\pm 2}(\mathbf{r},t) + 4n_i^2} - \displaystyle\sum_k C_k^{\pm}(\mathbf{r},t)\right) \\ \phi(t) = V_{\mathrm{bias}}(t) + \mathrm{const.} \end{cases}, \tag{1.89}$$

where the constant potential can be referred, eventually, to a built-in potential.

In case of Schottky rectifying contact a linear drop affects the potential across it. To solve the DD system the "Sze-Bethe theory" is often used, where current densities are re-written as functions of the surface recombination:

$$\begin{cases} \mathbf{J}_{\perp_n}(\mathbf{r},t) = q\,v_{\mathrm{surf}_n}(\mathbf{r},t)\Big(n(\mathbf{r},t) - n_{\mathrm{surf}}(\mathbf{r},t)\Big) \\ \mathbf{J}_{\perp_p}(\mathbf{r},t) = q\,v_{\mathrm{surf}_p}(\mathbf{r},t)\Big(p(\mathbf{r},t) - p_{\mathrm{surf}}(\mathbf{r},t)\Big) \\ \phi(t) = V_{\mathrm{bias}}(t) + V_{\mathrm{bi}} + \mathrm{const.} \end{cases}, \tag{1.90}$$

where $\mathbf{J}_\perp \triangleq \mathbf{J}\cdot\hat{\mathbf{n}}$, V_{bi} is the Schottky barrier built-in potential and where v_{surf}, n_{surf}, p_{surf} are the surface recombination velocity and carrier concentrations, respectively.

1.5.3 Discretization procedure

In order to implement an iterative solving method for the DD system – like Newton's one – and reduce the complexity of its coupled PDEs a spatial discretization is required. This procedure allows to relax the computational density by considering only average unknowns, evaluated at the nodes of a pre-calculated optimized mesh grid. Also the mathematical complexity is reduced, by passing from PDEs to ODEs and, finally, to linear algebraic equations.

Among the various discretization procedures here we present one of the most used, a technique also exploited in all the semiclassical simulation described in this thesis. Following the 1934 work by B. Delaunay, a Russian mathematician

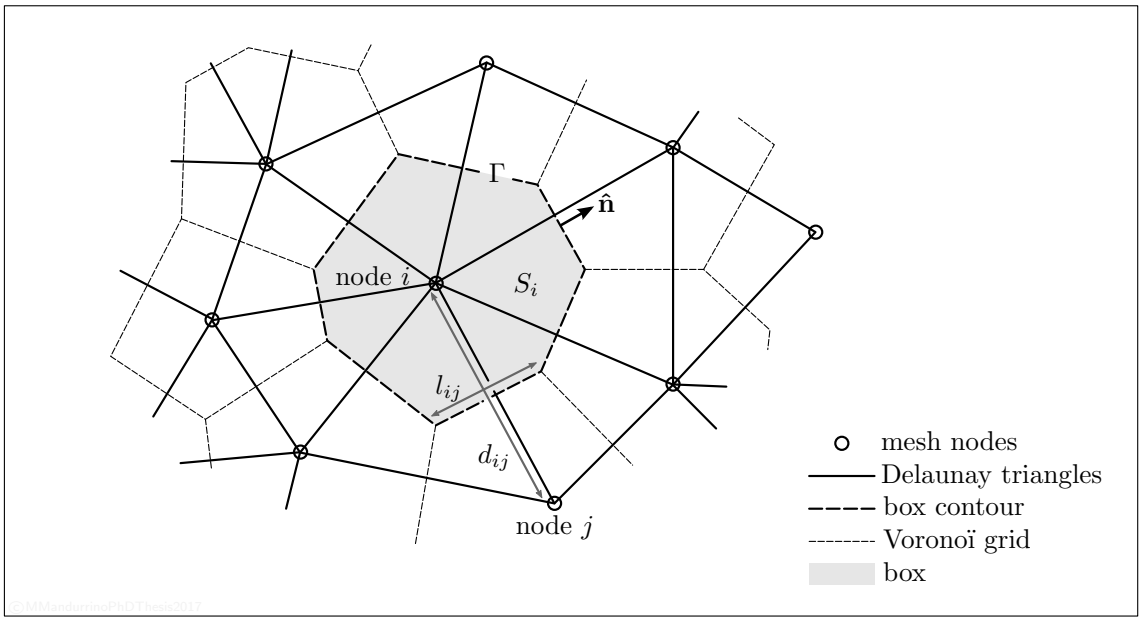

FIGURE 1.8. Delaunay-Voronoï finite boxes (FB) discretization scheme referred to the node i and one of its neighboors, the node j. Symbols are explained in the text.

(and mountain climber) [5], we make use of a procedure that generates a triangulation through non-obtuse triangles. Subsequently these triangular sub-domains are treated via finite boxes (FB) method [6], which computes the physical quantities under n-order approximation technique for each box and finally, by averaging, for each mesh node. At the first order all differential operators for scalar and vectorial quantities occurring in the DD model transform according to the following scheme (here in two dimensions)

$$\frac{\partial}{\partial t}\int_S x\,\mathrm{d}s \Rightarrow \frac{\partial x_i}{\partial t}\,S_i \tag{1.91}$$

$$\oint_\Gamma \mathbf{F}_\perp\,\mathrm{d}\gamma \Rightarrow \sum_j l_{ij}\,\langle \mathbf{F}_\perp\rangle_{ij} \tag{1.92}$$

$$\int_S c\,\mathrm{d}s \Rightarrow c_i\,S_i \tag{1.93}$$

where x and c are, respectively, a generic scalar unknown and constant, $\mathbf{F}$ is a generic vectorial unknown, S_i represents the area S (or the volume V, if we operate in 3D) of a finite box surrounding the i-th mesh node and Γ its contour, l_{ij} is the side of the box surrounding node i and located between i and j and, finally, $\langle \mathbf{F}_\perp\rangle_{ij}$ is the average value of the normal vector $\mathbf{F}_\perp$, which has its perpendicular foot lying along l_{ij}. Figure 1.8 shows a possible criterium through which FB is implemented: in this case we chose to represent the so-called Voronoï scheme, according to which boxes are defined by connecting the bisectors of each Delaunay triangle.

Through the Delaunay-Voronoï discretization procedure, and by applying a first-

order finite difference (FD) approximation, the DD system at a generic node i reduces to

$$\begin{aligned}\frac{q}{\epsilon}\left(\sum_k p_i(\mathbf{r},t) - n_i(\mathbf{r},t) + C^{\pm}_{k,i}(\mathbf{r},t)\right) S_i &= \sum_j l_{ij}\,\langle \mathcal{E}_\perp\rangle_{ij} \\ &\approx \sum_j l_{ij}\,\frac{\phi_i(\mathbf{r},t)-\phi_j(\mathbf{r},t)}{d_{ij}} \end{aligned} \tag{1.94}$$

and

$$\frac{\partial n_i(\mathbf{r},t)}{\partial t}\,S_i = \frac{1}{q}\sum_j l_{ij}\,\langle \mathbf{J}_{\perp_n}\rangle_{ij} - U_{n,i}(\mathbf{r},t)\,S_i\,, \tag{1.95}$$

where d_{ij} is the length of the segment connecting nodes i and j and where in the second line we used the 1-st order FD scheme

$$\langle \mathcal{E}_\perp\rangle_{ij} = -\frac{\partial \phi_{ij}(\mathbf{r},t)}{\partial \mathbf{r}} \approx \frac{\phi_i(\mathbf{r},t)-\phi_j(\mathbf{r},t)}{d_{ij}}\,, \tag{1.96}$$

being $\mathbf{r}$ a generic vector in space. Since at interfaces between boxes normal current densities $\mathbf{J}_{\perp_{n,p}}$ are conserved by construction, one can write:

$$\begin{aligned}\frac{1}{q}\langle \mathbf{J}_{\perp_n}\rangle_{ij} &\approx \frac{1}{q}\nabla \mathbf{J}_{\perp_n} \\ &= -\mu_n\, n(\mathbf{r})\,\frac{\partial \phi_{ij}(\mathbf{r},t)}{\partial \mathbf{r}} + D_n\,\frac{\mathrm{d}n(\mathbf{r},t)}{\mathrm{d}\mathbf{r}} \\ &\approx -\mu_n\, n(\mathbf{r},t)\,\frac{\phi_i(\mathbf{r})-\phi_j(\mathbf{r},t)}{d_{ij}} + D_n\,\frac{\mathrm{d}n(\mathbf{r},t)}{\mathrm{d}\mathbf{r}} \end{aligned} \tag{1.97}$$

where total derivatives have been used in place of partial derivatives thanks to the assumption of constant $\mathbf{J}_{\perp_n}$ between two nearby nodes. By solving Eq. (1.97) in the unknown n we obtain the non-linear solution

$$n(\mathbf{r},t) = n_i(\mathbf{r},t)\;[1-g\,(\mathbf{r},\Delta_{ij})] + n_j(\mathbf{r},t)\,g\,(\mathbf{r},\Delta_{ij}) \tag{1.98}$$

in which

$$\Delta_{ij} = q\,\frac{\phi_i(\mathbf{r},t)-\phi_j(\mathbf{r},t)}{k_{\mathrm{B}}T} = \frac{\phi_i(\mathbf{r},t)-\phi_j(\mathbf{r},t)}{V_{\mathrm{T}}} \tag{1.99}$$

and

$$g\,(\mathbf{r},\Delta_{ij}) = \frac{1-\mathrm{e}^{\frac{\mathbf{r}\,\Delta_{ij}}{d_{ij}}}}{1-\mathrm{e}^{\Delta_{ij}}}\,. \tag{1.100}$$

According to the so-called Scharfetter-Gummel (SG) approximation, ODEs can be spatially linearized through

$$\frac{1}{q}\langle \mathbf{J}_{\perp_n}\rangle_{ij} \overset{\mathrm{SG}}{\approx} \frac{D_n}{d_{ij}}\,\left[n_j(\mathbf{r},t)\,\mathsf{B}\,(\Delta_{ij}) - n_i(\mathbf{r},t)\,\mathsf{B}\,(-\Delta_{ij})\right]\,, \tag{1.101}$$

where

$$\mathsf{B}(\alpha) = \frac{\alpha}{e^{\alpha} - 1} \tag{1.102}$$

is the Bernoulli function. Finally:

$$\frac{\partial n_i(\mathbf{r},t)}{\partial t} = \sum_j \frac{D_n\, l_{ij}}{d_{ij}\, S_i} \left[n_j(\mathbf{r},t)\, \mathsf{B}\left(\Delta_{ij}\right) - n_i(\mathbf{r},t)\, \mathsf{B}\left(-\Delta_{ij}\right) \right] - U_{n,i}(\mathbf{r},t) \tag{1.103}$$

and

$$\frac{\partial p_i(\mathbf{r},t)}{\partial t} = -\sum_j \frac{D_p\, l_{ij}}{d_{ij}\, S_i} \left[p_i(\mathbf{r},t)\, \mathsf{B}\left(\Delta_{ij}\right) - p_j(\mathbf{r},t)\, \mathsf{B}\left(-\Delta_{ij}\right) \right] - U_{p,i}(\mathbf{r},t) \tag{1.104}$$

while, concerning fixed charges:

$$\frac{\partial C^{\pm}_{k,i}(\mathbf{r},t)}{\partial t} = -\sum_{j \neq k} \left(p_{k,ji}(\mathbf{r},t) - p_{k,ij}(\mathbf{r},t) \right). \tag{1.105}$$

The system composed by Eq. (1.94), (1.103), (1.104) and (1.105) represents the FB discretized version of the DD model for each i-th node of the Delaunay grid. From this set of equations is now possible to extract the approximated current at the contacts

$$\begin{aligned} I_{\text{cont}} \approx \sum_{i,j} \Bigg[& \frac{q\, D_n}{d_{ij}} \left(n_j(\mathbf{r},t)\, \mathsf{B}\left(\Delta_{ij}\right) - n_i(\mathbf{r},t)\, \mathsf{B}\left(-\Delta_{ij}\right) \right) \\ & + \frac{q\, D_p}{d_{ij}} \left(p_i(\mathbf{r},t)\, \mathsf{B}\left(\Delta_{ij}\right) - p_j(\mathbf{r},t)\, \mathsf{B}\left(-\Delta_{ij}\right) \right) \Bigg]\, d_{\text{cont}}\, l^*_{ij} + \\ & + q \left(-\frac{\partial n_i(\mathbf{r},t)}{\partial t} + \frac{\partial p_i(\mathbf{r},t)}{\partial t} - U_{n,i}(\mathbf{r},t) + U_{p,i}(\mathbf{r},t) \right) S_i\, d_{\text{cont}}\,, \end{aligned} \tag{1.106}$$

where d_{cont} is the contact length an

$$l^*_{ij} = \begin{cases} l_{ij} & \text{for internal nodes } j \\ \dfrac{l_{ij}}{2} & \text{for nodes } j \text{ on the boundary} \end{cases}. \tag{1.107}$$

So far we have seen how to rewrite the drift-diffusion PDE system into a system of ODEs via spatial discretization, subsequently linearized in space thanks to the SG approximation. Also time evolution has to be treated under the numerical standpoint. In order to do that, simulators exploit the so-called *quasistationary* procedure, where time-variant quantities are considered as fixed step-by-step quantities in a sort of discretized time scale in which at each instant the quasi-equilibrium approximation is assumed. Iterating this procedure for a certain number of states one can reproduce the time variation of any parameter of interest, like in case of voltage or current ramps at contacts.

Chapter 2

Narrow Gap Infrared Photodetectors (IRPDs)

This chapter will focus on one of the most prolific and interesting material within the electronic industry working on infrared (IR) light, both in terms of *quantum efficiency* and, in turn, by the standpoint of its commercial business and diffusion within the framework of solid state IR detection: the narrow-gap ternary alloy HgCdTe, also knows as "MCT" (from the acronym of mercury cadmium telluride, or also "Mer-Ca-Tel"). Besides introducing MCT, we will also briefly present the physical principles at the basis of light detection and, in particular, for what concerns the IR spectrum (conventionally, from $\lambda \simeq 0.7$ μm to $\lambda \simeq 1000$ μm).

The chapter is organized as follows: first, in Section 2.1, a review about the technological development in the field of IR detection is presented, and the reasons leading to modern electronics intended for these applications are also exposed. Then, in Section 2.2 a set of material properties, essential to understand the operating principles of MCT detectors, will be listed and commented. Finally, Section 2.3 will show the state of the art about HgCdTe-based electronic devices for IR detection, their physical and optoelectronic properties, the most relevant figures of merit and performance limitations, with a special attention to quantum processes, the main topic of Part II of this thesis.

2.1 Historical Overview

2.1.1 The origins

At its beginning, in the early XIX century, the IR detection was entirely relegated to thermometric measurements [7]. This means that the only way to explore this range of electromagnetic spectrum consisted in exploiting the *thermoelectric effect* discovered by T. J. Seebeck in 1821, which is correlated to the capability of a material

to change its electrical potential as a function of the thermal gradient or, if you want, as a function of the absolute temperature T, by choosing a reference within the material as the 'zero-point' of thermal distribution. The technological milestones of this era have been essentially fixed by three different inventions: the thermometer, introduced by Sir W. Herschel in 1800, then the thermopile, invented by L. Nobili in 1829, an finally the bolometer in 1881, by S. P. Langsley.

In parallel, discovering the *photoconductive effect* by W. Smith dated 1873 [8], which consists in the conductance variation of matter under lighting conditions (in the original study a sample of selenium was used), determined an increasing interest in capturing the electromagnetic radiation through solid-state principles in order to generate an electrical signal. In this sense, the first device exploiting the mechanism of *photoexcitation* to transduce the energy of incoming photons into an electrical potential can be considered the IR photo-electric detector composed of thallium, oxygen and sulphur developed by T. W. Case in 1917 and described in his 1920 work [9].

Following the same methodological approach of Case, in the following years, and especially throughout the World War I, other researchers focused their study on different materials, as lead or thallium sulfides, in order to improve the efficiency of photoelectric effect under IR illumination. This is the way the modern solid-state IR detection was born.

2.1.2 Extrinsic detection

During the World War II, with the theoretical achievements of Quantum Mechanics, initial difficulties affecting the first generation of devices made of polycristalline salts, especially by the fabrication point of view, were partially fixed thanks to the definitive introduction of bulk semiconductor alloys (in those early days, PbSe or PbTe) as the best choice for solid-state light detection. This, still active, fortunate season has been further strengthened by the invention of the transistor, due to J. Bardeen, W. B. Shockley and W. H. Brattain in December 1947, and also by setting up various techniques in order to introduce and control impurity elements in solids (as in Ge, for instance: Cu, Zn, Au or Hg). This particular innovation opened the way to a new mechanism in the family of photoexcitation processes, namely the *extrinsic absorption*. As represented in panel (a) of Figure 2.1, this mechanism involves the energy of states located within the forbidden gap that very often correspond to those dopant atoms we just referred to. In this way the electronic transition consequent to the released of the photon energy $E_{\mathrm{ph}} = h\nu$ can occur even for $E_{\mathrm{ph}} \leq E_{\mathrm{g}}$, since electron-phonon scattering here can also be 'partial' (*inelastic*), as well as 'total' (*elastic*), regardless the energy gap amplitude E_{g}.

The two major conflicts characterizing the XX century, the World War II and the Cold War, pushed scientists' mind in the direction of an increasing interest in IR detection, mainly driven by telecommunication needs for military purposes. These applications concentrated their attention on two specific wavelength windows for atmospherical signal transfer, named Mid-Wavelength Infra Red (MWIR) band, which is ranging from 3 to 5 μm, and Long-Wavelength Infra Red (LWIR) band,

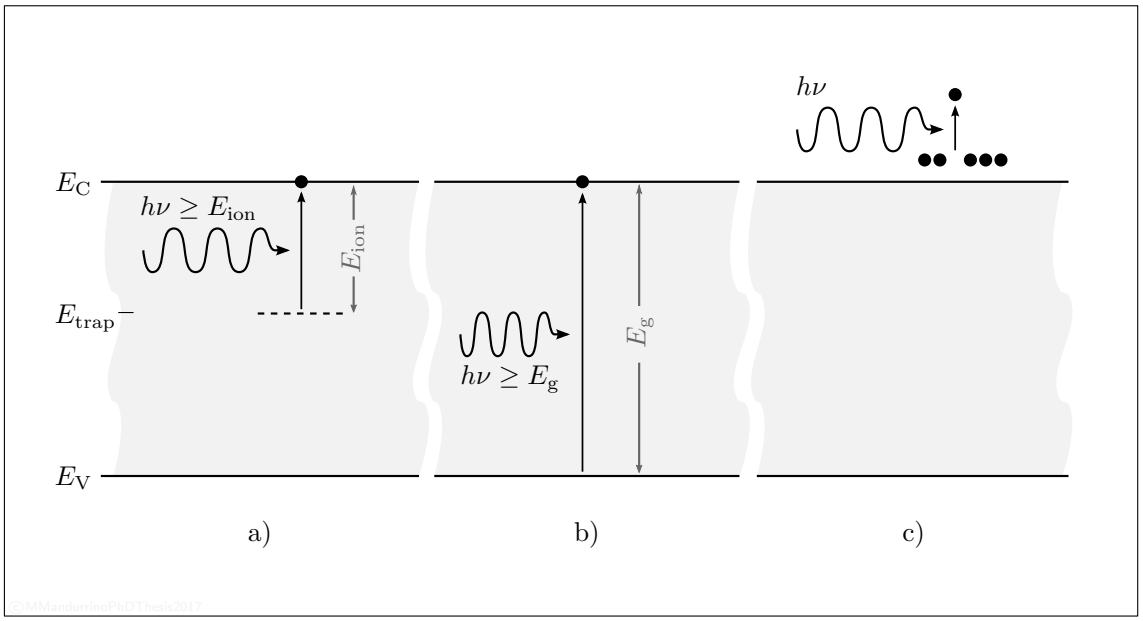

FIGURE 2.1. Representation of three fundamental photoexcitation processes in semiconductors: (a) *Extrinsic absorption* between a (donor) trap and conduction band. (b) *Intrinsic absorption*, i.e. direct transition between band edges. (c) *Free carrier absorption*.

from 8 to 14 μm. Besides these two, the Very Long-Wavelength Infra Red (VLWIR) band, from 15 to 30 μm, completes the set of a typical domain for the Ge-based photodetection exploiting the extrinsic absorption principle.

2.1.3 Intrinsic detection

Being trap-to-band transitions due to standard thermal fluctuations in atmosphere competitive with extrinsic absorption of IR radiation at the approximate wavelength of 3 μm, beyond this threshold solid-state IR detection is heavily characterized by stringent cooling requirements that must be overcome. To this purpose, a second generation of detectors, based on *intrinsic absorption*, depicted in panel (b) of Figure 2.1, were introduced in the market with the hope to get rid of all technological and physical issues related to the implantation and the subsequent controllability of impurities, and without the need of cooling systems. These efforts could, in principle, also produce a theoretical gain in terms of device reliability and, in turn, by the standpoint of quantum efficiency (a concept that will be properly developed in next sections). Then, to stimulate direct band-to-band transitions in LWIR and VLWIR sub-domains the use of narrow-gap semiconductors became essential. These materials are characterized by $E_g \leq 0.5$ eV or, in other words, they can exhibit an absorption decay of 3 dB at $\lambda_{\text{cut-off}} \simeq 2$ μm. In the first era (early 50s of the last century) III-V alloys have been employed: one of them, for instance, has been the compound InSb, a zincblende crystal with $E_g = 0.17$ eV at room

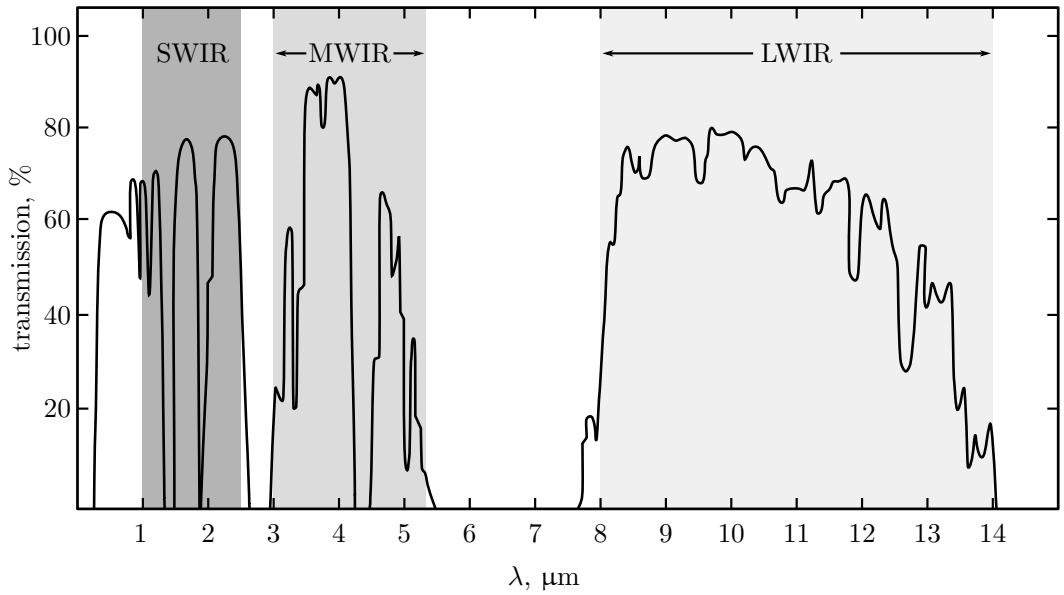

FIGURE 2.2. IR transmission spectrum in atmosphere for a 2 km long horizontal path at sea level (data published in Ref. [7]) where the three main ground atmospheric windows, SWIR, MWIR and LWIR, are highlighted. The lack of transmission around 6 μm and above 14 μm is due to the absorption bands of H_2O and CO_2, respectively.

temperature and with a gap broadening law of the form [10]:

$$E_g(T) = \alpha - \frac{\beta T^2}{T + \gamma}, \tag{2.1}$$

a typical behavior in narrow gap semiconductors, where in this specific case usual values are $\alpha = 0.24$ eV, $\beta = 6 \cdot 10^{-4}$ eV and $\gamma = 500$ K [10].

2.1.4 The modern era

After InSb, other semiconductor alloys from IV-VI and II-VI groups have been synthesized and exploited to cover the most part of the IR detectable spectrum. In the late 1950s and at the beginning of 60s also ternary compounds, as $Pb_{1-x}Sn_xTe$ (IV-VI), were introduced with the aim of having tailored bandgaps through a compositional engineering. Figure 2.3 shows the phase-space of some of the most interesting ternary and binary alloys for IR detection, represented in the plane lattice constant versus E_g. The plot allows us to better appreciate the matching properties of these alloys and, at the same time, their operating range in the electromagnetic domain.

Besides some of the most famous materials of this category as $InAs_{1-x}Sb_x$ (III-V), also MCT is represented in Figure 2.3. As shown, its precursors are the binaries HgTe and CdTe, that are linked by an almost vertical straight line representing all the possible states $Hg_{1-x}Cd_xTe$ (II-VI). By exploiting this very low lattice mismatch, in 1959 W. D. Lawson and co-workers [11] demonstrated for the first time

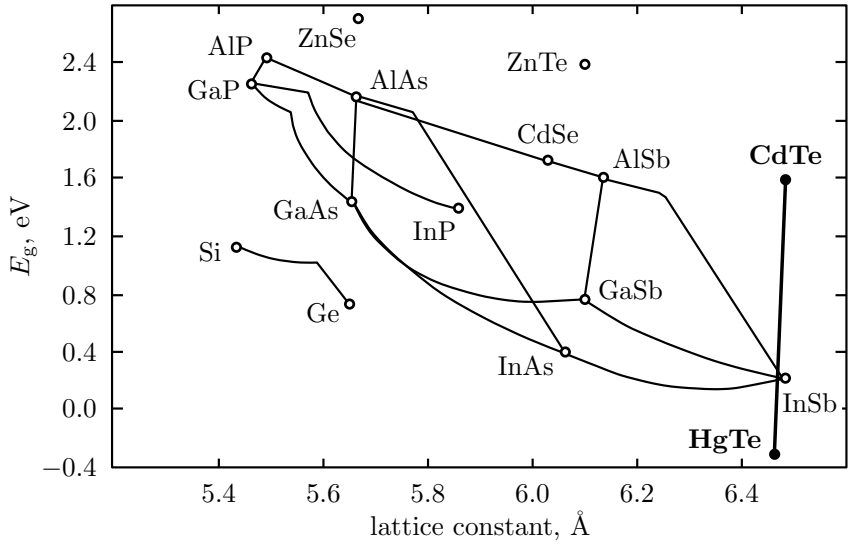

FIGURE 2.3. Phase-space of binary and ternary alloys for IR detection representing E_g versus the lattice constant. Points represent the binary compounds, while lines are the transitional states among them.

its producibility. In their paper the innovative properties of MCT have been highlighted, including the above mentioned capability of band gap engineering, from the configuration HgTe (with cadmium mole fraction $x = 0$), which ideally is a pure gapless semimetal alloy with $\lambda_{\text{cut-off}} \simeq 38$ µm, to the binary semiconductor CdTe ($x = 1$) connoted by $\lambda_{\text{cut-off}} \simeq 0.8$ µm. In Section 2.2 a detailed review of all the relevant physical properties of $Hg_{1-x}Cd_xTe$ will be provided.

Despite the high vapour pressure of Hg represents one of the most critical issues in the MCT fabrication technology, this alloy has probably become the most studied and exploited material of modern solid-state IR detection. Nevertheless, using HgCdTe rather than other semiconductors – as, among all, the well appreciated Si or GaAs – brings some particular benefits in terms of optical absorption and carrier mobility, especially with respect to other narrow-gap colleagues like InAsSb.

Thanks to less stringent cooling systems of IR detectors required by intrinsic generation rather then extrinsic one, and due to the progresses of new lithography standards, the late 60s have been accompanied by the advent of devices which have got rid of the detrimental competition between thermal generation and photogeneration. This ultimate generation of LWIR sensors, defined as Background-LImited Performance (BLIP) detectors, are characterized by a limitation exclusively coming from the noise generated by the system background flux and not by the sensor itself, as well explained by M. A. Kinch in his book [12]. So, these devices started to be produced in Focal Plane Array (FPA) configuration and with a single cryogenic stage. The 70s, instead, were distinguished by the CCD introduction due to W. S. Boyle, G. E. Smith and M. F. Tompsett [13, 14]. This invention coincided with the development of system and device architectures increasingly complex and com-

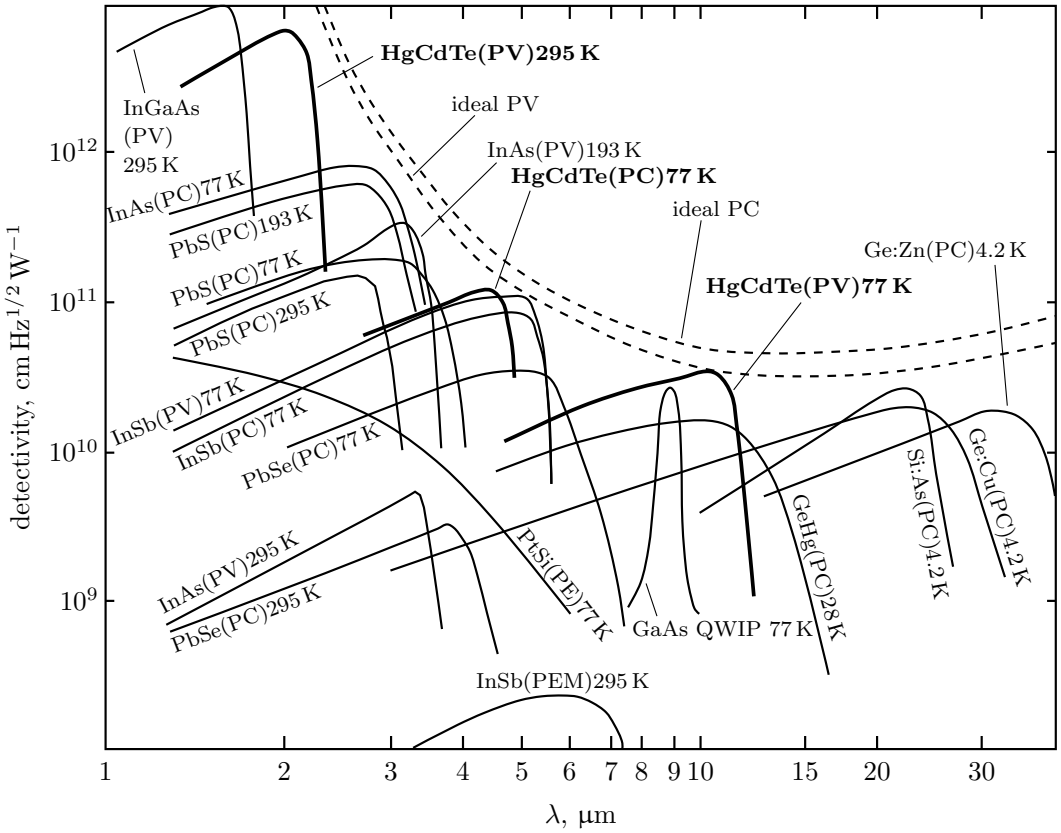

FIGURE 2.4. Detectivity curves measured for photovoltaic (PV), photoconductive (PC), photoemissive (PE) and photoelectromagnetic (PEM) devices at 300 K. Data published in Ref. [7].

pact, combining photon absorption and signal read-out in the same chip (monolithic detectors).

However, organizing electronics and detection all together very soon imposed even more attention to some physical features like tunneling effect within the photodiode junction, electromagnetic cross-talk between (CCD or FPA) pixels and field interactions involving the integrated sensor and the surrounding system. Considering also the high-impedance requirements in interfacing the detector with field-effect transistors (FETs) in order to properly manage the signal, what became good solutions are: PtSi Schottky barriers, InSb and HgCdTe photodiodes or PbSe, PbS and extrinsic Si photoconductors (PC).

At the beginning of 80s and later two elements driven the IR technology towards the use of MCT: on the one hand the invention of SPRITE (Signal PRocessing In The Element) devices exploiting the Time Delay and Integration (TDI) technology in a single elongated detector element, which pushed the interest in HgCdTe-based PC detectors; on the other hand the need of low-power dissipation and high-impedance went in the direction of photovoltaic (PV) configurations, leading to 2D-array devices still in use (and under continuous development).

TABLE 2.1. Infrared Photodetectors

TYPE	TRANSITION	SIGNAL	EXAMPLES
intrinsic	interband	photoconductive	PbS, PbSe, InSb, HgCdTe
		photovoltaic	InSb, InAs, PbTe, HgCdTe, PbSnTe
		capacitive	InSb, HgCdTe
extrinsic	from impurity to CB	photoconductive	Si:In, Si:Ga, Si:As, Ge:Cu, Ge:Hg
free carriers	intraband	photoemissive	Schottky barriers GaAs/CsO, PtSi, Pt_2Si, IrSi
		photoconductive	InSb electron bolometer
QWs	to/from quantized levels	photoconductive	InGaAs/AlGaAs, InAs/InGaSb, HgTe/CdTe, GaAs/AlGaAs
		photovoltaic	InSb/InAsSb, InAs/InAsSb

Comparison between different technological solutions for IR detection (adapted from Ref. [7]).

In their most recent version, these devices have been optimized for near-room-temperature operating regimes. This is what is called 'High-Operating Temperature' (HOT) technology. To incorporate specific performance demands, also multi-color detectors are under development. They represent, probably, the ultimate frontier of high-quality IR detection for night vision applications.

Besides the structure engineering, very active on the side of Quantum Well Infrared Photodetectors (QWIPs) or quantum dots, by the standpoint of materials and systems choice the framework is rather steady. The only real competitors of HgCdTe, which remains the best compromise between technological issues and performances with a billion-dollar business, are Schottky barriers on Si, SiGe heterojunctions, AlGaAs multiple QWs, GaInSb strain layer superlattices, high temperature superconductors and thermo-detectors, some of which are listed in Table 2.1. There, the most important characteristics of some examples of modern IR photon detectors are reported, classified by type, physical mechanism (see the column 'transition') and output signal.

2.2 MCT Material Properties

This section is devoted to the analysis of some relevant physical parameters and properties of $Hg_{1-x}Cd_xTe$ alloys, which in the last 30 years has represented and still represents one of the most interesting materials for IR detection. Thanks to its noticeable performance in terms of quantum efficiency, absorption coefficient, detectivity and responsivity [15], and since the high operating temperature, especially for LWIR detection, the gap tunability and, in turn, the possibility to have

tailored cut-off wavelengths, mercury cadmium telluride has become *the* material for IR sensors, and its properties now are well known. In particular, it is widely accepted that HgCdTe-based photodetectors can exhibit both better performances and more robust reliability by the standpoint of doping stability and regularity of composition and wafer surface with respect to traditional Si extrinsic devices [7], which are characterized by a wider (and indirect) bandgap. All these features allowed the spread of MCT photodetectors in all fields of IR industry, from military purposes to astronomy, passing through medical diagnostics and quality control in food industries.

As a matter of fact, the direct bandgap of $Hg_{1-x}Cd_xTe$ – null (or, better, negative) in HgTe and going up to about 1.6 eV in CdTe – is responsible for the high absorption coefficient in the IR domain but, at the same time, presents the drawback of being rather narrow, seen from the quantum standpoint. Indeed, since 60s of the last century the presence of quantum transport effects in semiconductors – among which, tunneling – was perfectly known, as attested by A. G. Chynoweth and H. S. Sommers works [16, 17]. Then J. A. del Alamo and R. M. Swanson in 1986 [18] and, soon later, D. K. Blanks [19] and Y. Nemirovsky [20] in 1988 and 1989, respectively, merged previous theories pointing out the high relevance of direct band-to-band tunneling (BTBT) transitions in narrow-gap semiconductors, especially in both forward and reverse biased HgCdTe junctions.

Not only direct tunneling is influenced by the narrowness of bandgap: also thermal-assisted processes, especially the Shockley-Read-Hall (SRH) [21], are enhanced by the gap narrowing with decreasing temperature $E_g(T)$, as occurs in the phenomenological law described by Eq. (2.1). And, in general, all band-trap-band processes might be driven by T, as well as by the impurity energy E_{trap}. A good representative of this family is trap-assisted-tunneling (TAT), which combines both energetic and spatial transitions. Since in HgCdTe, as we will see later, the curvatures of conduction band (CB) and light-hole (LH) band differ from a negligible quantity (in first approximation) then the maximum probability of TAT and SRH processes probabilities occurs for transitions via near-midgap states. As stated in Ref. [7, 12, 22], having the most used n- and p-type dopants for HOT applications ionization energies of few units/tens of meV (levels are quite far from band edges) then we can conclude that in HgCdTe the presence of defects rather than doping is the main responsible for generation-recombination (GR) mechanisms via intergap states, like SRH and TAT.

Even if controlling the formation of defects in HgCdTe during fabrication (epitaxy, MOCVD or MBE) is still a crucial issue, detecting in the LWIR range mainly depends on the bandgap, that must be as narrow as possible. This enhances another class of processes, i.e. the already mentioned direct transitions. By the engineering standpoint one solution is to properly design the shape of bands in correspondence of the junction, in order to reduce the longitudinal electric field. Unfortunately, a compromise has to be chosen in order to find the trade-off between leakage currents, due to band-to-band transitions, and the increasing of other detrimental phenomena which are dominant at low fields and room temperature – like dynamical resistance and Auger GR – and responsible for shot noise and Johnson noise occurrence [23].

Here the temperature comes back again as a subtle parameter that allows to mediate between two competitive leakage mechanisms: tunneling and Auger.

Recent experience shown up the features that an optimized HgCdTe structure for IR should have. First of all a proper passivation should be adopted to prevent surface GR leakage currents [24], then the active region should be lightly doped (to mitigate the junction field). By the structure design point of view, moreover, several solutions have actually been proposed. We will analyze some relevant device structures in the next section but, for the time being, one can say that, among the most promising, those worthy of note are the active regions realized with wide-gap barriers or double-junctions. An interesting example from the latter class of devices can be found in Ref. [25], where a p^+-ν-n^+ structure is proposed through experimental characterization and numerical modeling. The reverse bias operating principle of such device is based on removing more holes in unit time from the ν absorbing active region through the p^+-ν interface (which, therefore, is called *extraction junction*) rather than from the n^+ layer via the ν-n^+ *exclusion junction*. This yields a hole concentration in the active region lower than p at thermal equilibrium, which finally results in a drop of the Auger recombination rate and in the subsequent higher operating temperature (up to about 145 K in the LWIR window). This increase of T also allows a slight bandgap widening which implies a beneficial reduction of band-to-band transitions like BTBT.

Several other mechanisms are reported in the literature as source of dark/leakage current in HgCdTe-based photodiodes: surface state current, thermal diffusion of minority carriers, ohmic current and impact ionization from high energy carriers. But presently only tunneling and Auger seem to be the major limitations to devices performance. To clarify this behavior we will now analyze the main solid-state and electronic issues, as well as the most relevant HgCdTe properties: its crystal arrangement, the band structure and some important features relevant for carrier transport.

2.2.1 Lattice growth and structure

The two primary II-VI compounds which constitute MCT, CdTe and HgTe, are both made of a transition metal (Cd and Hg) bound to a tellurium atom, a metalloid element with six valence electrons ($5s^2\,5p^4$). The lattice constants of an $Hg_{1-x}Cd_xTe$ unit cell follows the empirical law [7]

$$a_0(x) = 6.4614 + (8.4\,x + 11.68\,x^2 - 5.7\,x^3)\,10^{-3}, \tag{2.2}$$

where a_0 is expressed in Ångström unit, going from $a_0(0) = 6.4614$ Å in HgTe to $a_0(1) = 6.4758$ Å of CdTe, with a negligible mismatch between them. In the resulting zincblende cell of a given ternary alloy (i.e. fixing the mole fraction x), two interpenetrating face-centered-cubic ("fcc") sub-cells displaced by a vector $\left(\frac{a_0}{4}, \frac{a_0}{4}, \frac{a_0}{4}\right)$ coexist. These two sub-lattices are such that Te always occupy one of the two, whereas Cd and Hg (whose configurations are $5s^2$ and $6s^2$, respectively) the other one, sharing the sites around Te almost randomly although x has been technologically fixed during the growth process. Moreover, being four the net number of

atoms per unit cell – two cations (Cd^+ or Hg^+) and two anions (Te^-) – there are five possible configurations of Cd and Hg as neighbor sites of Te, all located along tetrahedral directions (see Figure 2.5 and Table 2.2).

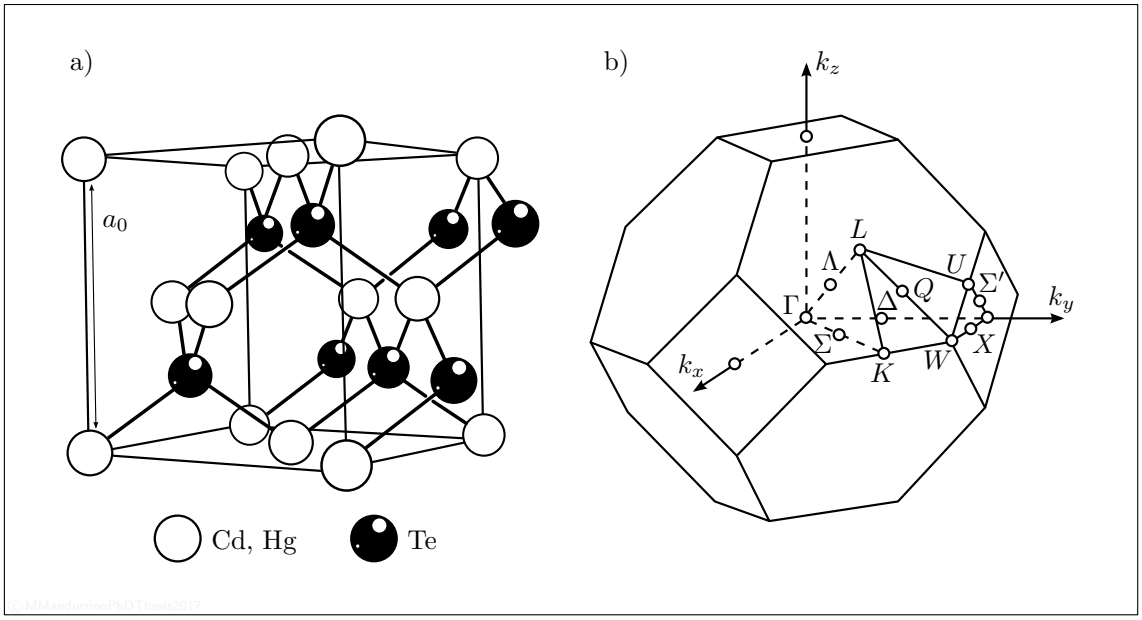

FIGURE 2.5. (a) Conceptual representation of the HgCdTe zincblende "fcc" lattice structure. (b) First Brillouin zone (FBZ) and its standard symmetry points. Adapted and re-edited from Ref. [26].

Since in Te the lowest unfilled electronic state is less energetic then the topmost filled states of Cd and Hg, the bonds Cd-Te and Hg-Te are mainly covalent, with a slight ionic component due to a deeper potential located near the anion that allows a net charge exchange. Figure 2.5 shows, in the left panel, a representation of the two merged sub-lattices in a standard conformation, i.e. without accounting for statistical deviations from the expected "fcc" configuration. However, it is worth noting that both long-range and short-range fluctuations in composition may occur. In particular, a Bernoulli distribution $Z(x)$ of charges depending on the mole fraction

TABLE 2.2. Cation-cation complexes in HgCdTe

ARRANGEMENT INDEX, i	IONIC COMPLEX	CELL COMPOSITION
0	four Hg-ions	HgTe
1	four Cd-ions	CdTe
2	two Cd-ions and two Hg-ions	$Hg_{0.5}Cd_{0.5}Te$
3	three Cd-ions and one Hg-ion	$Hg_{0.25}Cd_{0.75}Te$
4	three Hg-ions and one Cd-ion	$Hg_{0.75}Cd_{0.25}Te$

Classification of Te-neighbor complexes occurring in HgCdTe which determine the cell composition.

x with the form of Eq. (2.3) can be observed [26]:

$$Z(x) = \binom{4}{i} x^i \, (1-x)^{4-i}, \tag{2.3}$$

where $i = 0, 1, 2, 3, 4$ indexes the five configurations of neighbor cations around Te (see Table 2.2). From this distribution the probability of finding Cd-Cd or Hg-Hg complexes arises, which are respectively expressed by:

$$\begin{cases} P_{\text{Cd-Cd}}(x) = x + \gamma\,(1-x) \\ P_{\text{Hg-Hg}}(x) = (1-x) + \gamma\,x\,. \end{cases} \tag{2.4}$$

In Eqs. (2.4) the short-range order (SRO) parameter $\gamma \in [-1, 1]$ varies depending on the energy factor

$$\varepsilon = \frac{2\,E_{\text{Hg-Cd}}}{E_{\text{Hg-Hg}} + E_{\text{Cd-Cd}}}\,, \tag{2.5}$$

in which the bonding energies of the two binary complexes are involved (see also Table 2.3).

TABLE 2.3. Lattice Arrangements in HgCdTe

ENERGY PARAMETER, ε	SRO, γ	ALLOWED/ENHANCED CONFIGURATION
0	1	only $i = 0$ (HgTe) or $i = 1$ (CdTe) allowed
< 1	> 0	$i = 0$ or $i = 1$ populations enhanced
1	0	random fluctuations as described by Eq. (2.3)
> 1	< 0	$i = 2$ (both HgTe and CdTe) population enhanced
$\gg 1$	-1	ordered crystal

List of main HgCdTe lattice configurations in terms of cation-cation complex energy and probability (adapted from Ref. [26]).

What determines one of the most crucial parameters in MCT alloys, the cadmium mole fraction, is the growth process: in this regard, just consider that only an uncertainty of $\Delta x < 0.003$ is tolerated in order to manage cut-off fluctuations of detectable LWIR window as $\Delta\lambda_{\text{cut-off}} < 1$ µm [27]. This requirement could seem too stringent but, as we will see in a moment, it can be even overcome.

When Lawson and co-workers synthesized the first HgCdTe sample, in 1959, they did grow the two CdTe and HgTe crystals separately. Successively they mixed them in a vacuum chamber, heated slightly above the melting temperature, until the ternary state of the alloy was steadily reached. Nowadays there are essentially four primary methods to epitaxially grow HgCdTe from its constituents: bulk crystal growth, liquid phase epitaxy (LPE), molecular beam epitaxy (MBE) and vapor phase epitaxy (VPE).

Belonging to the first class of bulk processes we find the "Bridgman method" [28] (somehow similar to the "Czochralski method" used for Si or GaAs growth) which allows to produce pretty big ingots of material from liquid to solid phase through a furnace where, starting from a seed of nucleation, a slowly shifting temperature gradient is applied along the forming structure when it comes out of the furnace, providing a progressive crystallization. Among epitaxial methods we find the "solid state re-crystallization" which exploits a first stoichiometric mixture of Hg, Cd and Te close to the Te-HgTe eutectic point ($T^{\mathrm{e}}_{\mathrm{HgTe}} \simeq 970$ K) and then a fast rise of temperature as high as above the melting point ($T^{\mathrm{m}}_{\mathrm{HgTe}} > 1050$ K), where HgCdTe then grows spontaneously. Finally a rapid quench cooling is provided to the furnace in order to obtain a pretty fast crystallization ($t_{\mathrm{cr}} \simeq 250$ µm/h) that allows to lower the compositional fluctuations up to $\Delta x < 0.002$.

The second epitaxial process is the LPE. It essentially consists in the precipitation of a saturated precursor solution on the substrate surface. The mole fraction (again, $\Delta x < 0.002$) can be predicted and managed by controlling both the relative concentration of each reagent and the processing temperature (620 K to 820 K). As in all epitaxial techniques, the growth temperature is sensibly lower with respect to bulk methods. Since energies involved in the process here are reduced, higher levels of homogeneity with reasonably fast growths (some tens of µm/h) are ensured, and also the inclusion of dopant atoms (if any) in the grower becomes viable.

When the fabrication of HgCdTe thin films is needed, one of the most used techniques is the MBE. Even if growth rate is rather low (few units of µm/h) the fluctuations control can be highly precise, thanks to the value of both precursors sticking coefficients and reaction temperature (less than 470 K): in fact, Δx typically covers the range from 0.01 down to 0.0006. In this technique typical sources are CdTe, Te_2 and Hg, while substrates can be CdZnTe, Si or GaAs (with a ZnTe buffer layer).

The last important epitaxial process is the VPE. Developed in 1968, it originally consisted in the ISOthermal Vapour Phase Epitaxy (ISOVPE), a quasi-equilibrium technique in which a precursor $\mathrm{HgTe}^{(\mathrm{vap})}$ was transported at $T = 670$-870 K from a source and then condensed on the substrate. However, at present day the most used method is the Metal-Organic Chemical Vapor Deposition (MOCVD) where the substrate (among all, CdZnTe, sapphire or GaAs) is exposed at about 620 K to some precursors carried by H_2 like $\mathrm{Hg}^{(\mathrm{met})}$, dimethylcadmium (DMCd) plus diisopropyltelluride (DIPTe) or diethyltelluride (DETe). Reactive process takes place in non-equilibrium conditions, which provide a dynamical control of all growth factors, dopant fluxes as well as the preparation and passivation of surfaces/interfaces. Unfortunately the high controllability of composition ($\Delta x_{\mathrm{MOCVD}} \sim \Delta x_{\mathrm{MBE}}$) comes at the expense of growth velocity: only 2-10 µm/h.

2.2.2 Energy dispersion

The compositional fluctuations we previously seen are responsible for the absence of a true potential periodicity in $\mathrm{Hg}_{1-x}\mathrm{Cd}_x\mathrm{Te}$ alloys, an important requisite to apply the Bloch theorem and, in turn, to derive an approximated form of the MCT band

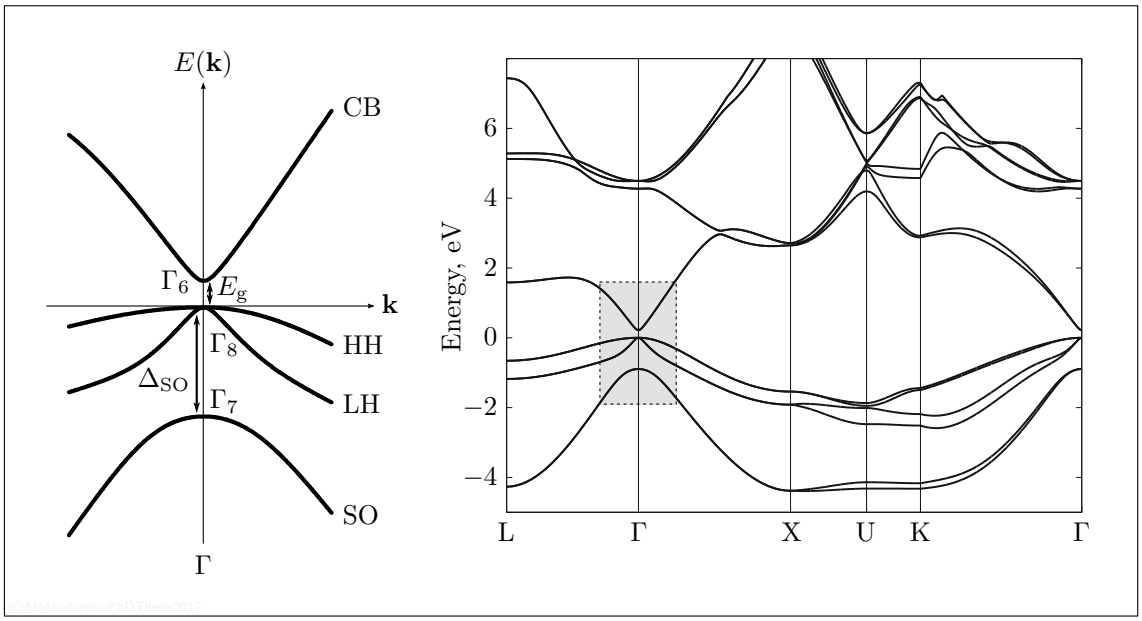

FIGURE 2.6. Left: MCT band structure in the FBZ around Γ, where we highlighted the bottom of CB, i.e. Γ_6, the top of both LH and HH valence bands and the top of the SO band, respectively Γ_8 and Γ_7. Finally, also some characteristic gaps as $E_g = \Gamma_6 - \Gamma_8$ and $\Delta_{SO} = \Gamma_8 - \Gamma_7$ are indicated. Right: Full-band $Hg_{0.7}Cd_{0.3}Te$ electronic structure calculated via Monte Carlo simulation (from Ref. [29]). The grey rectangle corresponds to the region of bands magnified on the left side.

structure. In absence of that, the VCA framework comes in our help, making use of a virtual-potential model interpolating throughout all compositional states, with the form

$$U_{VCA} = x\,U_{Cd} + (1 - x)\,U_{Hg} \tag{2.6}$$

and weighted on the potentials at Cd and Hg sites (U_{Cd} and U_{Hg} respectively), where x is the usual Cd mole fraction.

However, it has been pointed out that VCA fails, especially far from the characteristic points of the reciprocal lattice [26]. In those regions of the **k**-space an alternative approach may consist in using the Coherent Potential Approximation (CPA), which is as effective as high-computational-cost.

For the reasons mentioned above, one of the most proficuous strategies to compute the MCT band structure consists in the $\mathbf{k} \cdot \mathbf{p}$ method with perturbation corrections (see 1.1). Before analyzing the results of this formalism, we have to notice that in MCT the top of valence band is essentially composed by p-symmetry states (mostly p-Te ones), which also include a spin-orbit term. Whereas, the s-symmetry at the bottom of CB – which is twofold degenerated – is mainly due to s-Cd and s-Hg states. Moreover, since the atomic numbers of Hg, Cd and Te are rather high, the description of orbitals and the equation of motion should account also for relativistic

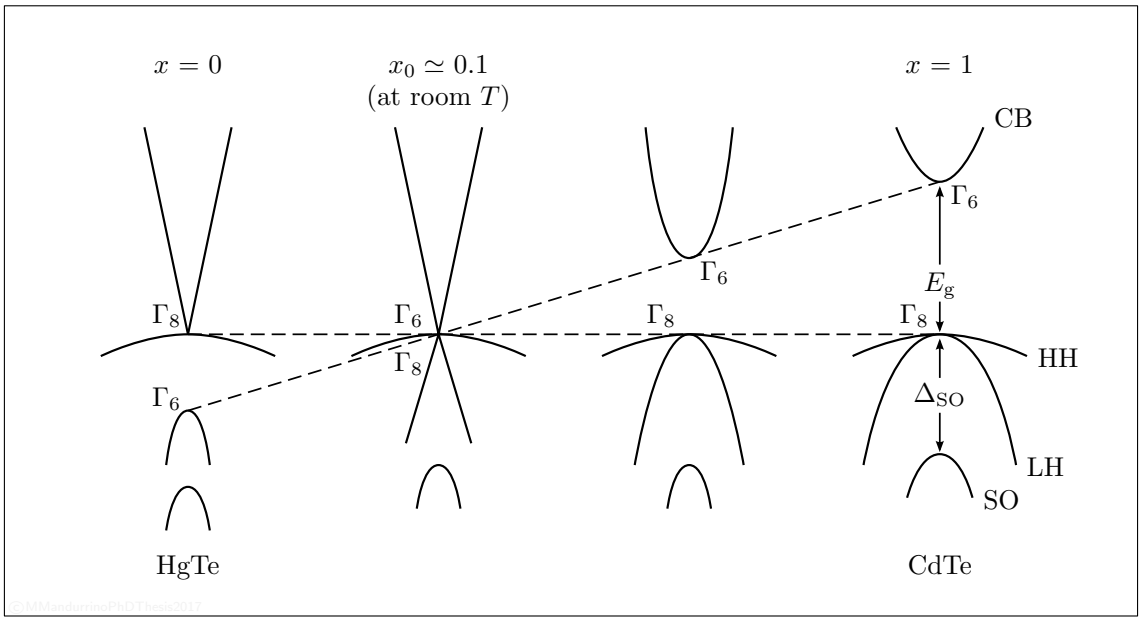

FIGURE 2.7. Band structure and energy gap E_g modulation in MCT as a function of cadmium mole fraction x. Going from left to right we start from a semimetal HgTe configuration, then the bands with Γ_6 and Γ_8 points reverse their sign and position at $x = x_0$ and a gap increasingly opens up to the CdTe configuration, which is purely semiconducting. Note that at points Γ_6 and Γ_8 the dispersion law $E(\mathbf{k})$ is quite linear around $x = x_0$. Adapted and re-edited from Ref. [26].

corrections. The resulting MCT Hamiltonian has the typical form

$$H = \overbrace{\left(-\frac{\hbar^2}{2m}\Delta_{\mathbf{r}} + U(\mathbf{r})\right)}^{H_0} + \overbrace{\left(\frac{mc^2 + U(\mathbf{r}) - E}{4m^2c^2}\hbar^2\Delta\right)}^{H_{\text{mv}}} + \overbrace{\left(-\frac{\hbar^2}{4m^2c^2}\nabla U(\mathbf{r})\cdot\nabla\right)}^{H_{\text{D}}} + \underbrace{\left(\frac{\hbar^2}{4m^2c^2}(\nabla U(\mathbf{r})\times\mathbf{p})\cdot\overline{\sigma}\right)}_{H_{\text{so}}} \tag{2.7}$$

where H_0 is the non relativistic contribution, H_{mv} and H_{D} describe the mass-velocity and Darwin interactions, respectively, H_{so} models the spin-orbit effect (which includes also a $\mathbf{k}\cdot\mathbf{p}$ perturbation term) and $\overline{\sigma}$ is the spin operator. Solving the secular equation (1.12) with the Hamiltonian reported in Eq. (2.7) means finding the eigenvalues of the problem and then predict the trend of the energetic bands (i.e. compute the value of the energy E for each wavevector $\mathbf{k}$ of the reciprocal space in the first Brillouin zone).

Now it is possible to compute, in a first-order perturbation approximation, the dispersion equations from the Hamiltonian. The four main expressions, valid around the FBZ center, are hereinafter reported [30]:

$$E_{\text{C}}(x,\mathbf{k}) = \frac{E_g(x)}{2}\left(1 + \sqrt{1 + \frac{8\mathbf{k}^2Q^2}{3E_g^2(x)}}\right) + \frac{\hbar^2\mathbf{k}^2}{2m_0} \tag{2.8}$$

$$E_{\mathrm{HH}}(x,\mathbf{k}) = -\frac{\hbar^2\mathbf{k}^2}{2m^*_{\mathrm{HH}}(x)} \tag{2.9}$$

$$E_{\mathrm{LH}}(x,\mathbf{k}) = -\frac{E_{\mathrm{g}}(x)}{2}\left(1-\sqrt{1+\frac{8\mathbf{k}^2Q^2}{3E_{\mathrm{g}}^2(x)}}\right)+\frac{\hbar^2\mathbf{k}^2}{2m_0} \tag{2.10}$$

$$E_{\mathrm{SO}}(x,\mathbf{k}) = -\Delta_{\mathrm{SO}}(x)+\frac{\hbar^2\mathbf{k}^2}{2m_0}-\frac{\mathbf{k}^2Q^2}{3(E_{\mathrm{g}}^2(x)+\Delta_{\mathrm{SO}}(x))} \tag{2.11}$$

where Q is the momentum matrix element, $\Delta_{\mathrm{SO}}(x) = E_{\mathrm{HH}}(\Gamma,x) - E_{\mathrm{SO}}(\Gamma,x)$ and m_0 is the electron rest mass. Typical values and trends of these parameters are $Q = 8.4\cdot10^{-8}$ eV and $\Delta_{\mathrm{SO}}(x) = 1.08 - 0.12x$ eV.

First, one should notice that both E_{g} and Δ_{SO} are x-dependent, as graphically reported in Figure 2.7. Moreover, observing the curvature of CB and LH for small $\mathbf{k}$, it can be inferred that $m^*_{\mathrm{C}} \simeq m^*_{\mathrm{LH}}$. Getting all band effective masses from Eqs. (2.8)–(2.11), one comes at the same conclusion. In addition, it is worth noting the high heavy-hole effective mass that, generally, is assumed $m^*_{\mathrm{HH}} \simeq 0.5\ m_0$

Being m^*_{C} (as well as m^*_{HH}) a crucial parameter in terms of GR processes, it deserves a special attention. In particular, two main regimes may be distinguished according to the alloy composition:

$$E_{\mathrm{C}}(\mathbf{k}) \simeq \begin{cases} \mathbf{k}^2 \\ \dfrac{E_{\mathrm{g}}}{2}+\dfrac{\hbar^2\mathbf{k}^2}{2m_0}+\sqrt{\dfrac{2}{3}}Q\,\mathbf{k} \end{cases} \quad\Rightarrow\quad m^*_{\mathrm{C}} \simeq \begin{cases} \dfrac{\hbar^2E_{\mathrm{g}}}{2Q^2}\,, & \text{for } E_{\mathrm{g}}^2 \gg \dfrac{8}{3}\mathbf{k}^2Q^2 \\ m_0\,, & \text{for } E_{\mathrm{g}}^2 \ll \dfrac{8}{3}\mathbf{k}^2Q^2 \end{cases}. \tag{2.12}$$

Equations above mean that for wider gaps (i.e. for high cadmium mole fraction x) CB is approximately parabolic around Γ, and the effective electron mass is linear with E_{g} (that, in turn, implies that m^*_{C} proportionally increases with x). On the other hand, if the gap is narrower than a certain threshold the dispersion law for CB becomes linear with respect to $\mathbf{k}$ and the effective mass is close to the electron rest mass m_0. As occurs for graphene, when CB and VB are linear, symmetric and touch themselves only by one point in correspondence of Γ the electron mass becomes the highest possible. Unfortunately, in the case of MCT this symmetry is tainted by the first valence band (for a comparison, see Figure 2.7 at $x \simeq x_0$), then no relativistic transport nor ultra-high mobilities are admitted here.

By the experimental point of view, there are several empirical models describing the trend of $m^*_{\mathrm{C}}(x)$. A first noticeable result comes from M. H. Weiler [27] and consists in

$$\frac{m_0}{m^*_{\mathrm{C}}(x)} = -0.6+\frac{E_{\mathrm{p}}}{3}\left(\frac{2}{E_{\mathrm{g}}(x)}+\frac{1}{E_{\mathrm{g}}(x)+\Delta_{\mathrm{SO}}}\right), \tag{2.13}$$

where $E_{\mathrm{p}} = 19$ eV and Δ_{SO} is assumed to be equal to 1 eV. A further implementation of Eq. (2.13) came recently thanks to J. Chu and A. Sher [26]. It reads:

$$\frac{m_0}{m^*_{\mathrm{C}}(x)} = \frac{1}{0.05966}\,\frac{E_{\mathrm{g}}(x)+0.667}{E_{\mathrm{g}}(x)(E_{\mathrm{g}}(x)+1)}. \tag{2.14}$$

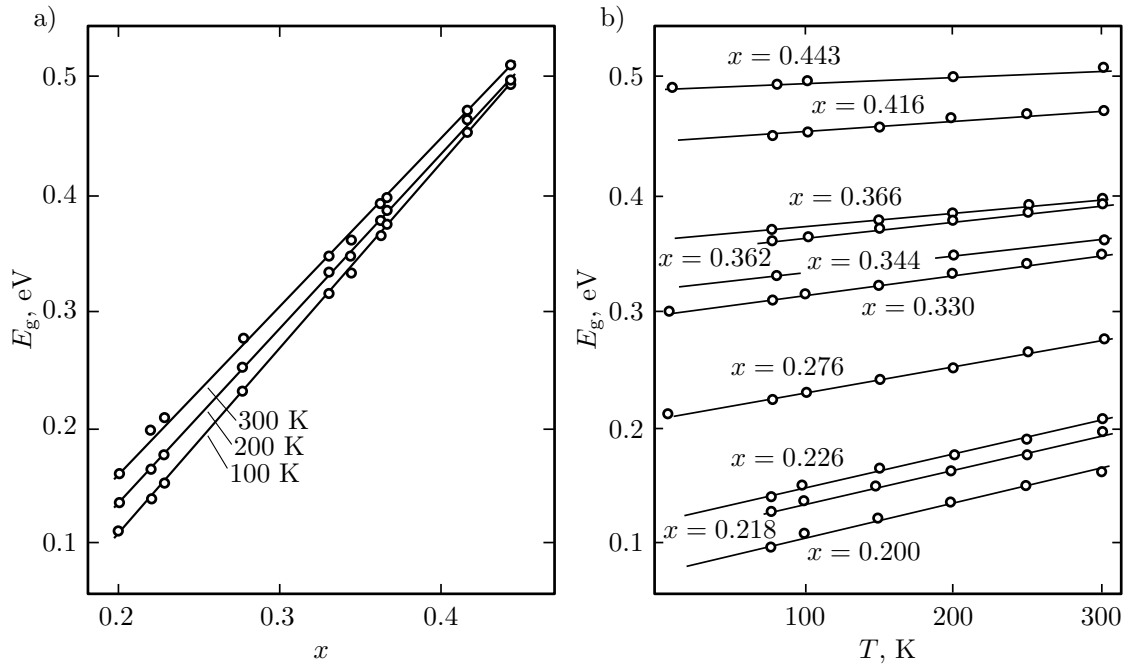

FIGURE 2.8. Experimental trend of energy gap in MCT, where dots denote measurements and lines are fitting curves. (a) $E_g(x)$ plotted at various temperatures. (b) $E_g(T)$ for different mole fractions. Data published in Ref. [26].

Finally, what we need to fully appreciate the meaning of Eqs. (2.13) and (2.14) is the behavior of the energy gap as a function of x. As already said, and according to Figure 2.7, HgTe is a semimetal alloy with negative gap (few meV) due to the fact that Γ_6 lies at the uppermost state of a negative-concavity band, whereas the higher point Γ_8 represents the lowest bound of a positive-concavity band. This determines $\Delta_{SO} = \Gamma_6 - \Gamma_8 < 0$. While x increases, the modulus of Δ_{SO} progressively decreases until, for $x = x_0$, Γ_6 and Γ_8 change the sign and their respective bands invert the curvature. In this configuration, which may occur for $x_0 \simeq 0.15$ at $T = 77$ K or $x_0 \simeq 0.1$ at $T = 300$ K, these two points are in contact at $E(\mathbf{k} = 0) = 0$. From now on, a gap opens and any further increase of x implies an increase of E_g too. This is the $Hg_{1-x}Cd_xTe$ semiconducting regime. With a non-linear dependency, this trend leads to $x = 1$, where the unit cell is only CdTe-based and the gap is about 1.5-1.6 eV, depending on T.

As a matter of fact, not only cadmium fraction can affect the gap amplitude, but also temperature does so. In other words, both $m_C^* = m_C^*(x, T)$ and $E_g = E_g(x, T)$ are functions of temperature and compositional stoichiometry. Several empirical models have been developed to describe these two functions and in particular the latter, which is the one through which the former subsists.

Calculating interband recombination rates of photoluminescence processes through carrier lifetime τ measurements and determining another physical parameter of MCT which depends on x and T, i.e. the absorption coefficient α (in cm^{-1}), one can

write [31]:

$$\alpha(x,T) \simeq \frac{\sqrt{2c}}{\tau}\sqrt{1-\frac{\lambda}{\lambda_{\text{cut-off}}(E_{\text{g}}(x,T))}}\left(\sqrt{\frac{m_{\text{C}}^{*}(E_{\text{g}}(x,T))\,\lambda}{m_0\,\hbar}}\right)^{3}, \tag{2.15}$$

where c is the speed of light in vacuum, λ the wavelength of the employed electromagnetic radiation and

$$\lambda_{\text{cut-off}}(E_{\text{g}}(x,T)) \simeq \frac{1.24}{E_{\text{g}}(x,T)}\,. \tag{2.16}$$

From Eq. (2.15) some models can be experimentally derived. Among the most common, one is for sure the one developed in 1982 by G. L. Hansen, J. L. Schmit and T. N. Casselman [32], a mixed polynomial expression of third-order in x and first-order in T:

$$E_{\text{g}}(x,T) = -0.302 + 1.93x - 0.81x^2 + 0.832x^3 + 5.35\cdot 10^{-4}\,(1-2x)\,f(T)\,, \tag{2.17}$$

where $f(T) = T$. Few years after, in 1990, D. G. Seiler and coworkers [33] published a slightly different model, via magneto-optical experiments at low temperature, in which the thermal factor is replaced by the expression

$$f(T)|_{x=\overline{x}} = \frac{-1822 + T^3}{255.2 + T^2}\,, \tag{2.18}$$

valid only with mole fraction $\overline{x} \in [0.2, 0.3]$.

2.2.3 Carrier transport

Photogeneration critically depends on thermal, optical and transport properties of the material under investigation. In particular, density of states (DOS) affects carrier concentrations that, finally, determines the transition rates. This is true not only for what concerns optical generation but also for the remaining spectrum of competitive GR processes like tunneling, SRH or Auger. Nevertheless, if densities essentially derive from alloy composition or thermodynamical conditions, what is intrinsically due to microscopical properties are carrier mobilities. Being MCT electron and hole effective masses (and mobilities) x-dependent, we saw that they also tune together with energy gap. This is particularly evident for very narrow gaps ($x \simeq x_0$), where bands become linear and mobilities increase drastically. Moreover, as we already observed, near the Γ symmetry point both m_{C}^{*} and m_{LH}^{*} are in general pretty low while m_{HH}^{*} is rather high, yielding electron mobility two or three orders of magnitude greater than heavy-holes: typically for MCT photodetectors it is assumed [30]

$$\mu_{\text{h}}(x,T) \simeq \frac{\mu_{\text{e}}(x,T)}{100}\,, \tag{2.19}$$

where a unique notation for all holes mobilities has been used (by virtue of the fact that GR processes are mainly governed by electrons and heavy-holes). Furthermore,

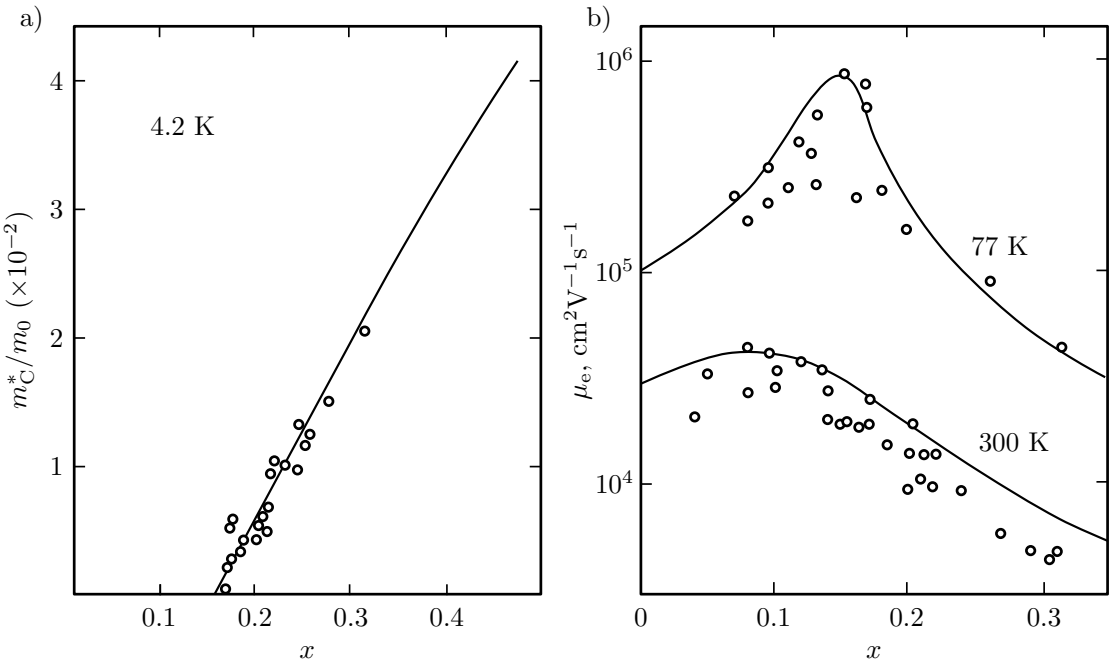

FIGURE 2.9. (a) Measured $m^*_C(x)/m_0$ ratio plotted as a function of the cadmium fraction, with its related empirical model (line) represented by Eq. (2.13). (b) Experimental trend of MCT electron mobility with respect to x for two different temperatures, accompanied with theoretical curves described by Eq. (2.20). Data published in Refs. [7] and [26], respectively.

μ_h at room temperature is commonly assumed to be between 40 and 80 cm^{2}V^{-1}s^{-1}, while at 77 K it is almost one order of magnitude higher.

From a microscopic point of view, MCT mobilities are limited by scattering mechanisms: up to 50 K μ_{hh}, as well as μ_e, is dominated by scattering with dislocations whereas for intermediate T with ionized impurities. At higher temperature, and above 200 K, polar phonon scattering (especially from LO phonons) governs while μ_{lh}, instead, is mainly driven by acoustic phonon scattering.

Concerning electrons, several expressions are used to describe the mobility trend according to thermodynamic and compositional conditions. One of the most used in the range $0.2 \leq x \leq 0.6$ and for $T > 50$ K – also reported in Ref. [7] – is

$$\mu_e(x,T) \simeq \frac{9 \cdot 10^8 \left(\frac{0.2}{x}\right)^{7.5}}{T^{\left[2\cdot\left(\frac{0.2}{x}\right)^{0.6}\right]}} \tag{2.20}$$

which at room temperature is approximated by the polynomial expression

$$\mu_e(x) \simeq \frac{1}{8.754 \cdot 10^{-4} x - 1.044 \cdot 10^{-4}}\,. \tag{2.21}$$

By the experimental standpoint, below 200 K and for $0.285 \leq x \leq 0.290$, μ_e has been measured in the range 2 to $4 \cdot 10^4$ cm^{2}V^{-1}s^{-1} while at 77 K and $x = 0.2$ it rises to $2 \cdot 10^5$ cm^{2}V^{-1}s^{-1}. Also minority carrier mobility can be crucial for the transport.

In p-type MCT samples μ_e does not differ too much from the value found in n-type ones, provided the acceptor concentration is $N_A < 10^{15}$ cm^{-3}, otherwise electron mobility drastically goes down.

Regarding the intrinsic concentration $n_i(x,T)$, another important feature to define the state of a semiconductor, we report one of the most exploited approximations derived from $\mathbf{k}\cdot\mathbf{p}$ theory:

$$n_i(x,T) \simeq \left(A + B\,x + C\,T + D\,x\,T + F\,x^2 + G\,T^2\right)\cdot 10^{14}\, E_g^{\frac{3}{4}}\, T^{\frac{3}{2}}\, e^{-\frac{E_g}{2k_B T}}\,, \quad (2.22)$$

where $A = 5.242566$, $B = -3.5729$, $C = -4.74019\cdot 10^{-4}$, $D = 1.25942\cdot 10^{-2}$, $F = 5.77046$ and $G = -4.24123\cdot 10^{-6}$. This formula was determined by J. L. Lowney [34] by obtaining $E_g(x,T)$ through a 3-bands model with non-parabolicity effects and thanks to experimental magnetoabsorption data. Other alternative approaches (which essentially differ in the way of expressing $E_g(x,T)$) can also be found in Refs. [32, 35, 36].

To conclude this paragraph we mention some general properties about lifetimes that will be developed more deeply in Section 2.3. Doping, obviously, is one of the main issues affecting lifetimes: typically, for impurity densities in MCT lower than $5\cdot 10^5$ cm^{-3}, minority carriers lifetime spreads very widely in the range 10^{-10} to 10^{-5} s. Differences between n- and p-doped samples also occur. Depending on the growth technique, values from 0.4 to 8 µs are reported in the literature for n-HgCdTe at 80 K (see, e.g., Ref. [7]) when carrier concentrations are below 10^{15} cm^{-3} and for $x \simeq 0.2$. In p-doped materials, instead, we may find lifetimes from a broader spectrum, mainly governed by the temperature which allows a maximum for τ at intermediate T. The action of temperature results even more evident in n-type samples with the two-fold dominance of Auger (high-T) and SRH (low-T). It is straightforward that the highest growth quality the lowest impurity presence and, in turn, the highest lifetime experienced by carriers. Indeed, the highest τ have been observed in undoped samples grown via LPE and MOCVD.

The last important feature in terms of lifetimes is the nature of dopants: in fact it has been observed a τ in Au- and As-doped HgCdTe 2-5 times larger than in samples with the same concentration of native dopants. This is probably due to the activation process (annealing) that minimizes internal energies within the lattice.

2.2.4 Photon absorption

We have already spent some general words about the absorption coefficient α and its dependence on the cadmium mole fraction (see Eq. (2.15) and Section 2.2.2). Here we want to discuss the behavior of α as a function of the photon energy E_{ph}.

According to the theory of bands, since both CB and LH are quasi parabolic around Γ, we should observe $\alpha(E_{ph}) \sim \sqrt{E_{ph}}$ in correspondence of the absorption edge due to a square-root-like attenuation of the photogeneration transitions. Except for very high quality samples, this effect is almost never found. On the contrary, larger tails at $h\nu < E_g$ (see panel (a) of Figure 2.10) are typically observed. This is due in part by compositional fluctuations or impurities/defects, which respectively

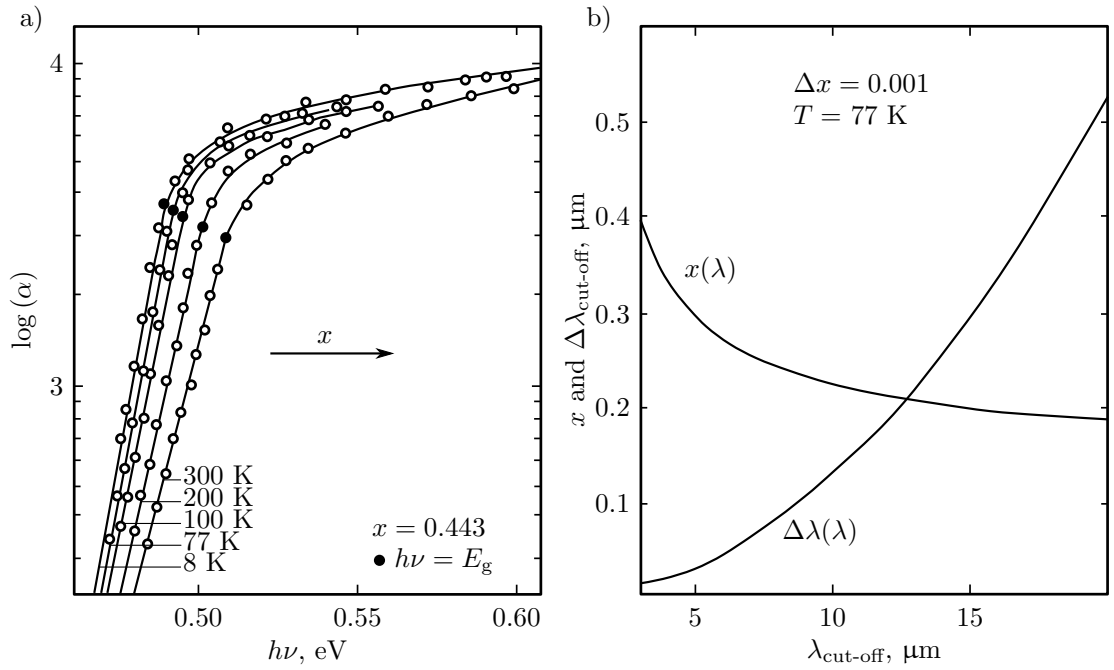

FIGURE 2.10. (a) Absorption coefficient in $Hg_{0.557}Cd_{0.443}Te$ as a function of photon energy and at different temperatures (data published in [26]). Below the knee at the edge (i.e. for $h\nu < E_g$) the Urbach-like trend is observable, as well as the Kane-like above it. Also the convergence of all curves to the same point, called "focal point", is pretty evident. This point is $x-$sensitive and in this particular case it has coordinates $(0.441; 10)$. Re-plotting these curves at fixed temperature but for increasing mole fraction we should note an horizontal shift towards the direction indicated by the arrow. (b) Experimental variation of $\Delta\lambda_{\text{cut-off}}$ and cadmium fraction as a function of the cut-off wavelength. Data published in Ref. [7].

determine local variations of E_g or the presence of intergap states, or it is imputable to inter-band transitions.

The first scientist who, in 1953, described and modeled this kind of behavior in solids was F. Urbach (see Ref. [37]). In its work he reported and described in great detail the exponential trend of the absorption coefficient of AgBr salt with respect to the photon frequency. Soon after, in 1957, W. Martienssen [38,39] extended these observations to some alkali-halide crystals, and in particular to KBr salt, by applying the Urbach model

$$\alpha_U(\nu) = \alpha_0 \exp\left(\frac{(h\nu - E_0)\,\sigma}{k_B T}\right), \tag{2.23}$$

where σ, α_0 and E_0 are fitting parameters. As demostrated in 1966 by D. T. F. Marple [40] and C. A. Hougen [41], Eq. (2.23) is still valid for II-VI materials like MCT, provided that

$$E_{0,\text{MCT}} = -0.355 + 1.77\,x \qquad [\text{eV}] \tag{2.24}$$

$$\ln \alpha_{0,\text{MCT}} = -18.5 + 45.68\,x \tag{2.25}$$

$$\sigma_{\text{MCT}} = 3.267 \cdot 10^4\,(1 + x)\,. \qquad [\text{K/eV}] \tag{2.26}$$

As represented in panel (a) of Figure 2.10, $\alpha_g \equiv \alpha(E_g)$ is x- and T-dependent (see black dots). Thus,

$$\frac{\sigma}{k_B T} = \frac{\ln \alpha_g - \ln \alpha_0}{E_g - E_0} \tag{2.27}$$

is verified. From this equation and from experimental data fit for HgCdTe samples one can also write the following Urbach-like version of the absorption trend at the edge:

$$\begin{cases} \alpha_g(x) = 500 + 5.6\,x & 0.17 \leq x \leq 0.443,\ T = 300\ \text{K} \\ \alpha_g(x) = -65 + 1.88\,T + (8694 - 10.31\,T)\,x & \text{otherwise} \end{cases} \tag{2.28}$$

valid for $E \lesssim E_g$, i.e. when band-to-band transitions start to dominate. Above this regime the Kane rule

$$\alpha_K(\nu) = A\sqrt{\frac{1+x}{B+T}}\sqrt{(h\nu - E_g)}\,, \tag{2.29}$$

is more suitable, where $A_{MCT} = 2.109 \cdot 10^5$ and $B_{MCT} = 81.9$.

As we already mentioned, the absorption strength depends on bands curvature around $\mathbf{k} = 0$ (and on m^*, obviously) and it decreases as the gap narrows. Since non-parabolicity effects can induce variations in the absorption mechanism via fluctuations of bands shape and energy gap, a new model has been recently developed (see Ref. [7]):

$$\alpha(\nu, x, T) \simeq \frac{\sqrt{(h\nu - E_g(x,T) + \gamma)^2 - \gamma^2}\,(h\nu - E_g(x,T) + \gamma)}{h\nu} \tag{2.30}$$

where γ is the hyperbolicity parameter. Eq. (2.30) has been explicitly written as a function of x and T through the energy gap because it is well known that the non-parabolicity degree increases as both x and T decrease.

One of the last parameters playing a role in the photon absorption is the refractive index n. In MCT, usually, a Kramers-Kronig description is used. However, since n is heavily T-dependent, in particular for $0.276 \leq x \leq 0.540$ and $4.2 \leq T \leq 300$ K (or, also, for $0.205 \leq x \leq 1$ and at room temperature), a specific phenomenological model [7, 22, 41, 42] is used for convenience:

$$n^2(\lambda, x, T) = A + \frac{B}{1 - \left(\frac{C}{\lambda}\right)^2} + D\,\lambda^2 \tag{2.31}$$

in which A, B, C and D are x- and T-dependent fitting parameters.

Regarding the static and high-frequency dielectric constants, ϵ_0 and ϵ_∞ respectively, it can be said that while the influence of temperature is very weak [7, 43], the cadmium mole fraction acts as follows:

$$\epsilon_\infty(x) = 15.2 - 15.6\,x + 8.2\,x^2 \tag{2.32}$$

and

$$\epsilon_0(x) = 20.5 - 15.6\,x + 5.7\,x^2\,. \tag{2.33}$$

If now we put together Eqs. (2.16) and (2.17), after some passage we obtain

$$\lambda_{\text{cut-off}} = \frac{1}{-0.244 + 1.566\,x - 0.65\,x^2 + 0.671\,x^3 + 4.31 \cdot 10^{-4}\,(1-2x)\,T}\,, \tag{2.34}$$

and also

$$\frac{\Delta\lambda_{\text{cut-off}}}{\Delta x} = \lambda^2_{\text{cut-off}}\,(1.566 - 1.3\,x + 2.013\,x^2 - 8.62 \cdot 10^{-4}\,T)\,. \tag{2.35}$$

In Figure 2.10, panel (b), Eqs. (2.34) and (2.35) are reported for a fixed compositional average fluctuation $\Delta x = 0.001$ and at $T = 77$ K.

2.2.5 Macroscopic properties

In this last introductory paragraph we will briefly discuss some of the variational features affecting the local composition of MCT alloys, property which is at the basis of all the mole fraction dependencies so far analyzed.

First, a special mention is reserved for substitutional impurity dopants. A common choice for acceptors in Cd or Hg sites is represented by elements from I-group (both IA- and IB-subgroups) like Ag, Cu or Au while, for donors, elements from IIIB-subgroup are typically implanted, as in the case of Ga, Al and In. To substitute the Te, elements from VIB- and VIIB-subgroup are often used as donors whereas elements from VB-subgroup are employed as acceptors. Furthermore, As and In are considered some of the most effective dopands for p- and n-type MCT alloys, respectively, while both point defects (like vacancies or interstitials) and lattice defects (dislocations and grain boundaries) have a twofold role, behaving as acceptors and sometimes as donors.

By the standpoint of lattice sites, there are several donors (B, Al, Ga, In, Si, O, Cl, Br, I) acting either as impurity and as standard activated dopants, and the same occurs for some acceptors (Li, Cu, Ag, Au, As, Sb, P). But, in general, the most difficult challenge in MCT growth is to reduce defects and impurities at all, being among the major sources of performance degradation. In fact, acting as trap-assisted or SRH recombination centers, they determine carrier lifetime and mobility decrease, current leakage and $1/f$-noise, allowing an increased dark current or, even, the formation of dead pixels. The role of some of these elements will be re-analyzed in Chapter 5 by discussing the experimental data from real devices and numerical modeling. Moreover, for dislocation levels above $5 \cdot 10^5$ cm^{-2}, the typical dimension of IR detectors implies a degradation threshold of about 12 dislocations per photodiode [30]. This can be avoided only controlling the presence of damages, contaminations and other lattice imperfections like stacking faults, mismatches, grain boundaries, precipitates, striations, pinholes, interstitials, terracing and inclusions during the growth process.

Finally, a special category of impurities consists in native defects which, in undoped HgCdTe, are typically represented by vacancies. Metals, instead, are considered native acceptors (this kind of donors is almost negligible). The growth

parameter determining the presence of native defects is essentially the mercury partial pressure, since hole density is inversely proportional to P_{Hg} [7]. To mitigate this effect the annealing procedure is mandatory, provided it is conducted at a precise temperature otherwise further issues, like p-to-n conversion, surface layers formation or strain effects, can appear in the final sample.

2.3 Detectors Performance and Figures of Merit

In the present paragraph we will conclude the introduction about IR detection by first describing the basic solid-state detection principles and then by listing some important detector features and limitations, both in terms of intrinsic factors and under the system conditions standpoint. Finally, some engineering and technological solutions to optimize the state-of-the-art devices will be given.

2.3.1 Light detection and photocurrent

In Section 2.1 we briefly analyzed the intrinsic and extrinsic absorption mechanisms. At the basis of these processes there is a generation of carriers due to the energy released by an absorbed photon. Moreover, in Section 2.2.4 we also discussed the fact that, even for energies $h\nu \lesssim Eg$, the absorption coefficient can assume non-null values (the so-called "tail" of the absorption edge).

Figure 2.11 represents two competitive optical transitions between CB and VB occurring in semiconductors. In panel (a) there is the process of our interest, the photoexcitation, i.e. the generation of an electron/hole couple stimulated by the absorption of a photon. This mechanism is very simple: the energy released (it must reasonably be not too much less than the energy gap) can promote a valence electron into the conduction band, by generating a heavy-hole at the initial state. The process can be direct, as in our picture, or not, depending on the presence of impurity states within the gap which may assist the transition.

Panel (b) of the same figure shows another process, inverse of the previous one: the radiative recombination. By its nature, it is competitive with the photogeneration. Nevertheless several authors (see, for instance, Refs. [12, 44]) point out that in HgCdTe photon-recycling may destructively add up to the radiative recombination producing an overall null effect. It is obvious that both of these mechanisms can not occur for each incoming/escaping photon. In other words, due to the presence of defects, material re-absorption, electro-optical leakages or photon-recycling, their efficiency can never reach the ideal limit of 100%.

Now consider the simplest prototype of an IR photodetector: a reverse-biased p-n junction. The positive voltage applied at the n-contact contributes to the drift current generated by the space charge field $\mathcal{E}$ and contemporarily acts against the diffusion current which, instead, would tend to compensate the gradient of minority carriers across the junction and between the two doped regions. This leads to an increment of the space charge ad to a depletion region widening. In presence of light illuminating the p-region all intrinsic and extrinsic absorption phenomena

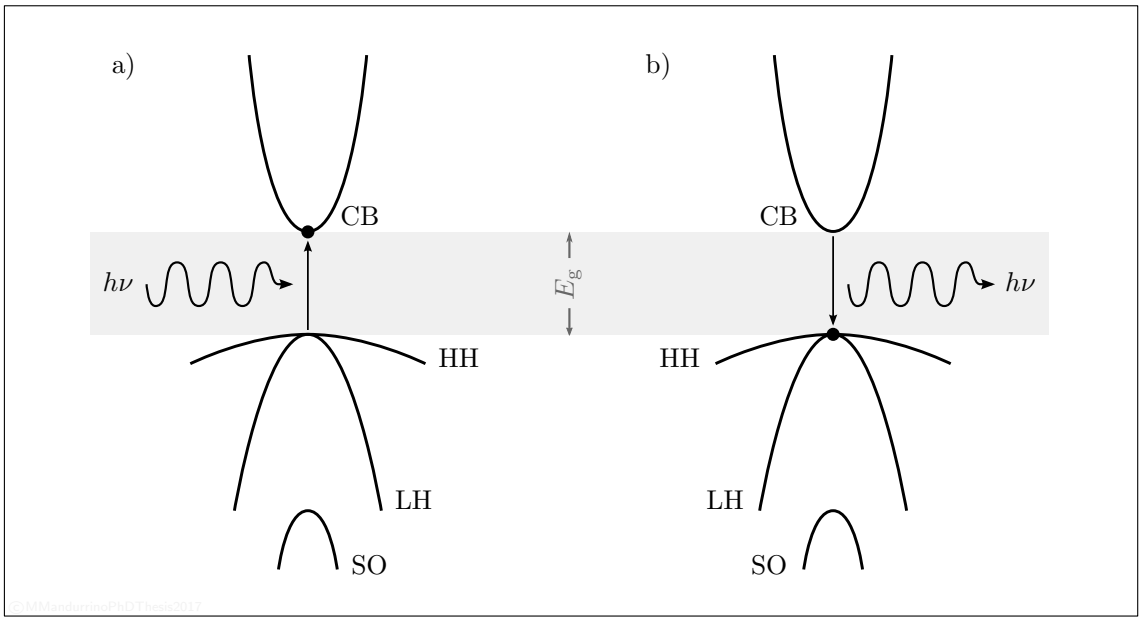

FIGURE 2.11. Semplificate scheme representing two of the most important optical transitions in direct-gap semiconductors, involving CB and HH band: (a) *photogeneration* and (b) *radiative recombination*. In each of these mechanisms, the absorption/emission of a photon with $h\nu \gtrsim E_g$ is expected due to the energy conservation principle.

taking place will contribute to the electron/hole pair photogeneration: minority electrons generated in the p-type region will diffuse within this region and here will recombine with holes, while the viceversa will occur for minority holes in the n-side. Pairs generated in the depletion region or not far than a diffusion length $L_{e,h}$ from it, instead, are separated and drifted into their respective sides. This produces a small detectable reverse current which depends on the absorption efficacy and on the quantum efficiency η of the generation process, which is the number of pairs generated per second N_0 over the number of incident photons per second N_{inc}. This photocurrent constitutes the signal of our p-n photodiode.

Let's analyze the first intrinsic limitation provided by the crystal: the absorption efficacy. As pointed out by Lambert and Beer in their famous law, the optical power flux of an incident electromagnetic radiation P_{inc} (in W) drops to 1/e at a *penetration depth* $z = 1/\alpha$ according to the expression

$$P_{abs}(z,\nu) = [P_{inc}\ (1 - r_\nu)]\ \left(1 - e^{-\alpha(\nu)z}\right) \tag{2.36}$$

where r_ν is the facet reflectivity of the material, α the absorption coefficient, N_{abs} is the number of absorbed photons per second, $P_{abs} = N_{abs}\, h\nu$ and $P_{inc} = N_{inc}\, h\nu$ represent, respectively, the absorbed and the incident power.

Once the radiation – or, better, a part of it – is absorbed, the generation mechanisms takes place. Then, as already mentioned, the photocurrent $I_{ph} = q\, N_0$ is generated with a particular internal quantum efficiency (IQE), that can be modeled as

$$\eta = \frac{N_0}{N_{inc}} = \frac{I_{ph}}{q\, N_{inc}} = \frac{I_{ph}\, h\nu}{q\, P_{inc}}\,, \tag{2.37}$$

which depends on the rate at which optical transitions occur and, in turn, on the presence of effects that could limit this mechanism.

Like the optical power, also the IQE experiences a z-dependence of the form

$$\eta(z,\nu) = (1-r_\nu)\left(1-\mathrm{e}^{-\alpha(\nu)z}\right) . \tag{2.38}$$

This equation suggests that, to maximize η, the reflectivity should be the lowest as possible so that most of the optical power can concentrate in the depletion region. Another way to improve the photogeneration is tu grow the junction reasonably close to the irradiated surface.

Panel (a) of Figure 2.12 shows a generic p-n photodiode and its bias circuit, while panel (b) schematically illustrates how $I(V)$ characteristics change with respect to the dark configuration (dashed) when the p-type region is illuminated. The usual operating regime of an MCT photodiode for IR detection is the plateau denoted by "II", that is at a low reverse bias (few Volts) where current saturates. Regions "I" (below the breakdown voltage V_{bd}) and "III" typically concern avalanche photodetectors (APDs) and solar cells operating regimes, respectively.

Panels (c) and (f), moreover, show the simplified structure of a p-n and p-i-n photodiode. (d) and (f) are the related band diagrams at a generic reverse bias and (e) and (h) their corresponding electric field profiles through the junction.

The ideal equation describing what shown in (b) without illumination is the well-known Shockley diode equation

$$I_{\mathrm{dark}} = I_0\left(\mathrm{e}^{\frac{q\,V_{\mathrm{bias}}}{k_{\mathrm{B}}T}}-1\right) , \tag{2.39}$$

where the *saturation current* I_0 is, in first approximation, a diffusion current of electrons towards the p-side and holes into the n-side [7]:

$$I_0 \simeq \frac{q\,D_{\mathrm{h}}\,p_n}{L_{\mathrm{h}}} + \frac{q\,D_{\mathrm{e}}\,n_p}{L_{\mathrm{e}}} , \tag{2.40}$$

with $D_{n,p} = \mu_{\mathrm{e,h}} k_{\mathrm{B}} T/q$ the minority carrier diffusion coefficients.

The width of the depletion region is bias-dependent and for a p-n junction can be written as

$$w_{\mathrm{dep}}\left(V_{\mathrm{bias}}\right) = \sqrt{\frac{2\epsilon\left[\frac{k_{\mathrm{B}}T}{q}\ln\left(\frac{N_{\mathrm{A}}N_{\mathrm{D}}}{n_i^2}\right)\pm V_{\mathrm{bias}}\right]}{qN_{\mathrm{A}}N_{\mathrm{D}}\left(N_{\mathrm{A}}+N_{\mathrm{D}}\right)}} , \tag{2.41}$$

where N_{A} and N_{D} are the acceptor/donor concentrations, ϵ is the dielectric constant, n_p and p_n are the minority electron/hole densities. Eq. (2.41) holds only within the so-called "short-diode approximation", i.e. for diffusion lengths $L_{\mathrm{e,h}}$ larger than the quasi-neutral region (where $\mathcal{E} \simeq 0$).

Neglecting other generation/recombination (GR) mechanisms, the ideal expression of the total current in light conditions consists in the sum of two terms, I_{dark} and $I_{\mathrm{ph}} = \eta\, q\, N_{\mathrm{inc}}$ (see Eq. (2.37)):

$$\begin{aligned} I_{\mathrm{tot}}\left(V_{\mathrm{bias}},T,N_{\mathrm{inc}}\right) &= I_{\mathrm{dark}} + I_{\mathrm{ph}} \\ &= I_0\left(\mathrm{e}^{\frac{q\,V_{\mathrm{bias}}}{k_{\mathrm{B}}T}}-1\right) + \eta\, q\, N_{\mathrm{inc}} . \end{aligned} \tag{2.42}$$

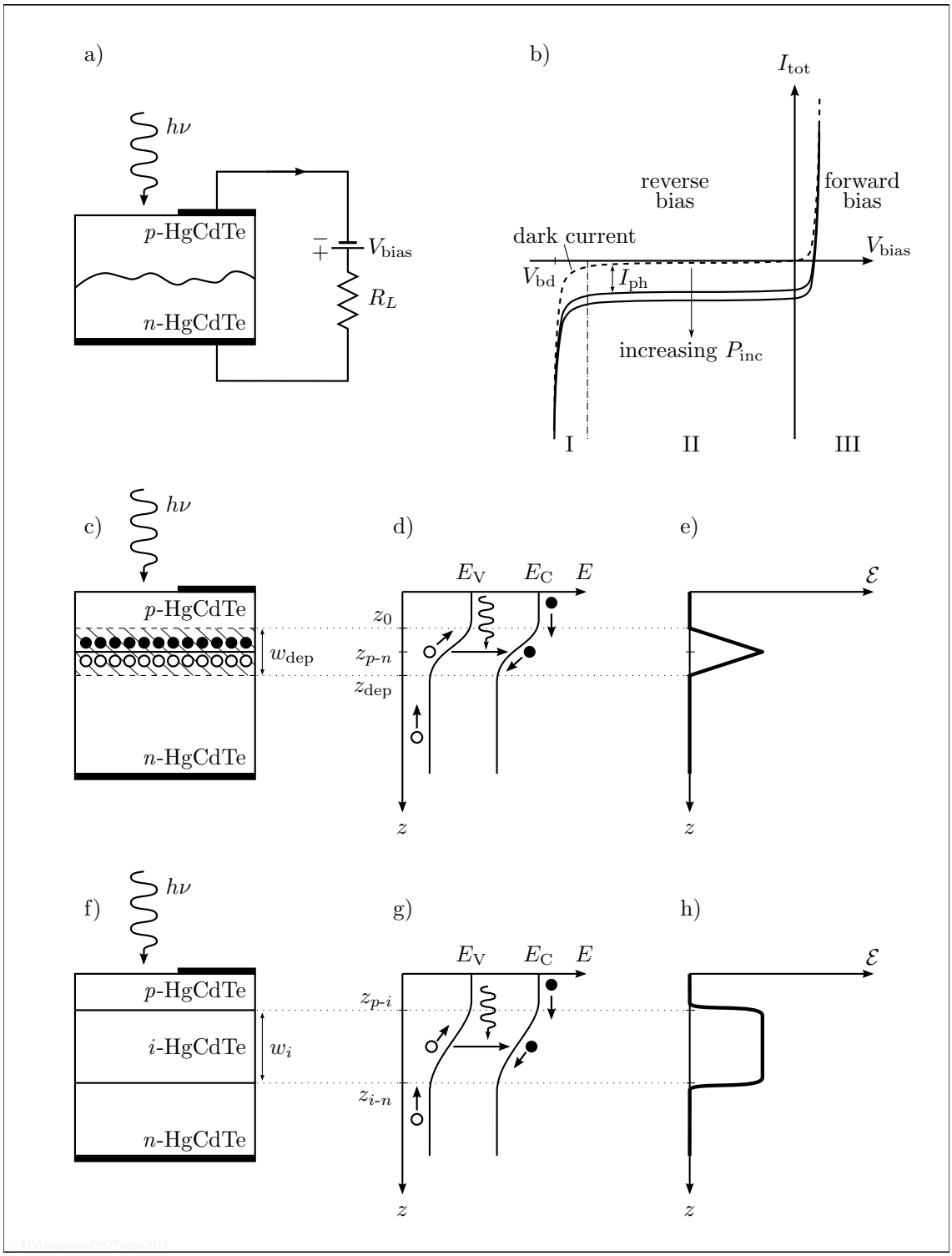

FIGURE 2.12. (a) Circuital scheme of a generic reverse-biased p-n photodiode, where V_{bias} and R_L are, respectively, the supply voltage and the load resistance. (b) Typical diode-like trend of $I(V)$ characteristics in reverse (I, II) and forward bias regime (III). It is well noticeable the current increase (in modulus) with respect to that one in dark conditions (dashed curve) when the number of incident photons increases. At high negative voltages (I) the diode breakdown occurs at V_{bd}. (c)-(f) Detailed representation of a p-n and p-i-n structure, respectively, with their (d)-(g) band diagrams and (e)-(h) electric field profiles along the junction. Striped area in (c) represents the depletion region and w_{dep} is its width. Adapted and re-edited from Ref. [45].

In presence of impurities or defects, as often occurs in real cases, other contributions should be added to the r.h.s. of Eq. (2.42). One of the first implications is that the zero-bias current will no longer be null as in the ideal case. Also the shape of $I_{\rm dark}(V)$ will change by reason of new transport or GR processes as trap-assisted tunneling (TAT) or band-to-band tunneling (BTBT), as we will extensively discuss in this work.

For instance, the presence of lattice defects at the external facets or in correspondence of interfaces between different layers produces typical energy levels in the bandgap that could generate the flow of so-called *surface leakage* currents. If we suppose a mid-gap state due to a surface defect acting as a SRH recombination center, its current contribution will be of the form [46]

$$I_{\rm surf} \sim q\, A\, n_i\, \sigma_{\rm trap}\, v_{\rm th}\, N_{\rm trap}\,, \tag{2.43}$$

i.e. proportional to the surface involved A and to the carrier thermal velocity $v_{\rm th}$ as well as to some defect-related properties like their cross sections $\sigma_{\rm trap}$ and density $N_{\rm trap}$. Furthermore, $I_{\rm surf}$ is also T-dependent, since $n_i(T) \sim \exp\left(-E_{\rm g}/2k_{\rm B}T\right)$. In order to prevent the formation of surface states devices are usually passivated through oxidation procedures.

Apart from surface states, also bulk regions are interested by trap-assisted transport phenomena, as we have already seen in Section 1.4. In particular, SRH generation within the depletion region can in principle heavily degrade the reverse current by increasing $I_{\rm dark}(V)$, especially at low temperature.

As has already been said, most of the carriers are photogenerated within the depletion region (plus one diffusion length around it). For this reason a certain compromise between $L_{\rm e,h}$ and the optimum absorption path $z_{\rm op} \simeq 1/\alpha(\nu)$ must be found in device designing, especially for HOT HgCdTe detectors where $z_{\rm op}$ is rather large. This can be obtained, for instance, by interposing an intrinsic layer between the two doped regions: see the *p*-*i*-*n* device in panel (f) of Figure 2.12. The intrinsic region can be either undoped or lightly doped, typically with $N_i \sim 10^{14}$-10^{15} cm^{-3} and its purpose is to take advantage of fully depletion along all the i-type layer width w_i, whatever $V_{\rm bias}$. In order to maximize the IQE, the junction should be buried at an optimum depth from the illuminated window and w_i is require to be as larger as possible. On the contrary, w_i must also be chosen such that the transit time of carriers determine a fast photoresponse. This can be achieved with a short intrinsic layer. So, again, a trade-off is mandatory.

Supposing to be in the ideal case of maximum efficiency $\eta = 1$, what constitutes the photocurrent components of a *p*-*i*-*n* photodiode is [46]:

$$\begin{aligned} I_{\rm ph}^{pin} &= I_{\rm drift}^{pin} + I_{\rm diff}^{pin} \\ &= (1-r_\nu) q N_{\rm inc} \left[\left(1 - {\rm e}^{-\alpha\, w_i}\right) + \left(\frac{\alpha\, L_{\rm e}}{1+\alpha\, L_{\rm e}} {\rm e}^{-\alpha\, w_i} + \frac{q\, D_{\rm e}\, n_p}{L_{\rm e}} \right) \right] \end{aligned} \tag{2.44}$$

Another design scheme for IR photodetectors passes through the separation of the absorbing layer from the active region. This is the case of n^+-p-π-p^+ multijunction APD detectors, where the multiplication layer is the p-type region while

the absorber is represented by the lightly doped intrinsic π-type material. In this way the high multiplication electric field occurs at the n^+-p junction, far from the illuminated window. Here the photocurrent in avalanche regime is mainly of the drift-type, and is

$$I_{\rm ph}^{\rm APD} = M\, I_{\rm ph}\,, \tag{2.45}$$

where $I_{\rm ph}$ is the low-bias current and M the *multiplication factor*, related to several physical details, and defined as

$$M^{-1} = 1 - \left|\frac{V_{\rm bias}}{V_{\rm bd}}\right|^m\,, \tag{2.46}$$

with m a material- and doping-dependent parameter of the APD device. It is worth noting that, besides this scheme there are many other structures designed for specific IR applications. Some few examples are: Schottky barrier photodiodes, Metal-Semiconductor-Metal (MSM) and Metal-Insulator-Semiconductor (MIS) devices or the promising, and relatively new, nBn structures [47–49].

2.3.2 Transport performance and limitations

What experiences the effects of various detrimental mechanisms in the photocurrent production mechanism is the charge transport. So, a good indicator about how fast and efficient the energy conversion is, can be represented by carriers lifetime. Let us start by analyzing the first process competing with photogeneration: the radiative recombination. For small Δn and Δp we can approximate radiative lifetimes with [50]

$$\tau_{n,\rm rad} = \frac{n_i^2}{B_{\rm rad}\, n} \tag{2.47}$$

and

$$\tau_{p,\rm rad} = \frac{n_i^2}{B_{\rm rad}\, p}\,, \tag{2.48}$$

respectively for n- and p-doped samples. According to the van Rosbroeck-Shockley theory [51] the MCT radiative coefficient entering in the related recombination coefficient B (see Eq. (1.65)) is expressed by

$$B_{\rm rad} = \frac{8\pi}{h^3c^3}\int_0^\infty \frac{\epsilon(E)\,\alpha(\nu)\,E^2}{e^{\frac{E}{k_{\rm B}T}}-1}\,dE\,. \tag{2.49}$$

Very often the high-frequency dielectric constant ϵ_∞ is used in place of $\epsilon(E)$ within Eq. (2.49) while $\alpha(\nu)$ is usually computed through the Bardeen expression

$$\alpha(\nu,x,T) = \frac{m_0\,q^2}{3\hbar^2}\sqrt{\left(\frac{2\,m_{\rm e}^*\,m_{\rm h}^*}{m_0(m_{\rm e}^*+m_{\rm h}^*)}\right)^3}\left(1+\frac{m_0}{m_{\rm e}^*}+\frac{m_0}{m_{\rm h}^*}\right)\sqrt{\frac{h\nu-E_{\rm g}(x,T)}{\epsilon_\infty(x)\,m_0\,c^2}}\,. \tag{2.50}$$

By plugging Eq. (2.50) into (2.49) and performing the integration one obtains [52]:

$$B_{\mathrm{rad}}(x,T) = 5.8\cdot 10^{-13}\sqrt{\epsilon_\infty(x)\left(\frac{m_0}{m_\mathrm{e}^*+m_\mathrm{h}^*}\right)^3}\left(1+\frac{m_0}{m_\mathrm{e}^*}\right) \times\sqrt{\left(\frac{300\ \mathrm{K}}{T}\right)^3}\left(E_\mathrm{g}^2(x,T)+3\,k_\mathrm{B}TE_\mathrm{g}(x,T)+3.75\,k_\mathrm{B}^2T^2\right) \tag{2.51}$$

that enters the general formula, valid under equilibrium condition approximation,

$$\tau_{\mathrm{rad}} = \frac{1}{B_{\mathrm{rad}}\,(n+p)}\,, \tag{2.52}$$

and where one should notice that also electron and hole effective masses (especially the former) depend on cadmium mole fraction.

The second mechanism, by importance, affecting I_{dark} in IR photodetectors is the Shockley-Read-Hall generation. Even if it is very complicated to predict SRH lifetimes – because this process depends on growth technology – for $x < 0.28$ and $T \simeq 300$ K typical τ_{SRH} should be only slightly greater than τ_{rad}, in the order of magnitude of 10^{-6} s. Due to its remarkable T-dependence, well reproduced by the empirical Sharfetter-like model [53, 54]

$$\tau_{\mathrm{SRH}} = \left(\tau_{\mathrm{min}}+\frac{\tau_{\mathrm{max}}-\tau_{\mathrm{min}}}{1+\left(\frac{N_\mathrm{A}+N_\mathrm{D}}{N_{\mathrm{ref}}}\right)^\gamma}\right)\left(\frac{T}{300\ \mathrm{K}}\right)^\alpha\,, \tag{2.53}$$

at 77 K radiative lifetime returns to be dominant, with τ_{rad} from 10 to 40 µs.

But the mechanism really governing dark current even at room temperature and with a very strong T-dependence, whatever the mole fraction x, is the Auger generation. As reported by A. R. Beattie and A. M. White [55] the possible Auger mechanisms in narrow gap semiconductors are essentially ten. In particular, in HgCdTe, two of these are dominant: "Auger 1" (A1, or CHCC), typical of n-type $\mathrm{Hg_{1-x}Cd_xTe}$ with $x \simeq 0.2$ and high-T, and "Auger 7" (A7 or CHLH), especially in p-HgCdTe but not only.

As one may notice in Figure 2.13, only processes A1 and A2 involves two electrons and one hole ("eeh") while the others are all "ehh"-like, which partially explains the dominance of A1 in n-type and A7 in p-type MCT. For each of them no phonon-assisted transitions have been accounted for, as well as radiative or trap-assisted ones.

Following the same formalism reported in Refs. [56, 57], assuming parabolic band dispersions and in absence of any external stimulus, we can write the GR rates for A1 as

$$\begin{cases} G_{\mathrm{A1}} = \dfrac{n'}{n_0}\,G_0 \\ R_{\mathrm{A1}} = \dfrac{n'^2\,p'}{n_0^2\,p_0}\,R_0 \end{cases}, \tag{2.54}$$

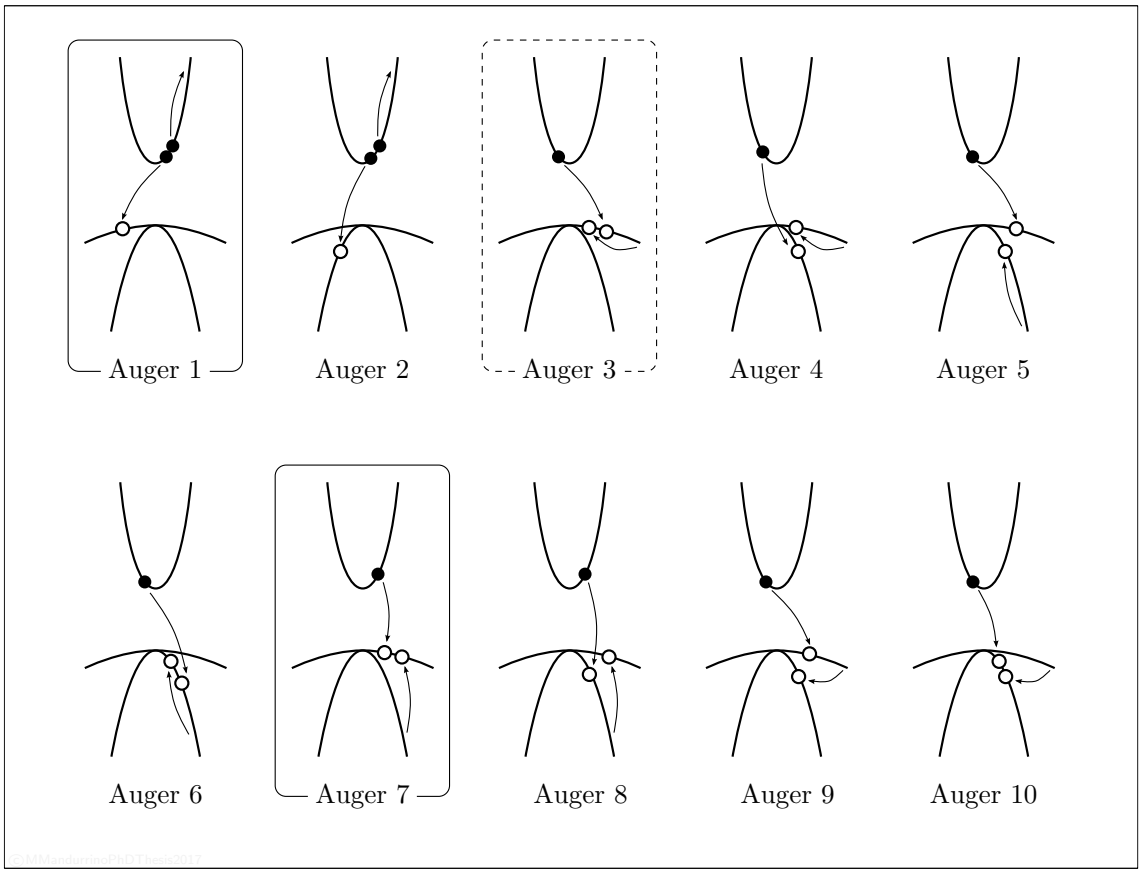

FIGURE 2.13. Simplified illustration of all the Auger recombination mechanisms occurring in direct narrow gap semiconductors, described for the first time in 1996 by A. R. Beattie and A. M. White [55]. "Auger 1" and "Auger 7" are dominant in MCT, closely followed by "Auger 3" (especially in p-HgCdTe). Adapted and re-edited from [45].

where G_0 and R_0 are the steady-state rates. Since, as we have already said in Section 1.4, in equilibrium conditions $\Delta n = \Delta p$ then it follows that

$$\tau_{\mathrm{A1}} = \frac{n_i^4}{n'\, p_0 (n' + p_0)} \frac{1}{G_{\mathrm{A1}}} \,. \tag{2.55}$$

By evaluating the transition probability and integrating it over the wavevector $\mathbf{k}$ under the hypothesis of negligible screening effects, G_0 (which in general dominates over R_0) can be expressed by

$$\begin{aligned} G_0(x,T) \;=\; & \frac{8(2\pi)^{\frac{5}{2}}\, q^4\, m_{\mathrm{e}}^*(x,T)\, |F_{1,2}|^2\, n_0}{h^3\, \epsilon^2 \sqrt{1 + M^*(x,T)}\, (1 + 2M^*(x,T))} \left(\frac{k_{\mathrm{B}} T}{E_{\mathrm{g}}(x,T)} \right)^{\frac{3}{2}} \\ & \times \exp\left(- \frac{1 + 2M^*(x,T)}{1 + M^*(x,T)} \frac{E_{\mathrm{g}}}{k_{\mathrm{B}} T} \right) . \end{aligned} \tag{2.56}$$

So it can be demonstrated that

$$\tau_{\mathrm{A1}}(x,T) = \frac{2n_i^2}{n_0(n_0 + p_0)}\, \tau_{i,\mathrm{A1}}(x,T) \,, \tag{2.57}$$

where

$$\tau_{i,\mathrm{A1}}(x,T) = \frac{n_i}{2\,G_{\mathrm{A1}}(x,T)} \tag{2.58}$$

is the intrinsic "Auger 1" lifetime,

$$M^*(x,T) = \frac{m_{\mathrm{e}}^*(x,T)}{m_{\mathrm{hh}}^*}\,, \tag{2.59}$$

ϵ_{r} is the relative dielectric constant and $|F_{1,2}|$ is the Bloch-wavefunctions overlap between the involved CB and HH band states (ranging from 0.1 to 0.3, as reported in literature [7]). By merging Eqs. (2.56) and (2.57) we obtain the so-called "BLB model" (from its developers A. R. Beattie, P. T. Landsberg and J. S. Blakemore).

By applying the same theoretical assumptions also to the process "Auger 7", A. R. Beattie and G. Smith [58] found that

$$\begin{aligned} \tau_{\mathrm{A7}}(x,T) &= \left[1 + \frac{2n_0\,\Delta p_{\mathrm{hh}}}{p_{0,\mathrm{hh}}\,\Delta n} - \frac{n_0\,\Delta p_{\mathrm{lh}}}{p_{0,\mathrm{lh}}\,\Delta n}\right]^{-1} \frac{\tau_{i,\mathrm{A7}}(x,T)}{2} \\ &\simeq \frac{2n_i^2}{p_0(n_0+p_0)}\,\frac{\tau_{i,\mathrm{A7}}(x,T)}{2} \end{aligned} \tag{2.60}$$

where the intrinsic "Auger 7" lifetime $\tau_{i,\mathrm{A7}}$ is a function of the threshold energy $E_{\mathrm{th}}(x,T)$ of this process. As it can be clearly seen by comparing Eqs. (2.57) and (2.60),

$$\frac{\tau_{\mathrm{A7}}}{\tau_{\mathrm{A1}}} \simeq \frac{n_0}{p_0}\frac{\tau_{i,\mathrm{A7}}}{\tau_{i,\mathrm{A1}}} = \frac{n_i^2}{p_0^2}\frac{\tau_{i,\mathrm{A7}}}{\tau_{i,\mathrm{A1}}} = \frac{n_i^2}{p_0^2}\,\gamma\,, \tag{2.61}$$

where the debated value of γ is spanning from 3 to 8 for $0.16 \leq x \leq 0.40$ and $50 \leq T \leq 300$ K. See, for instance, several experimental results in Refs. [55, 59, 60].

Once the Auger dominant lifetimes have been derived, we can write the total Auger generation rate as [7]

$$\begin{aligned} G_{\mathrm{A}}(x,T) &= \frac{np - n_i^2}{2n_i^2}\left(\frac{n}{(1+a\,n)\,\tau_{i,\mathrm{A1}}(x,T)} + \frac{p}{\tau_{i,\mathrm{A7}}(x,T)}\right) \\ &= \left(C_n(x,T)\,n + C_p(x,T)\,p\right)\left(np - n_i^2\right) \end{aligned} \tag{2.62}$$

where $a_{\mathrm{MCT}} \sim 10^{-18}$ cm^3 and coefficients $C_{n,p}$ are

$$C_n(x,T) = \frac{1}{2\,n_i^2\,\tau_{i,\mathrm{A1}}(x,T)} \tag{2.63}$$

and

$$C_p(x,T) = \frac{1}{2\,n_i^2\,\tau_{i,\mathrm{A7}}(x,T)}\,. \tag{2.64}$$

Eqs. (2.63) and (2.64) can be computed by using the BLB relation

$$\tau_{i,\mathrm{A1}}(x,T) = 3.8\cdot 10^{-18}\epsilon_\infty(x)^2\sqrt{1+M^*(x,T)}\,(1+2M^*(x,T))\,\frac{\exp\left(A(x,T)\right)}{B(x,T)} \tag{2.65}$$

or the semi-empirical law due to Kinch [60]

$$\tau_{i,\mathrm{A1}}(x,T) = \frac{2.12\cdot 10^{-14}\sqrt{E_\mathrm{g}(x,T)}\exp\left(\frac{E_\mathrm{g}(x,T)}{k_\mathrm{B}T}\right)}{|F_{1,2}|^2\,(k_\mathrm{B}T)^{3/2}}\,, \tag{2.66}$$

where

$$A(x,T) = \frac{1+2M^*(x,T)}{1+M^*(x,T)}\frac{E_\mathrm{g}(x,T)}{k_\mathrm{B}T} \tag{2.67}$$

and

$$B(x,T) = m_\mathrm{e}^*(x,T)|F_{1,2}|^2\left(\frac{E_\mathrm{g}(x,T)}{k_\mathrm{B}T}\right)^{3/2}\,. \tag{2.68}$$

Now, remembering that the total lifetime is related to the lifetimes of each single GR contribution (except for tunneling mechanisms that we will introduce in the following chapters) via Eq. (1.33), one can finally write

$$\frac{1}{\tau} = \frac{1}{\tau_\mathrm{rad}} + \frac{1}{\tau_\mathrm{SRH}} + \frac{1}{\tau_\mathrm{A1}} + \frac{1}{\tau_\mathrm{A7}}\,. \tag{2.69}$$

At low temperature Eq. (2.69) is dominated by τ_rad, especially for low mole fractions. While T increases, and in particular at near room temperature, τ_SRH becomes more important, then at high-T Auger lifetime definitely governs among all limitations to the ideal dark photoconduction.

Not only fundamental processes affect the device performance, but also technological or system-related features can do. Other important figures of merit belonging to a generic IR photodetector are the *responsivity* $\mathcal{R}$ (in A/W), which depends on the IQE η through the relation

$$\mathcal{R}(\nu) = \frac{I_\mathrm{ph}}{P_\mathrm{inc}}g(V_\mathrm{bias}) = q\frac{\eta(\nu)}{h\,\nu}\,g(V_\mathrm{bias})\,, \tag{2.70}$$

where $g(V_\mathrm{bias}) = I_\mathrm{ph}/\eta N_\mathrm{inc}$ is the detector gain.

The *Noise Equivalent Power* (NEP, in W units), i.e. the signal required - at fixed λ - to have a unitary signal-to-noise ratio (SNR), is also relevant being

$$\mathrm{NEP}(\nu) = \frac{1}{\mathcal{R}(\nu)}\sqrt{\frac{\langle I\rangle^2}{\Delta\alpha_f}}\,, \tag{2.71}$$

where $\langle I\rangle^2$ is the total mean square current and $\Delta\alpha_f$ is the frequency noise bandwidth or, in other words, the peak FWHM in the spectrum of α as a function of the operating electrical frequency f.

The last parameter, i.e. the *detectivity* D (in W^{-1}), is defined as

$$D(\nu) = \frac{1}{\mathrm{NEP}(\nu)} \tag{2.72}$$

and, similarly, the *normalized detectivity* D^* ($\mathrm{cm\,Hz^{1/2}\,W^{-1}}$) is

$$D^*(\nu) = D(\nu)\sqrt{A\,\Delta\alpha_f} = \frac{\sqrt{A\,\Delta\alpha_f}}{\mathrm{NEP}(\nu)}\,, \tag{2.73}$$

where A is the area of the active region.

By setting Φ as the background flux ($\mathrm{cm^{-2}\ s^{-1}}$), it can be demonstrated [7] that the normalized detectivity (hereinafter, only detectivity) is also

$$D^*(\nu) = \frac{\eta\,q}{h\,\nu}\left[4\frac{k_\mathrm{B}T}{R_0 A} + 2\,q^2\Phi\,\eta\right]^{-\frac{1}{2}}\,, \tag{2.74}$$

in which R_0 represents the zero-bias resistance. From Eq. (2.74) is pretty evident that two operating regimes can be defined, that is:

- the already mentioned BLIP, if $4\frac{k_\mathrm{B}T}{R_0 A} \ll 2\,q^2\Phi\,\eta$;
- the thermal noise limited performance, if $4\frac{k_\mathrm{B}T}{R_0 A} \gg 2\,q^2\Phi\,\eta$

from which emerges that in order to increase $D^*(\nu)$ the product $R_0\,A$ must be maximized (and, obviously, also the IQE).

Chapter 3

Wide Gap Light-Emitting Diodes (LEDs)

Here we conclude the general part of the present thesis by introducing the last area of interest concerning our main target of tunneling modeling in optoelectronic devices: this chapter is entirely devoted to presenting GaN-based light-emitting diodes (LEDs) and their relevant features seen both by the performance and physics-based description standpoint.

Similarly to what has been done about IRPDs, we articulate our discussion on light-emitters starting from the origins of lighting (Section 3.1). Then in Section 3.2 the most important technological issues and physical features about GaN and its alloys will be analyzed, having been this an important part of my work during the first year of Ph.D. Section 3.3, finally, examines the operating principles of the state-of-the-art InGaN/GaN blue LEDs and, besides this, their major operating figures of merit. This section will also allow us to introduce the so-called "efficiency-droop" problem (one of the most hotly debated issues about LEDs performances at high injection levels) partly related to some specific tunneling mechanisms.

3.1 Historical Overview

3.1.1 From SiC to nitrides

The history of solid-state lighting starts with the first assured testimony of a light-emitting diode produced in laboratory which dates back to the early XX century and, more precisely in 1907, when H. J. Round reported his observations about electroluminescence effect in a SiC macroscopic sample put in contact with a conductive electrode. Round's explanation for light emission was partly based on the electromotive force and, secondly, on thermoelectric effect. Nowadays we know that this kind of photoemission is due to a rectifying Schottky junction under forward bias

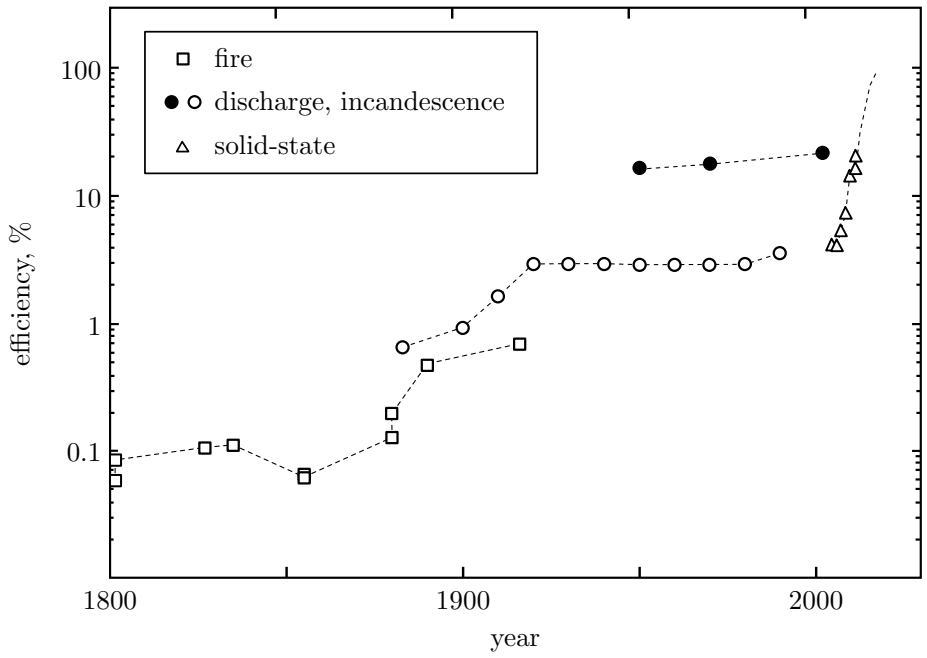

FIGURE 3.1. Evolution of lighting technologies from 1800 to nowadays in terms of devices efficiency. Adapted and re-edited from Ref. [61].

where, at high applied voltage, a strong injection of minority carriers occurs from the metal to the semiconductor by tunneling through the Schottky barrier. Then within bulk SiC the radiative recombination of these carriers takes place, giving rise to the light observed by Round.

Soon after, in 1928, also the russian scientist O. V. Losev reported similar experimental observations not only about SiC but also in ZnO. Moreover, he tried to give a more microscopic explanation to the physical origin of this phenomenon by studying the emission spectra and its threshold onset. In the following decades the fabrication parameters of SiC became more controllable (mainly thanks to the contribution by E. E. Violin) allowing the production of very-low efficient blue LEDs (0.005-0.03%).

What pushed a lot the development of these devices was the demonstration of a reliable production of III-V compounds, in the early 50s of the last century. The first alloy employed in LED development was GaAs. In the form of bulk substrates, GaAs was initially used to grow *p*-*n* junctions via VPE or LPE epitaxial procedures or through the horizontal Bridgman method already seen concerning IRPDs [28]. With Zn-diffusive doping, the first IR LEDs were based on these technologies [62–64]. Then, approximately in 70s, the era of Si-doped GaAs LEDs arrived mainly thanks to the work done by J. Woodall at IBM (NY), allowing the production of devices with external efficiencies up to 6%, five times higher than Zn-doped ones.

But the very breakthrough was represented by a step occurred few years before, in 1962, when N. Holonyak and S. F. Bevacqua [65] published on the first number of a newborn *Journal of Applied Physics* their recent discoveries about a compound,

$Ga(As_{1-x}P_x)$, and its capabilities to emit withint the visible spectrum of the electromagnetic radiation. Almost contemporarily a series of attempts to develop a reliable technology based on GaP materials for green LEDs started. And despite the big issue of obtaining acceptable efficiencies, at the end of 60s this targed was definitely achieved, although with values not greater than few percentage units, depending on the N-doping dose [66, 67].

3.1.2 Blue LEDs

The 70s have also been the era of orange and yellow emission, with several studies focusing on the radiative recombination properties of GaAsP under different N-doping levels [68, 69], even if the efficiency of these devices were heavily degraded by dislocations generated from the substrate/epilayer mismatch. Afterwards, this problem was partially fixed by introducing a significative technology step in the GaP-based LED production, consisting in an optically-active doping procedure (through N or O, see Refs. [70, 71]), allowing to reach high efficiency ($\sim$2% at room-T) through trap-assisted radiative emission, rather than via direct band-to-band transitions as in GaAsP-based devices [72, 73].

At that point the mass production of a new generation of trichromy-LED TVs was almost mature, except for the lack of a blue-light emitter. Several studies on the optical properties of GaN and its band gap were already indicating from 30s that it was the material in which to invest. As a matter of fact, in late 60s GaN was produced from ammonia and liquid gallium only in the form of powder or small crystals [74]. Who gave the pulse to develop new technological methodologies was P. Maruska, a guy from the Radio Corporation of America (RCA). In 1969 he realized the first epitaxy of a "vapour-deposited single-crystal-line GaN" film on sapphire using hydride vapor phase epitaxy (HVPE), as reported in his work [75]. Only two years later, the researches by Maruska, fostered and supported also by Pankove, led to the first blue-light-emitter, a MIS structure consisting of an In metal surface, a Zn insulation layer and a non-intentionally doped n-GaN [76].

The challenge of having high-quality single-crystal and impurity doped p-GaN remained unresolved, despite several attempts by Pankove himself and his coworkers with Zn and Mg. So the possibility to grow p-n structures has been practically impossible till more than a decade after: in 1989 H. Amano from Nagoya University demonstrated the Mg activation in GaN through "low-energy electron beam irradiation" [77] and, almost contemporarily, both I. Akasaki (scientific mentor of the young Amano, inventor of new MOVPE and MBE technique suitable for GaN) [78, 79] and S. Nakamura [80], Nichia Corporation (who introduced improvements in the VPE method), contributed to the development of reliable growth and doping procedures essential for p-n GaN diodes production.

In 1992 Akasaki presented at the "GaAs and related compounds" conference the first example of a blue LED made of GaN on sapphire, with an efficiency of $\sim$1%. This opened the door to a rapid diffusion of nitrides for lighting, such that just the year after Nakamura was presenting the first green and blue InGaN/GaN heterostructures [81, 82] with the historical achievement of efficiencies $\sim$10%.

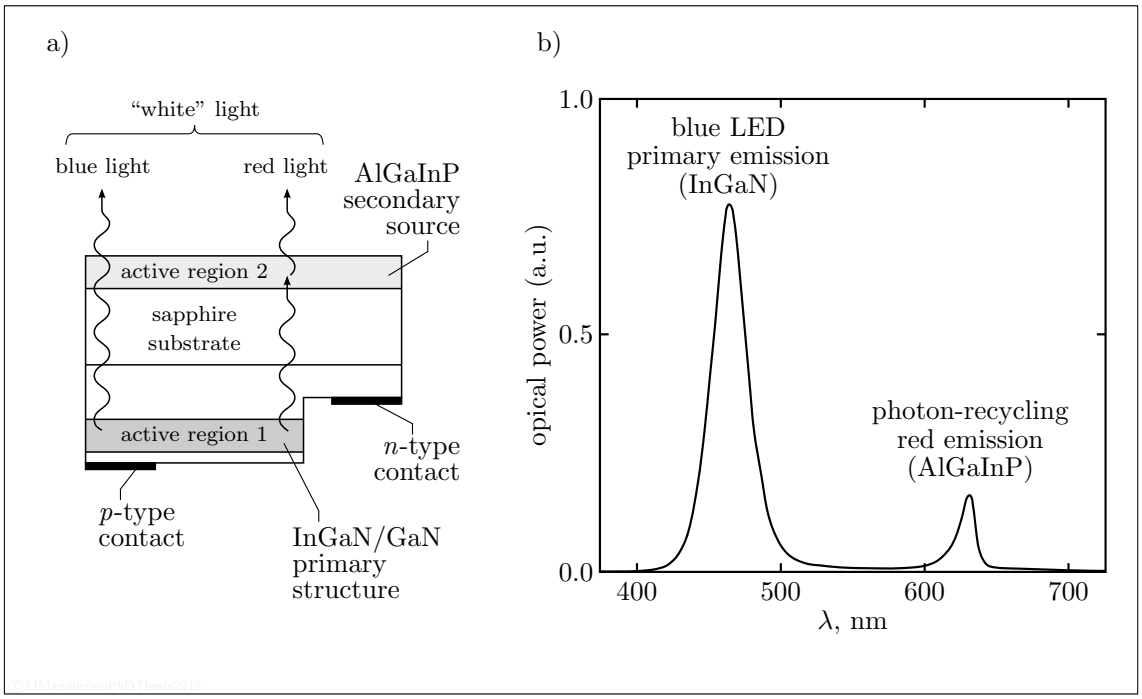

FIGURE 3.2. Photon-Recycling Semiconductor LED (SPR LED) as in Ref. [83] (adapted and re-edited pictures). (a) Schematic structure. (b) Room-temperature luminescence spectrum.

3.1.3 Today lighting

Besides displays and traffic signals, one of the most important applications of modern LEDs is the lighting technology, both for domestic and public utilise. Nevertheless, it has been very difficult for a long time to obtain one of the fundamental ingredients of this area, i.e. the emission of *white* light. This is so important that, for their parallel breakthrough results in the field of solid-state lighting, in 2014 I. Akasaki, H. Amano and S. Nakamura have been honoured with the Nobel Prize in Physics whith the following motivation:

> "*for the invention of efficient blue light-emitting diodes which has enabled bright and energy-saving white light sources*".

The involvement of three scientists mainly dedited to the development of blue LEDs in a so important award with such a kind of motivation has to be reconducted to the nature of white light itself.

Strictly speaking, white light does not actually exists, by the physical standpoint, since white is not properly a *color* with a specific wavelength as real colors. When a certain ratio of three or more wavelengths stimulates the photosensitive cells of human eye, the response of our brain experiences the feeling of a white light. What we can do with technology is to emulate this effect by using two or three different monochromatic sources. To this purpose, the primary emission is partially converted via *photon-recycling* into a second radiation with different wavelength through the use of phosphors, particular types of semiconductors (as

AlGaInP [83]) or with dyes. Figure 3.2 shows a famous example from early literature about Photon-Recycling Semiconductor (SPR) LEDs in which the main active region, consisting in a InGaN/GaN heterostructure with emission line at 470 nm and a second active layer, made of AlGaInP, which partially absorbs the primary radiation and than re-emits at 630 nm. The ratio between the two wavelengths is carefully engineered to obtain the so-called "white"-light-effect we mentioned.

Despite all the still ongoing technological and theoretical developments that started in the past century and although only solid-state lighting (SSL) can reach efficiencies close to 100% (see Figure 3.1), LEDs are still far from being at their ultimate stage. The business in GaN-based LEDs was about 7 billion dollars in 2008 and it expected to triplicate within next decade, but the technology of III-nitrides (essentially AlN, InN and GaN) is still young and costly. The main issues arise from the absence of a reliable bulk GaN industry. This represents a big limiting factor for its diffusion. This forces the growth of GaN on different substrates – as Si or SiC – whose lattice mismatch (∼17%) often induces additional costs and, worst, structural defects like dislocations which propagate along the growth direction from substrates to epi-GaN [84]. Defects, in turn, generate traps for electrons (and/or holes) which further make increase the production costs and the need for processes control.

In the power electronics industry, a rising application field of GaN, traps introduced via GaN-on-Si technology are responsible for the main performance degradation such as threshold voltage shift (originating from gate region traps), saturation current collapse (coming, instead, from access region traps) and reduction of on-state resistance (dynamic $R_{\mathrm{DS,on}}$). Thus, the development of a proper GaN-on-GaN technology would be auspicable, both in terms of both growth and doping quality, given also the thermal instability of dopant activation process suffered by the actual GaN-on-Si technology [85].

Within the SSL area, instead, defects are mainly at the origin of trap-assisted transport phenomena occurring under forward-bias conditions (see, e.g., Ref. [86]). These mechanisms can produce leakage currents like tunneling [87], a very common phenomenon in low-injection regimes but partly relevant also at high applied currents [88, 89].

Besides defects, other two big issues affect the GaN development for lighting in terms of quantum efficiency: one is the so-called *green gap*, i.e. the performance degradation in the range of emitted radiation between ∼500 and ∼600 nm [90] (from which comes the name). Another debated problem is the *efficiency droop* (or, only *droop*), i.e. the drop of the internal quantum efficiency as a function of injection currents and/or temperature (see Figure 3.3). While the first effect is essentially a matter of technology, and it is somehow compensated by a higher sensitivity of the human eye within the green window, the second one concerns more fundamental physical principles and its origin is still under debate, even because a solution for this problem could furtherly improve GaN-based LEDs reliability and at the same time reduce SSL costs.

As one may notice in Figure 3.3, panel (a), the green gap (often also called "yellow-green gap") is determined by the tail of InGaN-based emitters, decreasing for

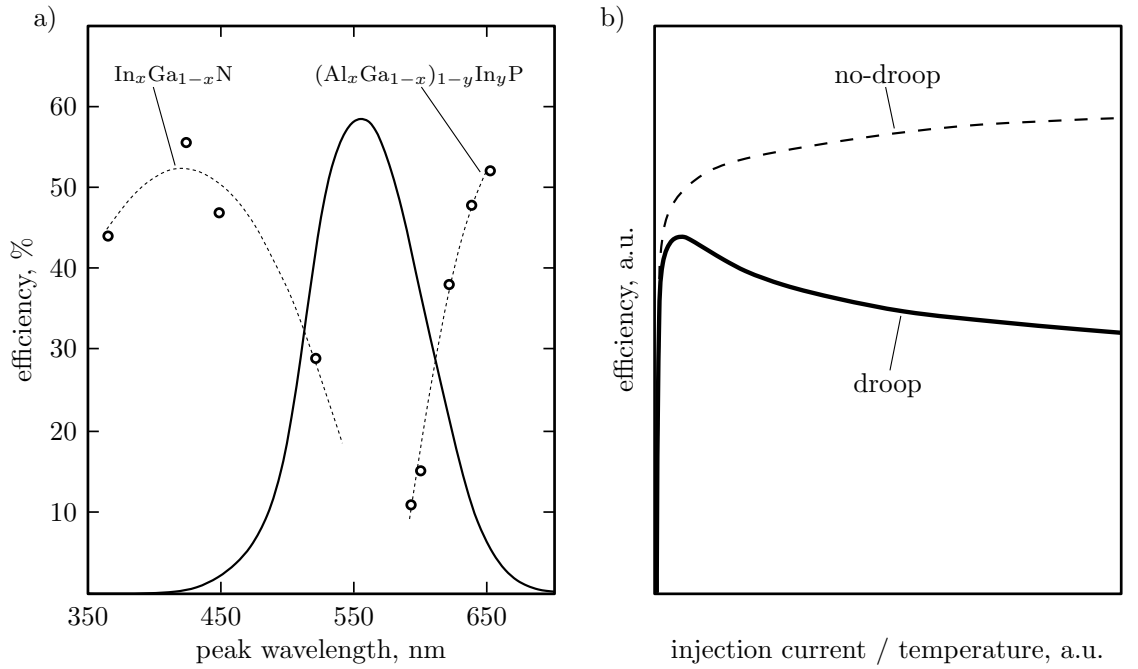

FIGURE 3.3. (a) Representation of the *green gap*: points are measurements while dotted lines represent fitting curves. The black-solid line is the human eye responsivity curve (adapted and re-edited from Ref. [91]). (b) Schematic illustration of the *efficiency droop*, where the ideal trend of efficiency without performance degradation is also represented.

peak intensities at ∼525 nm, and the raising trend of AlGaInP-based LEDs starting from approximately ∼600 nm. As already said, this lack is partially compensated by the human eye responsivity curve centered at ∼555 nm. In panel (b), instead, we find a simplified and schematic representation of the efficiency droop phenomenon. For increasing injection currents (or temperature) the real trend of the internal quantum efficiency (IQE) reaches a maximum, which depends on the device quality, then experiences a decrease. In the ideal case (dashed curve) the IQE should saturate at a certain value (again, depending on the particular manufacturing technology).

3.2 Properties of III-nitride materials

As light-emitters, LEDs are devices mainly working through radiative recombination mechanisms. In order to enhance it, their design requires that the emission is maximized in terms of efficiency. This means that, first, optical dispersions and carrier leakages have to be prevented and, on the other hand, that radiative recombination itself has to be optimized. This result is achieved by localizing the process strictly within the active region, exploiting the quantum confinement of carriers. To this purpose, LEDs are realized with the intercalation of thin layers of different materials with different energy gaps, which generate discontinuities and valleys along the potential called *quantum-wells* (QWs).

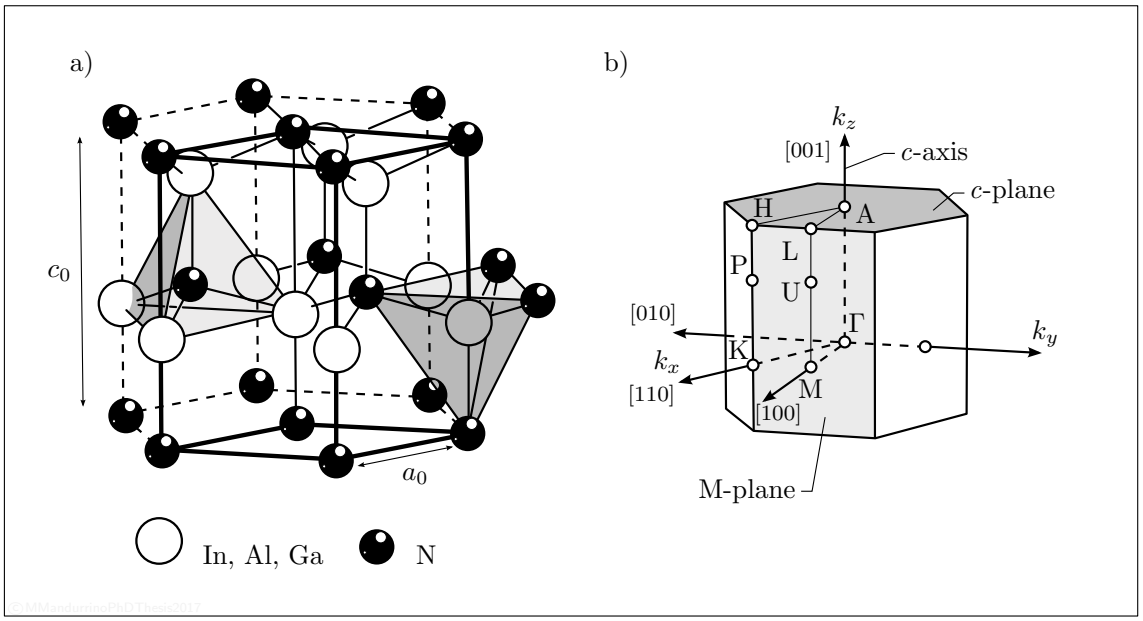

FIGURE 3.4. (a) Hexagonal cell of wurtzite (In-Al)GaN in the real space and its lattice parameters a_0 and c_0. Notice the two elementary tetrahedral coordinations: on the left the one centered on nitrogen and on the right the other one centered on a III-group element. (b) First Brillouin zone (FBZ) of wurtzite (In-Al)GaN in the reciprocal space, where the most important symmetry points and the main crystal planes and directions have been highlighted. The c-axis usually corresponds to the MOCVD growth direction.

In this section we focus on the properties of GaN, the direct band gap compound used as bulk material for blue LEDs fabrication, and the ternary alloys (still direct band gap semiconductors from III-V groups) $In_xGa_{1-x}N$, used as the QW active region material and $Al_xGa_{1-x}N$, a wider band gap material often employed to prevent the unwanted transport of electrons from the n-type material to the p-side of the diode (in the so-called Electron-Blocking Layer, or EBL).

3.2.1 Growth techniques and lattice structure

As already mentioned in our brief historical review, one of the most important techniques for bulk GaN production is the MOCVD and, in particular its MOVPE variant. By controlling the flow of precursor gases into the reactor chamber it is possible to manage both compositions and epitaxy velocity. Typically, in III-nitride-based production trimethylgallium (TMGa), trimethylaluminum (TMAl) and trimethylindium (TMIn) are used as main ingredients of the growth process, while H plays the role of carrier gas (as in NH_3, which introduces the nitrogen). Also dopants are added as organic precursors: for instance, the precursor of Mg, one of the most common p-type dopant in GaN, is dienyl magnesium (Cp_2Mg).

Since the final lattice composition is proportional to the flux ratio between components, in order to grow a certain $In_xGa_{1-x}N$ or $Al_xGa_{1-x}N$ ternary alloy with a precise indium or aluminum mole fraction x it is almost sufficient to linearly control

the molar ratio of their respecting precursors. So:

$$x = \frac{[\text{Al-In}]^{(\text{gas})}}{[\text{Al-In}]^{(\text{gas})} + [\text{Ga}]^{(\text{gas})}}\,, \tag{3.1}$$

where the label "gas" refers to the gaseous phase of metalorganic precursors that contain In or Al. The unique difference between the process involving the indium is the temperature, higher (and up to 1120 K) than in case of aluminum.

The III-N stable phase of GaN, AlN and InN at room temperature and atmospheric pressure is the wurtzite (see Figure 3.4) but also the zincblende can be obtained via epitaxial procedures. Both are characterized by a tetrahedral coordination where each site is surrounded by four nearest-neighbor atoms, but the wurtzite (in which we are interested for SSL applications) has an hexagonal cell while the latter is characterized by a "fcc" cell like in HgCdTe. As shown in Figure 3.4, panel (a), the elementary cell of (In-Al)GaN has two lattice parameters (which are summarized in Table 3.1). In the same image one may also notice the tetrahedral coordination around V- and III-group sites. Panel (b), instead, illustrates the FBZ (in reciprocal space) as well as some of the most characteristic crystal directions (like the growth c-axis) and planes. As usual, the Γ symmetry point lies at the center of the cell.

TABLE 3.1. Lattice parameters

MATERIAL	**PRIMARY PARAMETER**, a_0	**SECONDARY PARAMETER**, c_0
InN	3.533 Å [92]	5.693 Å [92]
AlN	3.112 Å [93]	4.982 Å [93]
GaN	3.189 Å [75, 94]	5.185 Å [75, 95]

Basic lattice parameters concerning 300 K wurtzite InN, AlN and GaN.

Since Ga-N bonds are polar and thanks to the fact that there is no inversion point in the lattice symmetry, such that it is impossible to find at coordinate (x, y, z) the same atom present at $(-x, -y, -z)$, III-N materials exhibit an intrinsic spontaneous polarity which manifest at cleavage planes, interfaces or external facets as net *polarization charges*.

Figure 3.5 shows the lattice constants a_0 of wurtzite GaN, AlN and InN and their relationship with the energy gap. The plot also represents the visible spectrum, for a comparison with the E_g of $In_xGa_{1-x}N$ and the indium mole fraction x. As already said, one of the hardest problems related to GaN growth is the substrate. The reason is shown by the example of 6H-SiC (wurtzite SiC), whose principal lattice constant is $a_{0,\text{SiC}} = 3.07$ Å, as one may notice. Often also n-Si(111) substrate (with minimum interatomic distance $d_{\text{Si(111)}} = 3.84$ Å and lattice constant $a_{0,\text{Si}} = 5.431$ Å) is used and, on top of it, intermediate layers (e.g., AlN) are grown before GaN in order to gradually relax the strain forces induced by lattice mismatch.

Not only $In_xGa_{1-x}N$ and $Al_xGa_{1-x}N$ lattice parameters are affected by the aluminum or indium content, but also the temperature play an important role in de-

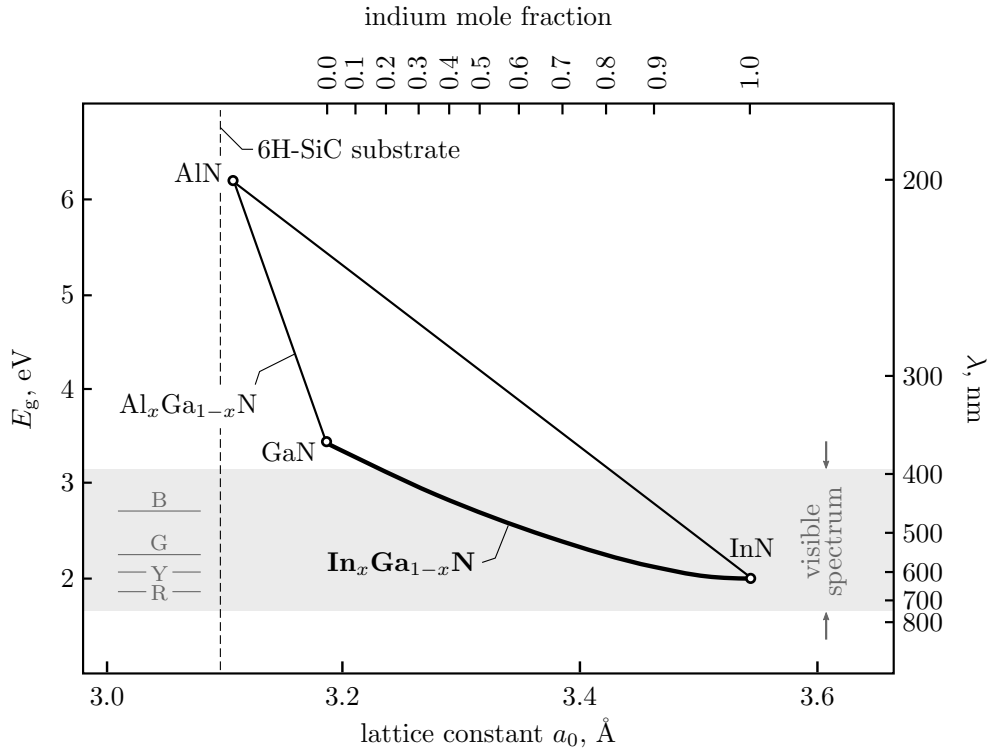

FIGURE 3.5. Multiple plot representing the room-T energy gap (and the emission wavelength) as a function of the primary lattice constant a_0 for ternary alloys AlGaN and InGaN. In this latter case, also the trend of E_g as a function of the indium mole fraction is highlighted. The grey area represents the visible region of the electromagnetic spectrum (with main color lines highlighted on the left). The graph also shows the lattice constant of a 6H-SiC (wurtzite SiC) substrate (vertical dashed line).

termining the lattice structure and its conformation. As reported by several authors (e.g., see Ref. [96]) the primary parameter follows the same law valid for GaN, so:

$$a(T) = a_0 \left(1 + A\,T + B\,T^2\right) , \tag{3.2}$$

where our a_0 is such that $a_0 \equiv a(0\,\mathrm{K})$ and in which A and B are fitting parameters. Moreover, a similar second-order trend can be demonstrated also for $c(T)$. Then, calculated via VCA or measured through diffraction experiments, the x-dependence of both a and c parameters in InGaN and AlGaN show to follow an almost linear law with small bowing factors [97].

3.2.2 Band structure

Never as important as in III-nitride materials intended for emitting QWs the study of energy dispersion is, since bands shape can heavily determine the probability of optical transitions in the LEDs' active region. Several approaches have been employed in the past to derive the trend of $E(\mathbf{k})$, but all of them are based on the $\mathbf{k}\cdot\mathbf{p}$ perturbation method under effective mass approximation (see Section 1.1), with the inclusion of spin-orbit and strain effects.

Generally, since most of the optical and transport mechanisms involves a reasonably small region around Γ, two CBs and six VBs are considered (8×8 $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian formalism), where the conduction states are assumed to be approximately symmetric with respect to Γ with a parabolic $\mathbf{k}$-dependence. The CB Luttinger-Kohn Hamiltonian for wurtzite III-nitrides has the form

$$H_{\mathrm{C}}(\mathbf{k},\varepsilon) = \frac{\hbar^2\mathbf{k}_z^2}{2m^*_{\mathrm{e}||}} + \frac{\hbar^2\left(\mathbf{k}_x^2+\mathbf{k}_y^2\right)}{2m^*_{\mathrm{e}\perp}} + a_{1\mathrm{c}}\varepsilon_{zz} + a_{2\mathrm{c}}\left(\varepsilon_{xx}+\varepsilon_{yy}\right), \tag{3.3}$$

where $m^*_{\mathrm{e}||}$ and $m^*_{\mathrm{e}\perp}$ are the electron effective mass (at the Γ point) parallel and perpendicular to the growth direction, respectively, $a_{1\mathrm{c}}$ and $a_{2\mathrm{c}}$ are the Bir-Pikus deformation coefficients [98] and ε represents the strain tensor components, which depend on the lattice constant and on the eventual strain force, where

$$\varepsilon_{\perp} = \varepsilon_{xx} + \varepsilon_{yy}, \tag{3.4}$$

$$\varepsilon_{\pm} = \varepsilon_{xx} - \varepsilon_{yy} \pm 2i\varepsilon_{xy} \tag{3.5}$$

and

$$\varepsilon_{\pm z} = \varepsilon_{xz} \pm i\varepsilon_{yz}. \tag{3.6}$$

Furthermore, for the valence bands we have

$$\begin{aligned} H_{\mathrm{V}}(\mathbf{k},\varepsilon) \;=\; & \Delta_1 L_z^2 + \Delta_2 L_z\sigma_z + \sqrt{2}\Delta_3\left(L_+\sigma_- + L_-\sigma_+\right) \\ & + \left(A_1 + A_3L_z^2\right)\mathbf{k}_z^2 + \left(A_2 + A_4L_z^2\right)\mathbf{k}_\perp^2 - A_5\left(L_+^2\mathbf{k}_-^2 + L_-^2\mathbf{k}_+^2\right) \\ & -2A_6\mathbf{k}_z\Big(\left[L_z,L_+\right]\mathbf{k}_- + \left[L_z,L_-\right]\mathbf{k}_+ + iA_7\left(L_+\mathbf{k}_- - L_-\mathbf{k}_+\right) \\ & + \left(D_1 + D_3L_z^2\right)\varepsilon_{zz} + \left(D_2 + D_4L_z^2\right)\varepsilon_{\perp} - D_5\left(L_+^2\varepsilon_- + L_-^2\varepsilon_+\right) \\ & -2D_6\left(\left[L_z,L_+\right]\varepsilon_{-z} + \left[L_z,L_-\right]\varepsilon_{+z}\right)\Big), \end{aligned} \tag{3.7}$$

where $\Delta_2 = \Delta_3 = \Delta_{\mathrm{SO}}/3$ depend to the spin-orbit splitting energy Δ_{SO} (already seen for MCT) and $\Delta_1 = \Delta_{\mathrm{CR}}$ the crystal-field splitting energy (an effect of degeneracy removal caused by charges distribution in the lattice producing the splitting of HH and LH sub-bands), the orbital and spin momentum operators $L_\pm$, $\sigma_\pm$ and $\mathbf{k}_\perp$, $\mathbf{k}_\pm$ are defined as

$$L_{\pm} = \frac{L_x \pm iL_y}{\sqrt{2}}, \tag{3.8}$$

$$\sigma_{\pm} = \frac{\sigma_x \pm i\sigma_y}{2}, \tag{3.9}$$

$$\mathbf{k}_{\perp} = \mathbf{k}_x^2 + \mathbf{k}_y^2, \tag{3.10}$$

and

$$\mathbf{k}_{\pm} = \mathbf{k}_x \pm i\mathbf{k}_y, \tag{3.11}$$

being valid the rule

$$\left[L_i,L_j\right] = \frac{L_iL_j + L_jL_i}{2}, \tag{3.12}$$

and where D_i are the Bir-Pikus shear deformation potentials (in eV) and A_i the inverse hole effective mass parameters.

Since the wurtzite configuration represents a particular case of the zincblende, some terms are derived from the latter as perturbation corrections so that one can finally write that [99]

$$A_1 - A_2 = -A_3 = 2A_4 = 4A_5 - \sqrt{2}A_6\,, \tag{3.13}$$

$$A_7 = 0 \tag{3.14}$$

$$D_1 - D_2 = -D_3 = 2D_4 = 4D_5 - \sqrt{2}D_6\,. \tag{3.15}$$

Now we have all the ingredients to derive the energy dispersion relations of wurtzite GaN. For what concerns the main CB, around Γ Eq. (3.3) yields a quasi-parabolic trend of the form

$$E_{\mathrm{C}}(\mathbf{k}) \simeq E_{\mathrm{g}}\,\frac{\mathbf{k}^2}{2m^*_{\mathrm{C}}} + \text{strain effects}\,. \tag{3.16}$$

In the case of valence bands it is convenient, due to the complexity, to reduce the problem in a matrix form. Following the theoretical approach described in Ref. [100] we chose a set of (normalized) basis functions

$$\phi_1 = \left(1/\sqrt{2}\right)|(X+iY)\,,\alpha\rangle \tag{3.17}$$

$$\phi_2 = \left(1/\sqrt{2}\right)|(X+iY)\,,\beta\rangle \tag{3.18}$$

$$\phi_3 = |Z,\alpha\rangle \tag{3.19}$$

$$\phi_4 = |Z,\beta\rangle \tag{3.20}$$

$$\phi_5 = \left(1/\sqrt{2}\right)|(X-iY)\,,\alpha\rangle \tag{3.21}$$

$$\phi_6 = \left(1/\sqrt{2}\right)|(X-iY)\,,\beta\rangle \tag{3.22}$$

that allow the diagonalization of the Hamiltonian in Eq. (3.7), where $|X\rangle$, $|Y\rangle$ and $|Z\rangle$ represent Bloch functions near Γ and $|\alpha\rangle$, $|\beta\rangle$ are spin-up and spin-down functions, respectively. The Hamiltonian, then, can be rewritten as

$$\mathbf{H}_{\mathrm{V}}(\mathbf{k}) = \begin{pmatrix} F & 0 & -H^* & 0 & K^* & 0 \\ 0 & G & \Delta & -H^* & 0 & K \\ -H & \Delta & \lambda & 0 & I^* & 0 \\ 0 & -H & 0 & \lambda & \Delta & I^* \\ K & 0 & I & \Delta & G & 0 \\ 0 & K & 0 & I & 0 & F \end{pmatrix}, \tag{3.23}$$

where $F = \Delta_1 + \Delta_2 + \lambda + \theta$, $G = \Delta_1 - \Delta_2 + \lambda + \theta$, $H = iA_6\mathbf{k}_z\mathbf{k}_+ - A_7\mathbf{k}_+$, $I = iA_6\mathbf{k}_z\mathbf{k}_+ + A_7\mathbf{k}_+$, $K = A_5\mathbf{k}_+^2$, $\Delta = \sqrt{2}\Delta_3$, $\lambda = A_1\mathbf{k}_z^2 + A_2(\mathbf{k}_x^2 + \mathbf{k}_y^2)$ and $\theta = A_3\mathbf{k}_z^2 + A_4(\mathbf{k}_x^2 + \mathbf{k}_y^2)$.

By using now Eqs. (3.13)–(3.15) the dispersion relation for valence states can be calculated by solving the equation

$$\det|\mathbf{H}_{\mathrm{V}}(\mathbf{k}) - E(\mathbf{k})\mathbf{I}| = 0\,, \tag{3.24}$$

in which $\mathbf{I}$ obviously represents a 6×6 identity matrix. In absence of split-orbit coupling interaction the first valence eigenvalues in Γ are simply given by

$$E_1(0) = \Delta_1 + \Delta_2 \tag{3.25}$$

and

$$E_{2,3}(0) = \frac{\Delta_1 - \Delta_2 \pm \sqrt{(\Delta_1 - \Delta_2)^2 + 8\Delta_3^2}}{2}\,, \tag{3.26}$$

as indicated in Figure 3.6. Under the same current hypotheses, the dispersion relations become

$$E_1(\mathbf{k}) = F' - K' \tag{3.27}$$

and

$$E_{2,3}(\mathbf{k}) = \frac{F' - K' + \lambda \pm \sqrt{(F' - K' - \lambda)^2 + 8H'^2}}{2}\,, \tag{3.28}$$

being

$$F' = \Delta_1 + (A_1 + A_3)\,\mathbf{k}_z^2 + (A_2 + A_4)\left(\mathbf{k}_x^2 + \mathbf{k}_y^2\right)\,, \tag{3.29}$$

$$K' = A_5\left(\mathbf{k}_x^2 + \mathbf{k}_y^2\right) \tag{3.30}$$

and

$$H' = \sqrt{(A_6^2\mathbf{k}_z^2 + A_7^2)\left(\mathbf{k}_x^2 + \mathbf{k}_y^2\right)} \tag{3.31}$$

from which the HH, LH and split-off hole effective masses can be obtained by deriving the energies $E_1(\mathbf{k})$, $E_2(\mathbf{k})$ and $E_3(\mathbf{k})$, respectively (for positive Δ_1, otherwise $2 \leftrightarrow 3$), with respect to $\mathbf{k}$. Moreover, one may notice from Eqs. (3.27)–(3.31) that $E(\mathbf{k})$ are almost isotropic in the plane $\mathbf{k}_x \times \mathbf{k}_y$, so for this reason Figure 3.6 shows only two axes of the $\mathbf{k}$-space. From panels (b) and (c) we can appreciate the effect of split-off interaction removing the degeneracy between HH and LH sub-bands (in the $\mathbf{k}_z$-plane) and also the one at LH-SO intersection point. Notice also that from Eqs. (3.16) and (3.27) it is possible to calculate the energy gap as $E_{\mathrm{g}} = E_{\mathrm{C}} - E_1(0)$. Typical values of GaN band parameters at 300 K are $E_{\mathrm{g}} = 3.39$ eV, $E_{\mathrm{A}} = 4.7$-5.5 eV, $E_{\mathrm{L\text{-}M}} = 4.5$-$5.3$ eV, $\Delta_{\mathrm{SO}} = 0.008$ eV and $\Delta_{\mathrm{CR}} = 0.04$ eV (see Ref. [100]).

Starting from the proper Luttinger and Bir-Pikus parameters of AlN, InN and their ternary alloys with gallium one could, in principle, calculate the energy dispersion for all these materials through the same *ab initio* procedure we briefly derived here. The mole fraction shapes InGaN and AlGaN bands through the x-dependence of the primary lattice constant and, in turn, via all the Hamiltonian terms as, among

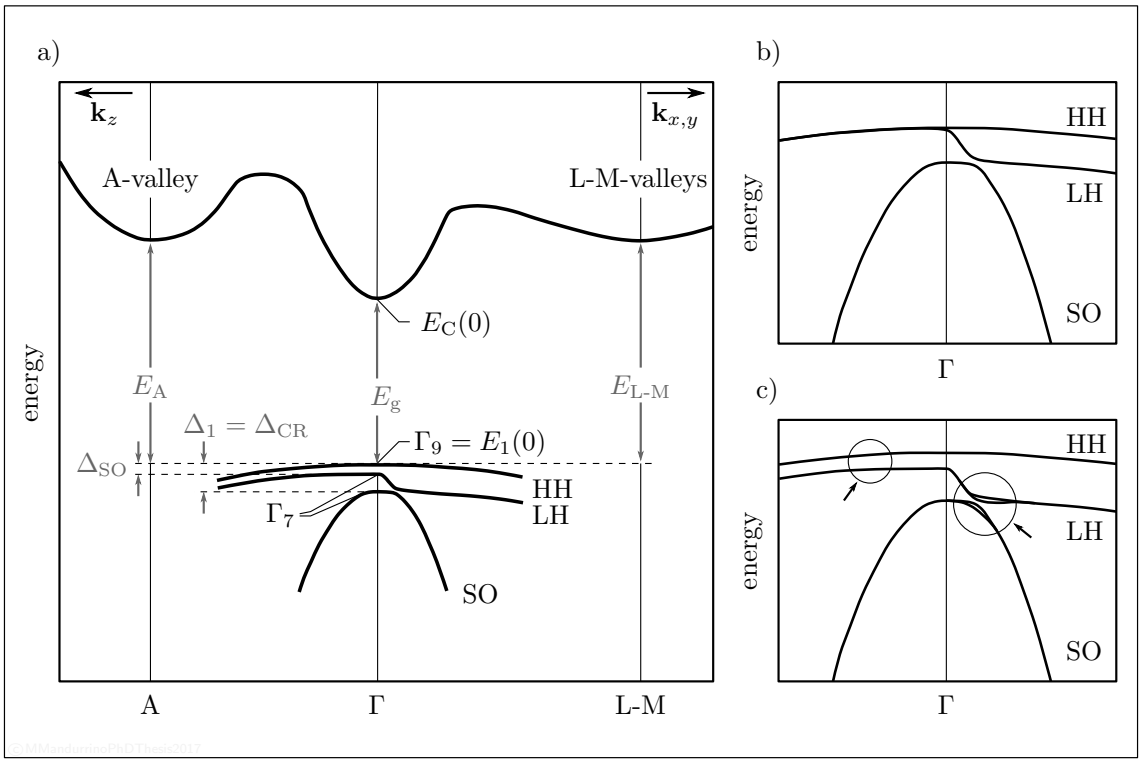

FIGURE 3.6. (a) Band structure of wurtzite GaN without strain effects. For the meaning of labels, see the text. (b)–(c) Zoom of the valence band structure without and with spin-orbit coupling effects, respectively. Notice that in the case with interactions some degeneracies are removed, as between HH and LH sub-bands in the space of vertical crystal momenta and the ones in LH and SO sub-bands near their intersection point in the $\mathbf{k}_x \times \mathbf{k}_y$ plane (as highlighted by arrows).

all, the strain tensor ε. The most common law for describing the dependence of the gap to the alloy composition and temperature is (see, e.g., Ref. [101])

$$E_{\mathrm{g,\circ GaN}}(x,T) = (1-x)\, E_{\mathrm{g,GaN}}(T) \;+\; x\, E_{\mathrm{g,\circ N}}(T) \;-\; x\,(1-x)\, b_{\circ\mathrm{GaN}}\,, \tag{3.32}$$

where the T-dependence

$$E_{\mathrm{g}}(T) = E_{\mathrm{g}}(0\ \mathrm{K}) - \frac{\alpha T^2}{\beta + T} \tag{3.33}$$

derives from the well-known *Varshni model* [102], b is the bowing factor already mentioned in this section, α and β are fitting parameters and the symbol "$\circ$" stands for In or Al, alternatively.

3.2.3 GaN, InGaN and AlGaN material libraries

Part of the first year of my Ph.D. has been dedicated to fitting out and optimizing the materials input files afterwards employed in (semiclassical) numerical simulations of tunneling in GaN-based blue LEDs. The present subsection represents a summary of self-consistent results derived from both theoretical and comparative studies, also

thanks to some experimental feedback provided by the colleagues and collaborators from Padua University. For a detailed treatment of implementing methods employed in semiclassical LED simulations see Appendix B.

The first important property of each semiconductor is the energy gap, since most of the relevant transport and recombination mechanisms occur at the center of the Brillouin zone. Concerning the implementation of Eqs. (3.32) and (3.33), the physical and fitting parameters we adopted are listed in Table 3.2.

TABLE 3.2. Energy gap parameters

MATERIAL	$E_g(0\,\mathrm{K})$, eV	α, eV/K	β, K	b, eV
GaN	3.507 [103]	$9.09 \cdot 10^{-4}$ [103]	830 [103]	–
InN	*0.735	$2.45 \cdot 10^{-4}$ [104]	624 [104]	–
AlN	*6.230	$17.99 \cdot 10^{-4}$ [104]	1462 [104]	–
InGaN	–	–	–	3 [105]
AlGaN	–	–	–	1 [105]

Physical and fitting parameters used to model the x- and T-dependence of energy gap in GaN, InGaN and AlGaN with Eqs. (3.32) and (3.33). *Note: values in accordance with Ref. [105]

By fixing x and changing T, or viceversa, the functions $E_g(T)$ and $E_g(x)$ for ternary alloys $In_xGa_{1-x}N$ and $Al_xGa_{1-x}N$ are represented in Figure 3.7.

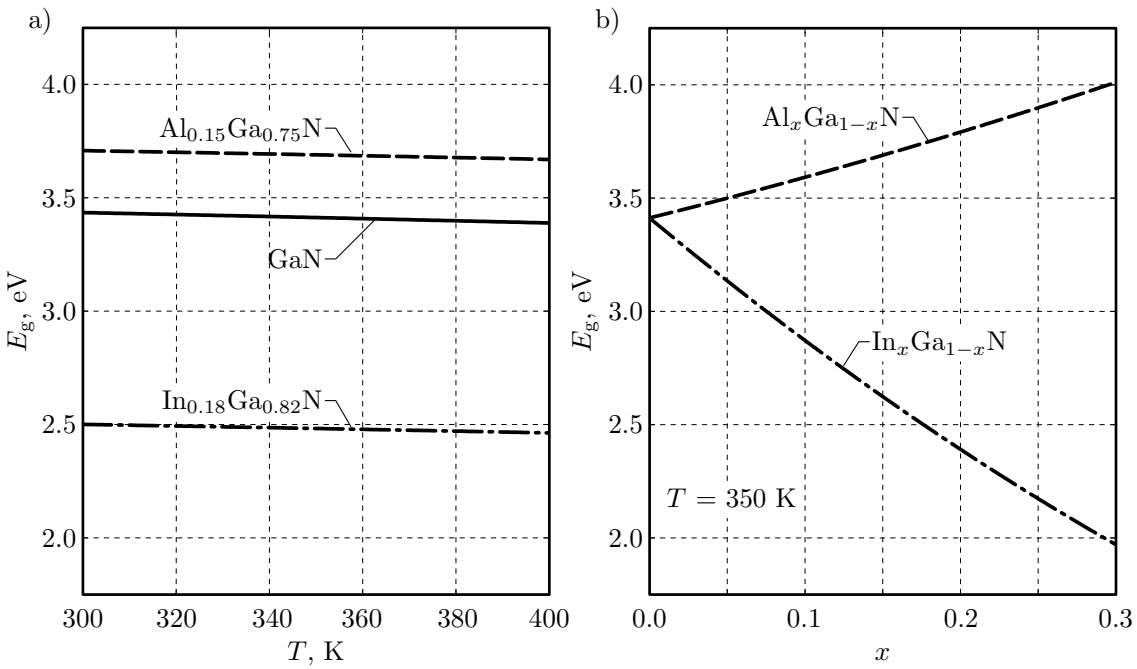

FIGURE 3.7. (a) GaN, InGaN and AlGaN energy gap versus T where we arbitrarily chose $x_{In} = 0.18$ and $x_{Al} = 0.15$. (b) $E_g(x)$ for InGaN and AlGaN at fixed temperature $T = 350$ K.

As one may notice, the increase of indium content reduces the gap (this is the reason why InGaN is actually the material for QWs in blue LEDs) whereas, increasing the aluminum fraction, E_g furtherly widens, and this is why AlGaN can often find a place in EBL for preventing electron leakage from the active region.

Knowing the energy dispersion and recalling Eq. (1.6) we can obtain another fundamental electronic property: the carrier effective mass. In performing semiclassical simulations the most relevant values are those of electrons and heavy holes. We report in Table 3.3 the masses used in this work and, for the sake of completeness, also some values extracted from literature.

TABLE 3.3. Effective mass

MATERIAL	SOURCE	m^*_e, m_0	m^*_{hh}, m_0	m^*_{lh}, m_0	m^*_{so}, m_0
GaN	Ref. [100] (1995)	0.18	1.74	1.74	0.15
	Ref. [106] (2001)	0.20	0.54-2.20	–	–
	Ref. [105] (2003)	0.18-1.23	0.54-2.20	–	–
	this work	0.20	1.50	1.50	–
InN	Ref. [107] (2003)	0.137	2.493	2.493	0.137
	this work	0.133	2.444	2.444	–
AlN	Ref. [107] (2003)	0.232	2.370	2.370	0.209
	this work	0.200	2.300	2.300	–

List of electron and hole effective masses, as reported in the recent literature and as chosen in this work.

It has to be noticed that, although only binary effective masses have been reported, the values concerning ternary alloys $In_xGa_{1-x}N$ and $Al_xGa_{1-x}N$ can be self-consistently deduced from them by using the fitting law [107] (see Figure 3.8)

$$m^*_{{}_xGa_{1-x}N}(x) = m^*_{GaN} + x\ (m^*_{{}_xN} - m^*_{GaN})\ . \tag{3.34}$$

Moreover, in Refs. [100, 105–107] hole effective masses are computed under the low-**k** approximation from Luttinger-Kohn parameters as follows:

$$m^*_{hh} = -m_0\,(A_1 + A_3)^{-1}\ , \tag{3.35}$$

$$m^*_{lh} = -m_0\left(A_1 + A_3\frac{E_2(0) - \lambda}{E_2(0) - E_3(0)}\right)^{-1} \tag{3.36}$$

and

$$m^*_{so} = -m_0\left(A_1 + A_3\frac{E_3(0) - \lambda}{E_3(0) - E_2(0)}\right)^{-1} \tag{3.37}$$

Since band structure is critically dependent on the method used for its calculation, it is straightforward the fact that effective masses, related to the band

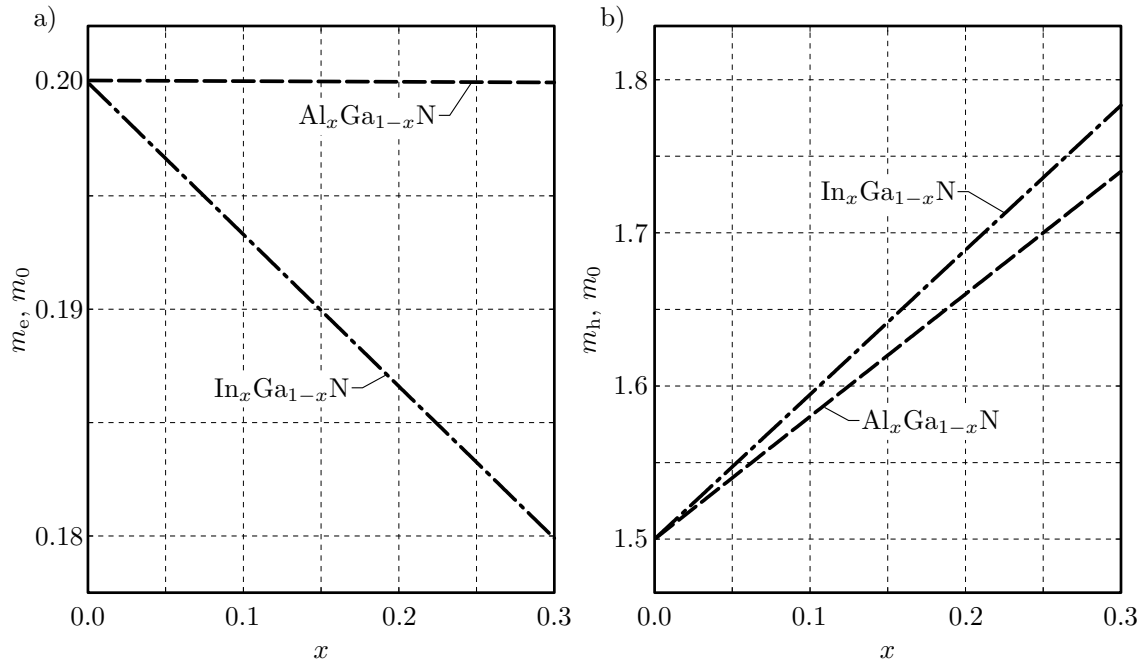

FIGURE 3.8. $In_xGa_{1-x}N$ and $Al_xGa_{1-x}N$ (a) electron and (b) hole effective mass according to Eq. (3.34)

concavity, can be method-dependent. And even if there are different experimental procedures to evaluate them, the literature is not fully agree about m^* values.

Another electrical parameter, still dependent on bands trend and in particular on E_g, is the electron affinity $\chi(x,T)$. Also in this case common Varshni/interpolating combined rules have been used to calculate ternary alloys parameters from those of binary compounds GaN, InN and AlN

$$\chi_{\mathrm{InGaN}}(x,T) = \chi_{\mathrm{GaN}} - \left(E_{\mathrm{g,InGaN}}(x,T) - E_{\mathrm{g,GaN}}(x,T)\right)\,\Delta_{\mathrm{off}} \tag{3.38}$$

and

$$\begin{aligned}\chi_{\mathrm{AlGaN}}(x,T) \;=\; & \chi_{\mathrm{GaN}} - \left(E_{\mathrm{g,AlGaN}}(x,T) - E_{\mathrm{g,AlGaN}}(x,300\ \mathrm{K})\right) \\ & - \left(E_{\mathrm{g,AlGaN}}(x,300\ \mathrm{K}) - E_{\mathrm{g,GaN}}(300\ \mathrm{K})\right)\,\Delta_{\mathrm{off}}\,,\end{aligned} \tag{3.39}$$

(see Figure 3.9) where $\chi_{\mathrm{GaN}} = 4.07$ eV (according to Refs. [95, 105]) and Δ_{off} is the conduction band offset at InGaN/GaN and AlGaN/GaN heterointerfaces calculated as $\Delta_{\mathrm{off}} = \Delta E_C/\Delta E_g$.

Before introducing the last electrical material parameter we now briefly see how the relative dielectric constant (or permittivity) ϵ has been treated. Again, we used an interpolating equation to deduce binary permittivities:

$$\epsilon_{\mathrm{oGaN}} = \epsilon_{\mathrm{GaN}} + x\,\left(\epsilon_{\mathrm{oN}} - \epsilon_{\mathrm{GaN}}\right)\,, \tag{3.40}$$

in which the values $\epsilon_{\mathrm{GaN}} = 9.5$, $\epsilon_{\mathrm{InN}} = 15$ and $\epsilon_{\mathrm{AlN}} = 8.5$ have been used in accordance with Ref. [108], [92] and [93], respectively (see panel (a) of Figure 3.10).

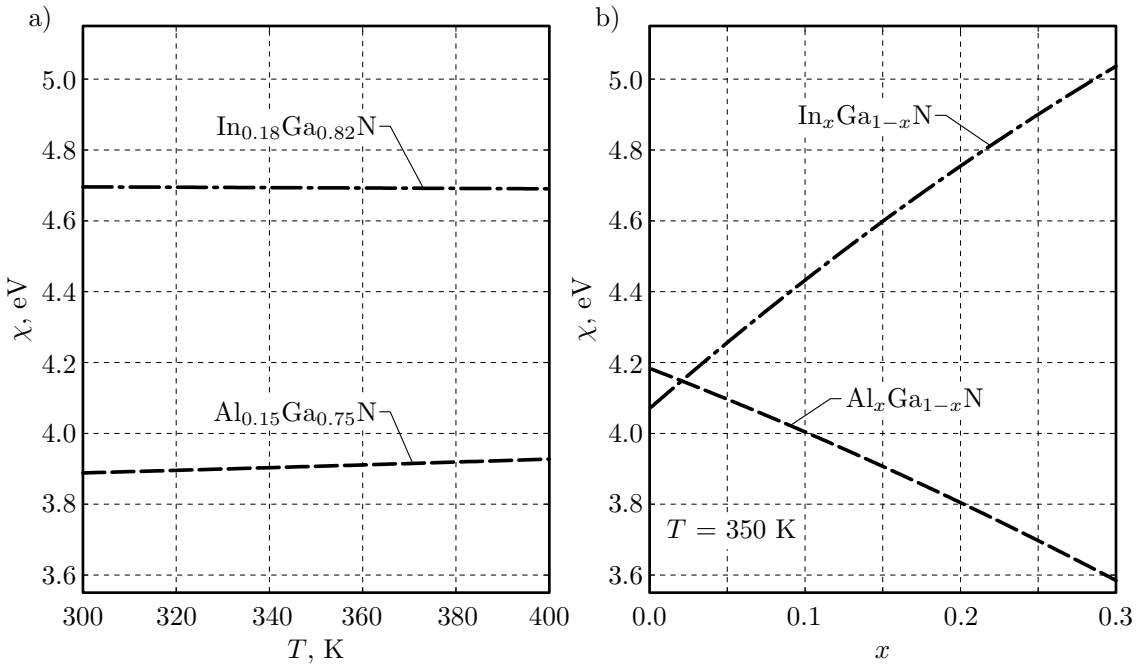

FIGURE 3.9. Electron affinity of InGaN and AlGaN described by Eqs. (3.38) and (3.39) and plotted (a) as a function of temperature, by fixing $x_{\rm In} = 0.18$ and $x_{\rm Al} = 0.15$, and (b) as a function of the mole fraction at fixed $T = 350$ K. In both cases $\Delta_{\rm off} \simeq 0.67$.

As previously anticipated, GaN-based semiconductors are characterized by an intrinsic polarization, which is due to the polarity of III-N bonds and to the lack of a central symmetry in the lattice. As pointed out by O. Ambacher *et al.* (see Ref. [109]), besides this spontaneous component also piezoelectric effects can contribute to the total polarization charge. So, the overall polarization vector can be expressed by

$$\mathbf{P} = \mathbf{P}_{\rm sp}(x) + \mathbf{P}_{\rm pe}(x)\,, \tag{3.41}$$

in which the general form of the 'piezoelectric' term, which is strain- and x-dependent, it is demonstrated to be

$$\begin{aligned}
\mathbf{P}_{\rm pe}(x) &= e_{33}(x)\varepsilon_z + e_{31}(x)\,(\varepsilon_x + \varepsilon_y) \\
&= e_{33}\frac{c(x) - c_0}{c_0} + e_{31}\,2\,\frac{a(x) - a_0}{a_0} \\
&= e_{33}\left(-2\,\frac{C_{13}}{C_{33}}\frac{a(x) - a_0}{a_0}\right) + e_{31}\,2\,\frac{a(x) - a_0}{a_0} \\
&= 2\frac{a(x) - a_0}{a_0}\left(e_{31} - e_{33}\frac{C_{13}}{C_{33}}\right)\,,
\end{aligned} \tag{3.42}$$

where a_0 and c_0 are the usual primary and secondary wurtzite lattice parameters at $T = 0$ K, ε are the strain vector components such that $\varepsilon_z(x) = (c(x) - c_0)/c_0$ and

$\varepsilon_x(x) = \varepsilon_y(x) = (a(x) - a_0)/a_0$, e_{13} and e_{33} are the piezoelectric coefficients and C_{13}, C_{33} are elastic constants.

It is worth noting the fact that the lattice constant here explicitly depends on the mole fraction. The most common approach to model this dependency exploits the so called Vegard's model [110]

$$a_{\circ_x \mathrm{Ga}_{1-x}\mathrm{N}}(x) = A + Bx\,, \tag{3.43}$$

where A and B depend on the material (as usual, the symbol "$\circ$" stands for Al or In).

Few years later with respect to these results, in 2002, Ambacher together with V. Fiorentini and F. Bernardini [111] obtained from first principle calculations very compact equations describing the nonlinear trend of polarization charges with respect to the alloy composition in III-V nitrides. First of all they pointed out that in InGaN and AlGaN Eq. (3.43) has coefficients $A_{\mathrm{InGaN}} = A_{\mathrm{AlGaN}} = 0.31986$, $B_{\mathrm{InGaN}} = 0.03862$ and $B_{\mathrm{AlGaN}} = -0.00891$. Moreover, they demonstrated that in III-nitrides the two components of the polarization vector can be rewritten as

$$\mathbf{P}_{\mathrm{sp},\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}}(x) = -0.042x - 0.034\,(1-x) + 0.038\,(1-x) \tag{3.44}$$

and

$$\mathbf{P}_{\mathrm{sp},\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{N}}(x) = -0.090x - 0.034\,(1-x) + 0.019\,(1-x) \tag{3.45}$$

concerning the spontaneous term, and via the Vegard's form

$$\mathbf{P}_{\mathrm{pe},\circ_x\mathrm{Ga}_{1-x}\mathrm{N}}(x) = x\,\mathbf{P}_{\mathrm{pe},\circ\mathrm{N}}\left(\varepsilon_{\mathrm{bas}}(x)\right) + (1-x)\,\mathbf{P}_{\mathrm{pe},\mathrm{GaN}}\left(\varepsilon_{\mathrm{bas}}(x)\right)\,, \tag{3.46}$$

being

$$\varepsilon_{\mathrm{bas}}(x) = \frac{a_{\mathrm{sub}} - a(x)}{a(x)} \tag{3.47}$$

the "*basal strain*", which depends on the substrate lattice parameter a_{sub}. Fiorentini and coworkers finally found that

$$\mathbf{P}_{\mathrm{pe},\mathrm{InN}}(x) = -1.373\,\varepsilon_{\mathrm{bas}}(x) + 5.264\,\varepsilon_{\mathrm{bas}}(x)^2\,, \tag{3.48}$$

$$\mathbf{P}_{\mathrm{pe},\mathrm{AlN}}(x) = \begin{cases} -1.808\varepsilon_{\mathrm{bas}}(x) + 5.624\,\varepsilon_{\mathrm{bas}}(x)^2 & \text{for} \quad \varepsilon_{\mathrm{bas}} < 0 \\ -1.808\varepsilon_{\mathrm{bas}}(x) - 7.888\,\varepsilon_{\mathrm{bas}}(x)^2 & \text{for} \quad \varepsilon_{\mathrm{bas}} > 0 \end{cases} \tag{3.49}$$

and

$$\mathbf{P}_{\mathrm{pe},\mathrm{GaN}} = -1.373\,\varepsilon_{\mathrm{bas}}(x) + 7.559\,\varepsilon_{\mathrm{bas}}(x)^2\,. \tag{3.50}$$

By using Eq. (3.43) within Eq. (3.47) and substituting the latter in Eqs. (3.48)–(3.50) and, in turn, into Eq. (3.46) one can obtain the overall formalism known as *Fiorentini-Bernardini model* [111,112] of polarization charges, that we implemented in our numerical simulations as described here.

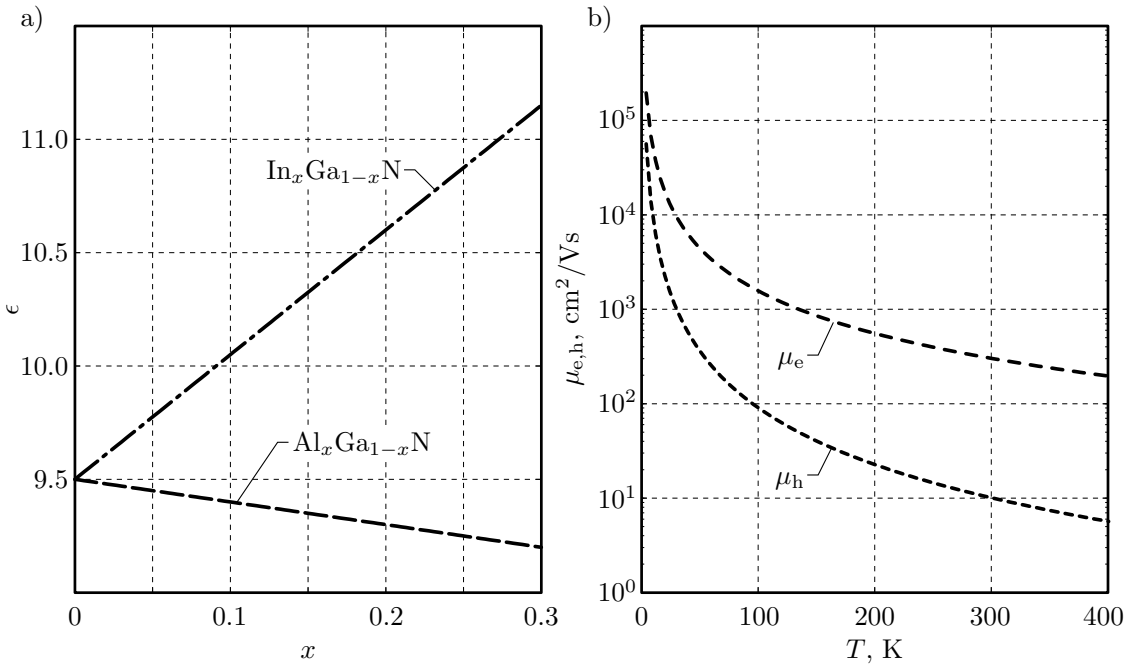

FIGURE 3.10. (a) Dielectric constant in $In_xGa_{1-x}N$ and $Al_xGa_{1-x}N$ alloys as a function of x, as stated by Eq. (3.40). (b) Temperature-dependent electron and hole mobility in GaN, InGaN and AlGaN calculated with Eq. (3.52).

Once the total polarization vector has been assembled, one can compute the polarization charge density (cm^{-2}) at a given surface or interface A as

$$\sigma_{\text{pol}} = -\frac{\text{div}\{\mathbf{P}\}}{A}\,. \tag{3.51}$$

As we will see talking about LEDs performance, σ_{pol} represents a crucial parameter in the electrical characterization of the diode. Furthermore Eq. (3.51) implies that, chosen an arbitrary surface A at given depth of a bulk III-nitride sample, a certain charge density σ_{pol} can be found depending on the local composition/symmetry of lattice sites intercepting the plane of the surface itself.

Due to its enormous potentiality, most of the possible applications of GaN remain still largely unexplored, nowadays, as emerged during a pleasant conversation I had the great honor to have with the Nobel Laureate Hiroshi Amano at the meeting "GaN Marathon 2016" in Padua. One of these areas of interest, in the author's opinion, could exactly involve the polarization properties we just discussed. Let us briefly see how. As pointed out in several works, GaN and its ternary compounds with indium and aluminum exhibit a particular robustness against particles and ionizing radiation damages (see, for instance, Refs. [113–115]). Besides this, P. Kivisaary *et al.* [116] have recently proved that, by combining the effect of intrinsic GaN polarization with the x-dependent piezoelectric charge induced in ternary alloys, an effective charge density can be produced. Such charge acts as a traditional impurity-based chemical doping (reaching 3D densities up to 10^{18} cm^{-3}) which can

be managed to produce true density profiles simply by exploiting compositional grading of indium or aluminum, both of n- and p-type (see also Ref. [117]). This feature could, in principle, overcome doping removal issues at high radiation fluences typical of Si-based high-energy particle detectors. For these reasons, after lighting and power electronics, the family of III-V nitrides could find a new application framework in high-energy physics, as already claimed by some authors [118, 119].

We continue our digression about III-N material properties introducing now some transport-related parameters. An important member of this class is without doubts the carrier mobility. In accordance with Ref. [120] electron and (heavy) hole mobilities ($\mathrm{cm^2V^{-1}s^{-1}}$) have been modeled as an exponential function of T:

$$\mu_{\mathrm{e,h}}(T) = \mu_{\mathrm{e,h}}(300\ \mathrm{K}) \left(\frac{T}{300\ \mathrm{K}} \right)^{\gamma_{\mathrm{e,h}}} , \tag{3.52}$$

where the values $\mu_{\mathrm{e,h}}(300\ \mathrm{K})$ have been chosen in accordance with actual literature (see Refs. [120–123]), i.e. $\mu_{\mathrm{e}}(300\ \mathrm{K}) = 300\ \mathrm{cm^2V^{-1}s^{-1}}$, $\mu_{\mathrm{h}}(300\ \mathrm{K}) = 10\ \mathrm{cm^2V^{-1}s^{-1}}$, $\gamma_{\mathrm{e}} = -1.5$ and $\gamma_{\mathrm{h}} = -2$ (see panel (b) of Figure 3.10).

Regarding GR mechanisms in III-nitrides, this work assumes parameters according to Ref. [120], i.e. $\tau_{\mathrm{SRH}} = 5 \cdot 10^{-8}$ s both for GaN and InGaN while $\tau_{\mathrm{SRH}} = 10^{-7}$ s in AlGaN. Radiative and Auger coefficients have been respectively adopted as $B = 2 \cdot 10^{-11}\ \mathrm{cm^3/s}$ and $C = 2 \cdot 10^{-30}\ \mathrm{cm^6/s}$ for binary and ternary compounds. Although several authors claim the x-dependence of these parameters (see, for instance, Ref. [124]), in our simulations and on the particular devices we will analyze this choice revealed to be sufficiently accurate, yielding values pretty consistent with those coming from such composition-dependent approaches.

3.2.4 Doping and defects

Impurities in III-nitride materials can be usually represented by chemical dopants or by native defects. As already pointed out in the introductive part of this chapter, p-type doping has been a challenge for a long time, at least until Amano's contributions in the 80s of the last century. This is a common issue – as well as any doping type – in wide band gap semiconductors. Besides chemical doping, induced through implantation processes, native defects can modify the material conductivity in both p- and n-type sense, acting with or against traditional dopants (where the latter is often called *compensation* effect).

Among the most common native defects one may find vacancies (way of example, nitrogen vacancy behaves like a donor), antisites or self-interstitials [125]. Moreover, also incorporated species, if not prevented during growth process, can act as effective dopants. This is the case of oxygen or silicon, which can form shallow donor levels. For this reason GaN and its alloys can be easily negatively doped.

For what concerns positive doping (p-type) of GaN, InGaN and AlGaN the element of choice is without doubts Mg. Even if this solution is now well established, it still remains challenging to reach high hole concentrations by Mg doping. This occurs because the reaction forming interstitial or N-substitutional magnesium, which generates acceptors, is less favorable than the competitive formation of

Ga-substitutionals. Furthermore, the above mentioned compensation by nitrogen vacancies (V_N) has the effect of mitigating the positive doping of Mg acceptors, especially in AlGaN alloys at high x_{Al}.

Negative (n-type) doping, instead, is commonly obtained with C, Si or Ge when used as substitutionals on cation sites. C, as well as Si, has been demonstrated to form shallow donor levels [126]. Nevertheless it can also act as shallow acceptor if on the N site [127].

By using Eq. (1.27), the following doping ionization energies have been adopted: $E_{D,ion} = 20$ meV for donors in GaN, InGaN and AlGaN and $E_{A,ion} = 199$ meV, 170 meV and 210 meV for acceptor levels in GaN, InGaN and AlGaN respectively (with concentrations in the range $0.15 \leq x_{In,Al} \leq 0.18$). Notice that, as widely discussed in literature, the Mg ionization energy is roughly proportional to the mole fraction x [128]: for instance, the AlGaN $E_{A,ion}$ ranges from 170 meV to 530 meV when x increases from 0 to 1 [129–132].

3.3 Operation Principles and Efficiency Problem

This conclusive section presents some important details concerning the physics behind light-emitters. After recalling the building-blocks of photons production, in Section 3.3.1 quantum confinement of carriers, needed for localizing radiative processes, will be introduced and explained. On the basis of this discussion, the framework regarding LED's efficiency will be introduced. So, in Section 3.3.2 all the concepts mentioned about GR and transport mechanisms will be applied to describe the standard regime of carriers flow and leakage. This will give us the right perspective to understand the meaning of quantum efficiency and the role of some relevant mechanisms which participate (concurring or competing) to the light-to-signal conversion.

3.3.1 Carrier confinement in QWs and light emission

As we learned by describing light absorption in MCT-based IRPDs, the power dissipation of an electromagnetic propagating wave within a solid medium follows the exponential-like Lambert's law. In order to optimize the optical absorption, photodetectors have to be designed in such a way that the region where the absorption and the subsequential photogeneration occur lies at an optimum depth from the illuminated facet. This allows to concentrate most of the electron/hole generation in a restricted volume: the active region. Specularly, light-emitters need to do the same and, since it would be no control in the propagation of photons emitted from anywhere in the device, LED structure has to be engineered to confine optical processes in an "optimal" region. Besides these macroscopical reasons, there is another – more subtle – motivation leading to clearly define the LED active region: being a wide band gap material, GaN would not be the best choice in generating monochromatic blue light (see Figure 3.5 for a fast reference). From these considerations the need for a specific region with optimized energy gap arises. To this purpose LEDs

are designed with one or more very thin layers (few nm) of narrower gap materials (in our case, InGaN) alternating with GaN which form band discontinuities and act as confining potentials for radiatively recombinating electron/hole pairs: the so-called quantum wells (QWs). This feature not only assures a geometrical control of photoemission but also introduces a proper filter for the emitted wavelength, being the energy gap the decisive tuner allowing to select a specific line in the emission spectrum.

In principle, the photons so generated should have all the same energy. In real life, they don't. Besides uncontrolled compositional fluctuations, which induce energy gap oscillations within the active region, some microscopic or quantum effects intervene. First of all, growing thin InGaN layers interleaved with GaN can be very critical by the mismatch standpoint, especially for high indium fractions. So, lattice strains may generate band deformations or energy gap deviations from standard trends. Second, the band structure of a bulk material (as described in Section 3.2.2) may significantly differ from that of a QW because the radius of potential interactions between different layers of different alloys become comparable with the dimension of the unit cell and its reciprocal representation. On the other hand, quantum confinement could be weak enough to generate deviation from the bulk conditions.

Apart from these considerations, to understand what regime characterizes electrons and holes in QWs one has to recall the behavior of fermions in a potential well. In order to achieve a full comprehension of this physical system the tool which comes to our aid is the well-known time-independent (1D) Schrödinger equation

$$-\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial z^2}\Psi_n(z) + U(z)\Psi_n(z) = E_n\Psi_n(z)\,, \tag{3.53}$$

that we wrote here at the FBZ center and along the confining direction z.

By solving this eigenvalues equation for the potential well we find a discrete set of solutions, represented by a certain number of eigenenergies E_n coupled with Bloch eigenfunctions $\Psi_n(z)$, being z the direction perpendicular to the InGaN/GaN interfaces (which usually corresponds to the growth c-axis). The number N of solutions depends on the well width L_{QW} and on its depth, i.e. the amount of potential discontinuity at the QW boundary interfaces. The simplest (ideal) case is represented by an infinite square well. In these conditions one finds that all electron/hole wavefunctions $\Psi_n(z)$ are pinched at the well boundaries and are described by standing waves. This means that all carriers are perfectly confined there because the probability density $|\Psi_n(z)|^2$ of those states is null outside.

In Figure 3.11 a more realistic case is represented. In panel (a) we find a schematic picture of a single quantum well (SQW) consisting of a double GaN/InGaN heterojunction. The electron potential $U_{\mathrm{e}}(z) = E(z)/q$ along the z-direction experiences a discontinuity which, at equilibrium, generates a finite square well where electrons are quasi-confined. We expressly said *quasi* confined because, as one may notice in panel (b), non null wavefunction tails propagate in the forbidden region along the z coordinate. This result is a direct consequence of the Schrödinger eigenvalue equation which allows full confinement only in the ideal case of a potential well with infinite depth. Then, for the same system, in panel (c) the 2D electron

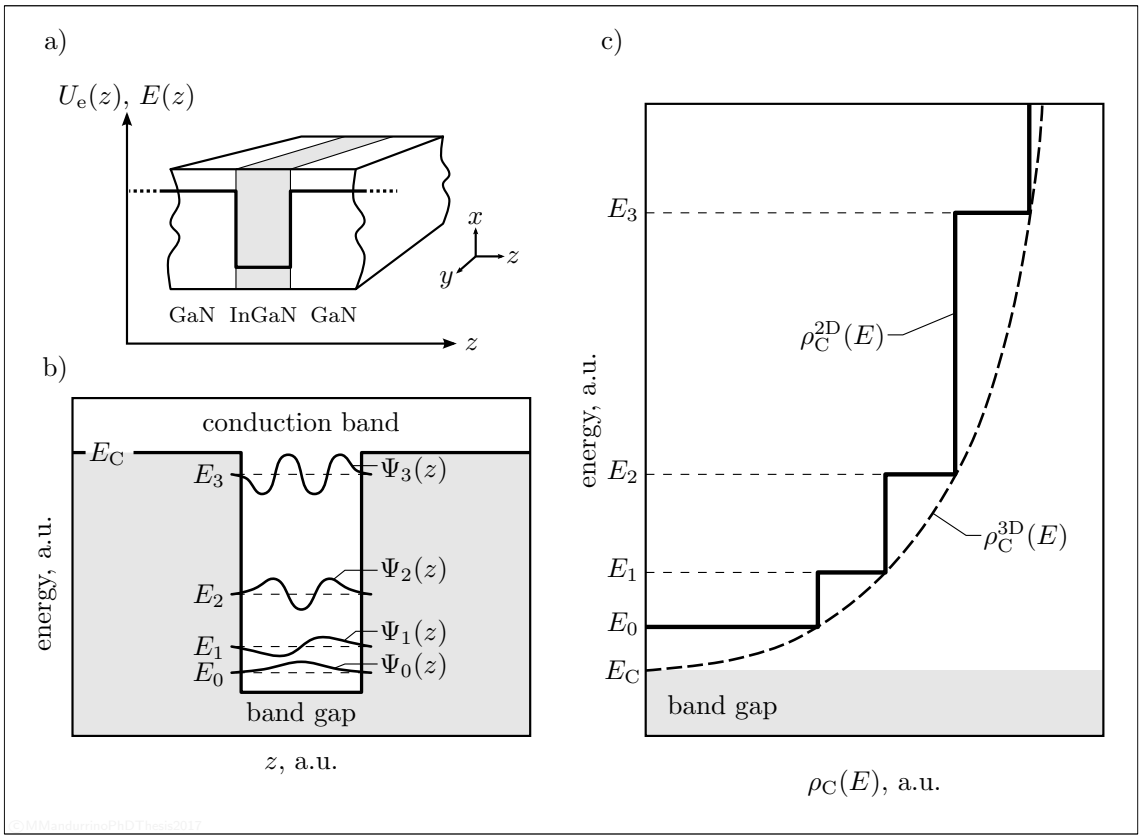

FIGURE 3.11. (a) Schematic picture representing an InGaN/GaN quantum well (QW). Also the electron potential (or energy) is represented. (b) Electron eigenenergies E_n and Bloch eigenfunctions $\Psi_n(z)$ in a finite square potential well, as the one generated by the structure in (a). The same occurs in valence band of quasi-confined holes. (c) Density of states (DOS) for confined electron sub-bands compared with bulk DOS.

density of states

$$\rho_C^{2D}(E) = \frac{m_e^*(E_n)}{\pi\hbar} \tag{3.54}$$

is plotted and compared with the standard 3D (bulk) DOS. Similar definition holds for holes. From this picture five important considerations arise:

1. Bloch wavefunctions are not well-confined. Nonetheless, in the scenario here depicted we can assume a full confinement since conduction band do not offer any available state for those quasi-bound electrons. Anyway, very often occurs that, out of equilibrium, band bending or in-gap impurities offer to these states a finite probability of tunneling from one of the two surrounding barriers towards the well.

2. Quantum confinement (although partial) acts only along the z-direction. This means that electrons and holes in QW can be safely considered particles in a "free" motion regime within the xy-plane (if we are under parabolic band and effective mass approximation).

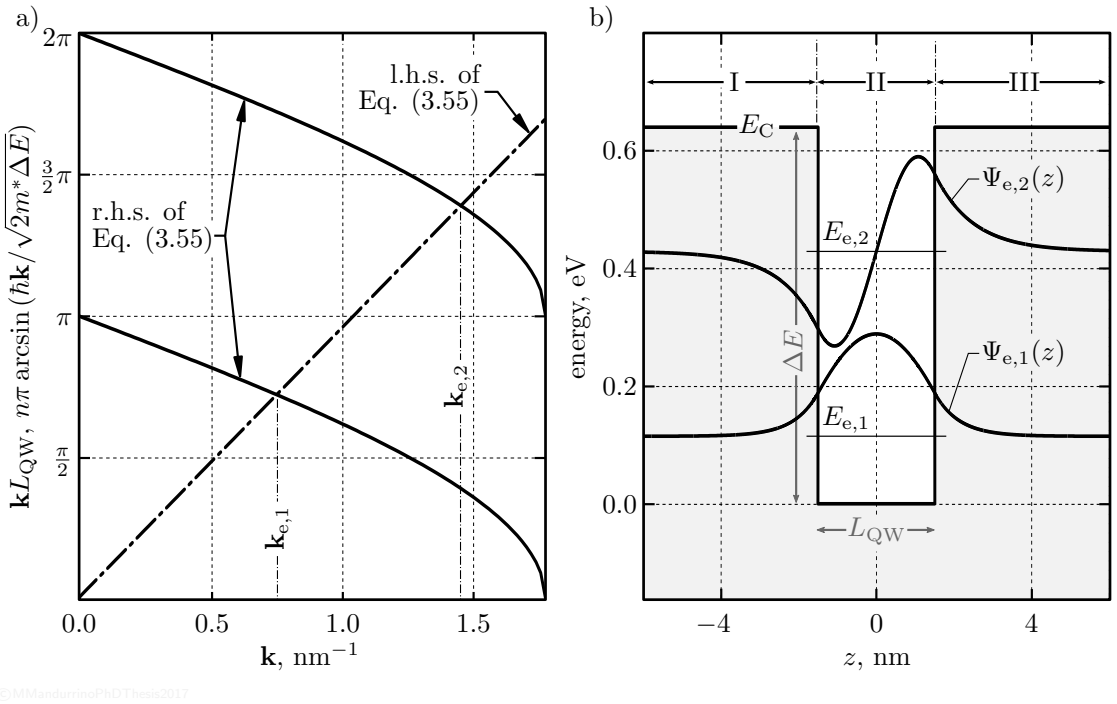

FIGURE 3.12. (a) Graphical solution of the electron Schrödinger equation for the InGaN/GaN structure of panel (b) implemented via a numerical routine which solves the trascendental system written in Eq. (3.55). (b) Eigenvalues and (properly normalized) eigenfunctions calculated with Eqs. (3.55) and (3.56). Notice the penetration of wavefunctions into the barriers I and III.

3. The ground-state energy E_0 does not correspond to the conduction band minimum E_{C} but is higher (viceversa for holes).

4. As noticeable in (b) and (c), the energy spacing between quantized levels is increasing with the increase of E, a feature deriving from the particular shape of the well.

5. Notice the wavefunctions parity trend with varying the principal quantum number n, which identifies the state. The first eigenfunction has no zeros and is an odd function, while the second has one zero and is an even function, and so on. Thus, n indicates also parity and number of zeros. Note: here we does not represented the spin degeneracy but, being fermions, it is obvious that, according to the Pauli exclusion principle, for each level E_n there are two electrons with opposite spins satisfying the Schrödinger equation.

Supposing to solve the Schrödinger problem for a finite square well, it can be demonstrated that (see, for instance, Ref. [133]) the energy spectrum is given by

$$\mathbf{k}L_{\mathrm{QW}} = n\pi - 2\arcsin\left(\frac{\hbar\mathbf{k}}{\sqrt{2m^*\Delta E}}\right), \quad \text{with} \quad n \geq 1\,, \tag{3.55}$$

where ΔE is the energy offset at the heterojunction measured from the bottom of the QW. By solving Eq. (3.55) one obtains a discrete set of $\mathbf{k}$ to be inserted into

wavefunctions of the form

$$\begin{cases} \Psi_{\mathrm{I}}(z) = C_1 \mathrm{e}^{\mathbf{k}z} \\ \Psi_{\mathrm{II}}(z) = C_2 \sin(\mathbf{k}z + \phi) \\ \Psi_{\mathrm{III}}(z) = C_3 \mathrm{e}^{-\mathbf{k}z} \end{cases}, \tag{3.56}$$

valid, respectively, for regions I, II and III (see Figure 3.12), where coefficients $C_{1,2,3}$ have to be determined via normalization and where $\sin\phi = \hbar\mathbf{k}/\sqrt{2m^*\Delta E}$.

By virtue of energy (or momentum) conservation, the photon emitted via spontaneous radiative recombination must have the momentum equal to the difference between the momenta of the electron and hole that recombined. Moreover, the photon momentum is some orders of magnitude lower than electron momentum, leading to quasi-vertical optical transitions. However, this does not exclude that vertical transitions at $\mathbf{k} \neq 0$ can take place (as we already shown in Figure 1.7) and, on the other hand, $\rho^{2\mathrm{D}}(E)$ is such that that high-energy sub-bands can be populated. Combining carriers statistics and their DOS, rough estimations of LED emission lineshapes can be obtained. In a very simple, case invoking Boltzmann statistics, the emission peak corresponds to an energy of about $E_{\mathrm{g}} + k_{\mathrm{B}}T/2$ with FWHM $\simeq 1.8\ k_{\mathrm{B}}T$.

Besides these effects related to the energy conservation, we have to consider also that neither electrons nor holes have ground-state energies equal to their respecting band minimum/maximum. This induces a wavelength shift in the emission spectrum. Indeed, if choosing a particular semiconductor for light emission depends on its E_{g} – since, neglecting thermal energy ($k_{\mathrm{B}}T \ll E_{\mathrm{g}}$), we have $E_{\mathrm{ph}} \simeq E_{\mathrm{g}}$ – this quantum effect has the consequence to induce an *effective energy gap* which is, in general, wider than bulk one (see Figure 3.13). Moreover, the influence of strains, polarization charges and external fields (as the one induced by an applied bias) can model the well shape out of equilibrium such as to enable other static effects which act on the emission spectrum. In Figure 3.13, for instance, one of these has been represented: accounting for band non-idealities, confined (2D) electron and hole density profiles may assume asymmetric distributions with respect to the well such that higher densities are found in correspondence of its deeper region (where $|E_{\mathrm{C,V}}|$ in the QW assumes lower values). So, this asymmetry can furtherly shift the emission lineshape by means of modulating the energy distribution of spontaneous radiative recombination.

Not only static phenomena may occur. In fact, when the LED is subjected to a rising applied voltage more and more carriers are injected into its active region. This increases the sub-bands occupancy whithin the well and, in turn, the separation between quasi-Fermi levels. This produces a blue-shift of the emission peak.

In Figure 3.13 we can see some hypothetical spontaneous radiative recombination processes occurring in an InGaN quantum well. Notice that, perpendicular to the band diagram, parabolic approximations of energy dispersion $E_n(\mathbf{k})$ have been drown. Moreover, also polarization is included, since E_{C} and E_{V} are tilted along the QW because of the presence of a polarization field. In case (1) we have the most simple recombination process, involving ground electron and hole quantized levels

in Γ (notice that $E_{\mathrm{e},0} - E_{0,\mathrm{h}} > E_{\mathrm{g}}$). Then, in cases (2) and (3) transitions are still vertical (no momentum exchange) but they occur at $\mathbf{k}' > 0$ and $\mathbf{k}'' > \mathbf{k}'$ respectively. These – less probable, according to the *selection rule* – transitions can contribute to the emission lineshape broadening. It is also worth stressing that representing confined eigenenergies $E_{\mathrm{e},0}$, $E_{\mathrm{e},1}$, $E_{\mathrm{h},0}$ and $E_{\mathrm{h},1}$ as single-line levels does not account for any degeneracy removal effect: first, no field-related spin splitting has been included, then nor any carrier-carrier or carrier-phonon scattering is accounted for (we

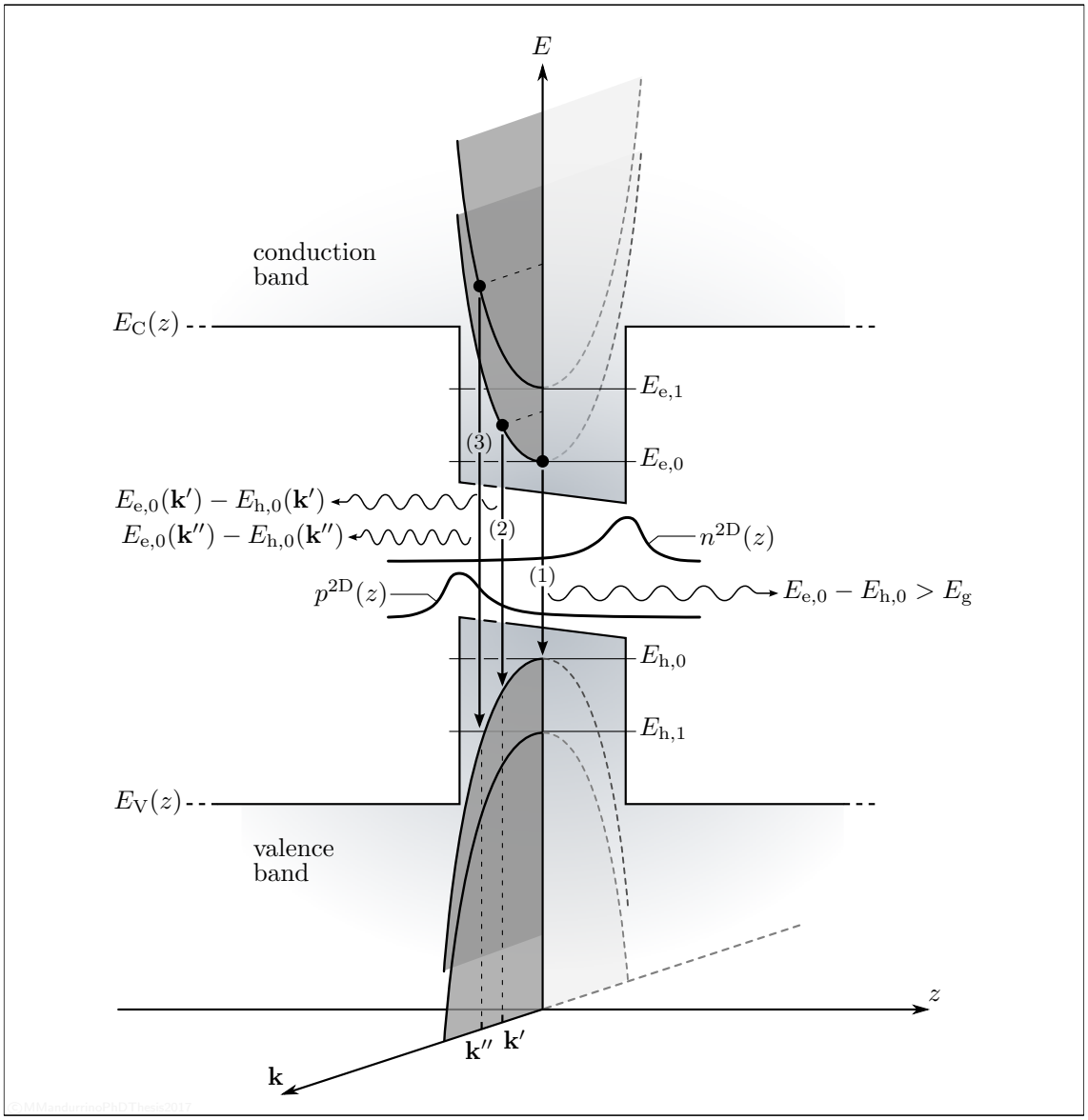

FIGURE 3.13. Spontaneous radiative recombination in a quantum well (QW), schematically represented in the band diagram space E versus z and, at the same time, developed for small $\mathbf{k}$ deviations from the FBZ center (see parabolas lying on the plane perpendicular to the band diagram). Three examples of emission processes are represented: (1) at the FBZ center ($\mathbf{k} = 0$); (2) between ground electron and hole states at $\mathbf{k}' > 0$; (3) between second electron level and first hole level at $\mathbf{k}'' > \mathbf{k}'$. Also confined carrier densities $n^{2\mathrm{D}}(z)$ and $p^{2\mathrm{D}}(z)$ have been plotted.

are in the so-called *ballistic regime*). All these effects, otherwise, would result in a broadening of each quantized level.

Finally, also notice how quantized (2D) electron and hole densities $n^{2D}(z)$ and $p^{2D}(z)$ does distribute along the confining z-direction: they depend on the sub-band carrier wavefunctions via their probability density as

$$n^{2D}(z) = \frac{k_B T}{\hbar L_{QW}} \sum_i |\Psi_{e,i}(z)|^2 \, \rho_C^{2D}(E_{e,i}) \, \ln\left[1 + e^{\frac{E_{Fn}(z) - E_{e,i}}{k_B T}}\right] \tag{3.57}$$

and

$$p^{2D}(z) = \frac{k_B T}{\hbar L_{QW}} \sum_i |\Psi_{h,i}(z)|^2 \, \rho_V^{2D}(E_{h,i}) \, \ln\left[1 + e^{\frac{E_{h,i} - E_{Fp}(z)}{k_B T}}\right] \tag{3.58}$$

(where i labels each sub-band and L_{QW} is the well width) thus, as predicted by the Schrödinger equation, peaks lies where the QW shape reaches deeper energy since carriers are more attracted there.

In order to better control carrier confinement and, thus, the radiative recombination, modern LEDs are designed with a certain number of quantum wells in sequence, a structure that is called multiple quantum well (MQW). This expedient also allows to obtain more than one emitted wavelength by using wells with different indium mole fraction [134–137].

3.3.2 Electronic transport and quantum efficiency

Quantum wells constitute only one part of an LED structure, although being the most important. As the name suggests, LEDs are essentially based on a p-n diode, which normally operates under forward-bias conditions. In particular, since the InGaN-based active region is usually lightly doped or not doped at all, we should more properly speak of p-i-n junction. The model best describing its dynamical behavior is the already mentioned Shockley diode equation, but in its simplified version valid under forward bias regime (and for $V_{bias} \gg k_B T/q$) and given by

$$I_{fwd}(V_{bias}) = I_0 \, \exp\left(\frac{q V_{bias}}{\eta k_B T}\right), \tag{3.59}$$

where I_0 is the diode saturation current written in Eq. (2.40), q is the electron elementary charge and

$$\eta = \frac{q}{k_B T} \left(\frac{\partial\{I_{fwd}(V_{bias})\}}{\partial V_{bias}}\right)^{-1} \tag{3.60}$$

here represents the so-called *diode ideality factor*, a macroscopic parameter which collects different physical mechanisms (GR, tunneling, ...) acting on the slope of the forward $I(V)$ diode characteristics. Theoretically η should be equal to 1, but experimentally this is almost never achieved and thus the ideality factor is very often greater than unity – up to several units – depending on the phenomenon which produces deviations from the ideal trend. In this sense deducing η through

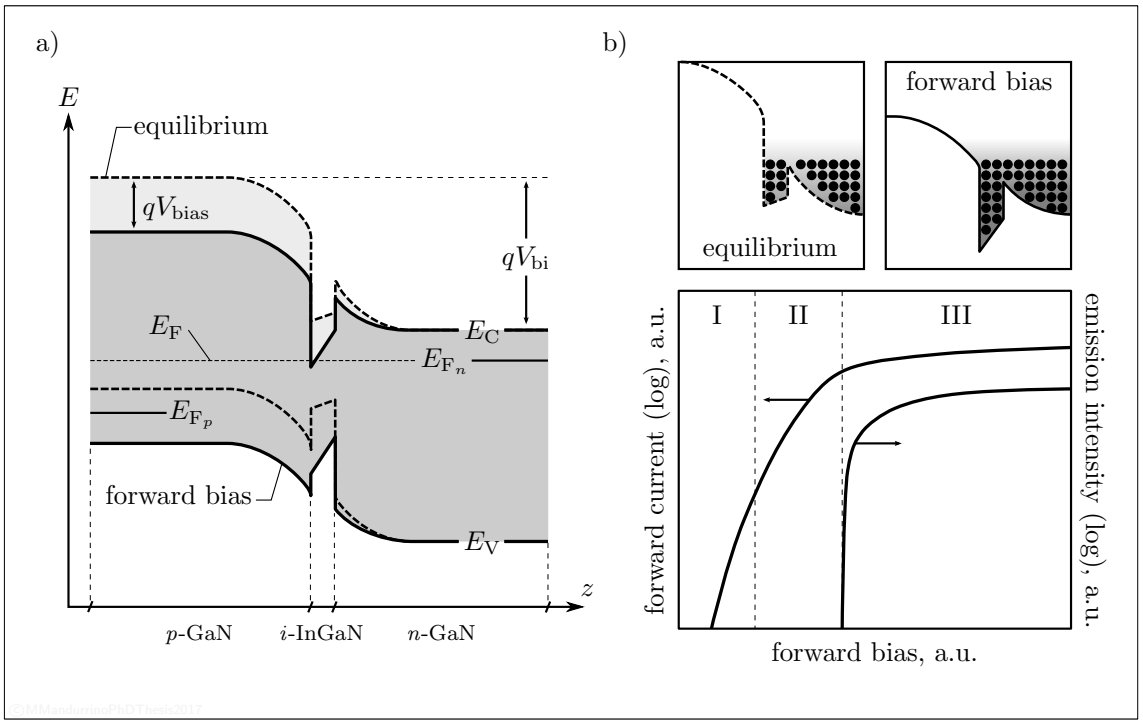

FIGURE 3.14. (a) Band diagram of a single quantum well (SQW) LED structure at equilibrium and under forward bias conditions. The most important energies have been highlighted, as Fermi energy E_{F} (at equilibrium), quasi-Fermi levels $E_{\text{F}_{n,p}}$ (out of equilibrium), the built-in energy qV_{bi} (where V_{bi} represents the built-in junction potential) and the bias-dependent band bending qV_{bias}. (b) In the upper part: simplified QW electron filling at equilibrium and forward bias, showing the carrier density increase for increasing bias. In the lower part: $I(V)$ and output power characteristics. Regimes "I", "II" and "III" are described in the text below.

the $I(V)$ analysis can be used as a tool to identify the fingerprint of specific leakage mechanisms.

Panel (a) of Figure 3.14 represents both equilibrium and forward bias band diagram of a generic SQW InGaN/GaN blue LED. The diffusion-type transport of standard *p*-*n* junctions here is heavily modified by reason of quantum confinement just discussed, which limit the radiative recombination no more to the usual diffusion lengths $L_{\text{n,p}}$ but only to the active region width (in this case the well width L_{QW}). As a consequence carriers in the well reach higher densties than in usual diodes and, in turn, radiative recombination is stronger since R_{rad} is roughly proportional to the product pn. As the bias goes up – see upper sub-figures of panel (b) – carrier density within the active region furtherly increases until radiative mechanisms become dominant (the same trend holds for valence band, here not depicted).

Taking a look to the lower sub-figure, one can observe that the electric characteristics experience three different regimes. In "I" the transport is mainly affected by series resistance and SRH processes. Here the slope suffers from the presence of high defect densities and non-radiative recombination occurring at SRH centers. Then in regime "II" SRH is still present but also defect-related tunneling processes, which require higher electric fields than simple non-radiative transitions, begin to

be more and more relevant with the bias increase. Finally in "III", when the onset of radiative mechanisms occurs (the so-called *optical turn-on*) spontaneous emission definitely drives the characteristics. Besides radiative recombination also Auger mechanisms are heavily carrier density dependent, as we already stated in this work. Indeed, at very high injection regimes (here not represented) three-particle mechanisms intervene such that the radiative one begins to experience carrier losses and, consequently, efficiency decrease.

In an LED we can define two types of efficiency. Accounting only for microscopic radiative mechanisms, we find the internal quantum efficiency IQE given by the ratio between the number of photons emitted from the QW per unit time and the number of electrons injected into the whole device per second. So, symmetrically to what we said for IR photodetectors:

$$\eta_{\text{int}}^{\text{LED}} \equiv \text{IQE} = \frac{qP_{\text{int}}}{I_{\text{inj}}\hbar\nu}\,, \tag{3.61}$$

where P_{int} is the optical power generated in the active region and I_{inj} the injected current. If we include also the absorption properties of GaN, which partially absorbs generated photons, one has to define an overall external efficiency (EQE). This quantity considers the capability to extract the light from the device to the surrounding free space and makes use of a new quantity called extraction efficiency, which is the ratio between the number of photons propagating outside per second and the number of photons emitted by the active region per second. Thus, the EQE is given by:

$$\begin{aligned} \eta_{\text{ext}}^{\text{LED}} \equiv \text{EQE} &= \eta_{\text{extraction}}^{\text{LED}} \cdot \text{IQE} \\ &= \frac{P_{\text{out}}}{P_{\text{int}}} \cdot \frac{qP_{\text{int}}}{I_{\text{inj}}\hbar\nu} \\ &= \frac{qP_{\text{out}}}{I_{\text{inj}}\hbar\nu}\,, \end{aligned} \tag{3.62}$$

where $P_{\text{out}} = N_{\text{out}}\hbar\nu$, being N_{out} the number of photons emitted into free space per unit time.

As pointed out by G. Verzellesi in Ref. [123], if only radiative processes take place then the IQE would be 1, and all injected carriers would produce photons. In this case $I_{\text{inj}} = V_{\text{QW}} q B_{\text{rad}} n^2$ and Eq. (3.61) becomes

$$\text{IQE} = \frac{V_{\text{QW}} B_{\text{rad}} n^2}{I_{\text{inj}}/q}\,, \tag{3.63}$$

where V_{QW} is the QW active volume. When other GR mechanisms participate and compete to the LED operation, the IQE must be re-written as

$$\text{IQE} = \frac{B_{\text{rad}} n^2}{A_{\text{SRH}} n + B_{\text{rad}} n^2 + C_{\text{Aug}} n^3}\,, \tag{3.64}$$

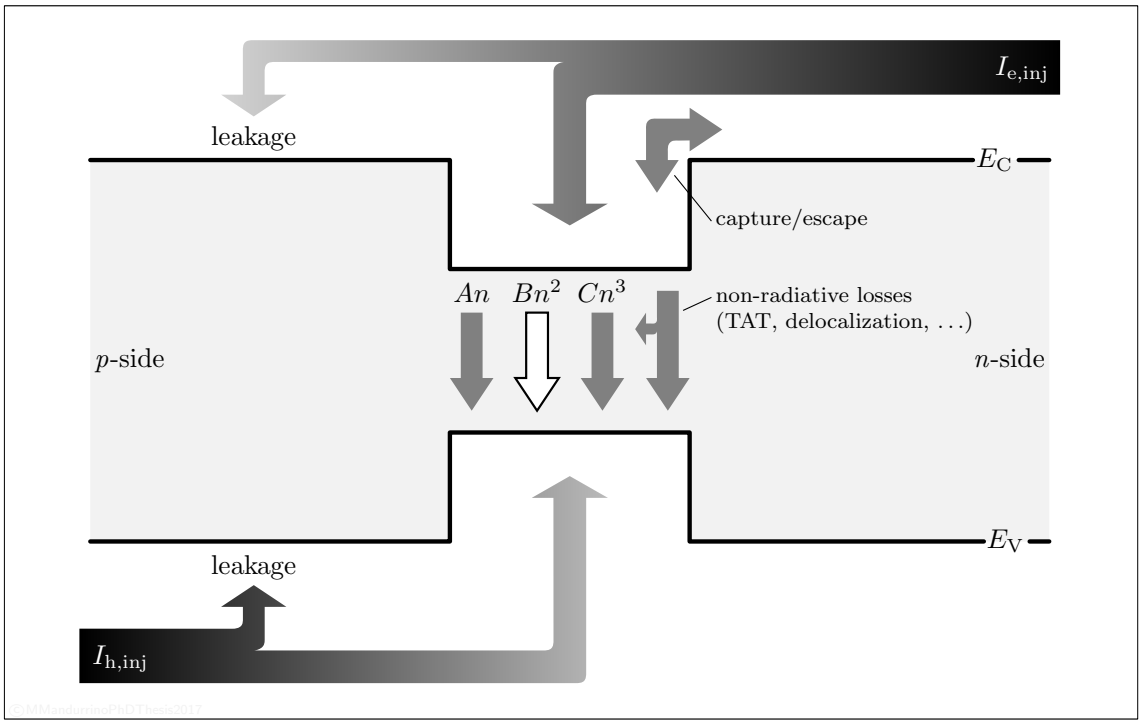

FIGURE 3.15. Mechanisms competing with spontaneous radiative recombination in a generic double heterojunction (DH) single quantum well (SQW) LED structure (adapted and re-edited from Ref. [123]).

in which $A_{SRH} = 1/\tau_{SRH}$ (in s^{-1}). Here we account only for SRH, radiative and Auger processes. If we include also leakage, escape mechanisms from QW, tunneling or carrier delocalization the order of n in the denominator would increase thanks to additional summands.

As schematically depicted in Figure 3.15, several processes concur to the net balance between injected carriers and emitted photons. On the one hand only $B_{rad}n^2$ contributes to the light emission, on the other hand there are several mechanisms undermining this process. As we said, each of them has a different field- and carrier-dependence, resulting in a marked dominance within a particular bias regime. At low injection $A_{SRH}n$ term is the most relevant leakage channel. Nonetheless, high defect densities can affect the LED performance also at higher bias and may also assist tunneling-related inter-band transitions. Other material features, like carrier delocalization, or mechanisms, as in case of QW carrier capture/escape, can contribute to decrease the radiative efficiency. In particular, the latter represents an interesting example of intrinsic quantum effect due to the finite probability to have excited or relaxing carriers traveling via phonon-assisted transitions across the heterojunctions and between energy sub-bands belonging to states in the well and in the surrounding barrier.

The process most indicated as the main competing with radiative recombination and among the most debated causes for the efficiency droop is Auger. Direct Auger transitions have a strong temperature dependence and are characterized by a

threshold onset energy [138]. Theoretical calculations point out that expected values for the coefficient C_{Aug} in InGaN QWs range from 10^{-32} (or less) to 10^{-30} cm^6s^{-1}. Indirect (phonon-assisted) Auger transitions instead have no threshold and may affect the performance also at low temperature and even at room-T, with a power-law dependence [139]. In this case values in the order of 10^{-31} cm^6s^{-1} [140] have been theoretically obtained. In general, Auger coefficients exhibit enhanced values in QWs with respect to bulk materials due to the $\mathbf{k}_z$-conservation removal at the heterojunctions. However it is commonly accepted that already since $C_{\text{Aug}} \sim 10^{-31}$ cm^6s^{-1} the Auger influence on droop becomes appreciable [141–143].

Among the candidates responsible for droop we find also trap-assisted tunneling (TAT). In particular N. I. Bochkareva *et al.* [89] claimed that *horizontal* processes from the well towards defects in the p-side barrier are correlated with an IQE decrease in SQW InGaN/GaN structures.

Part II

Tunneling in Direct Band Gap Semiconductor Devices

Chapter 4

Tunneling: From Quantum Theory to Modeling

As we have seen in the case of quantum wells (QWs), when one or more spatial dimensions are remarkably reduced every classical description of fermions becomes completely unreliable. This is the reason why, for instance, Bloch states are unavoidable in describing quasi-confined electrons and holes in LEDs active region. Moreover such a topic is even more crucial because of the actual trend of shrinking devices size in all electronics application fields.

However, not only system dimensions do affect our description of carriers. As a matter of fact, junctions made of narrow gap semiconductors may produce bands bending such that the path between two states in conduction and valence band facing each other on either side of the gap could be comparable with the electron/hole wavelength. No classical framework could never account for quantum effects occurring in such a case, unless suitable corrections are adopted.

This theoretical chapter is devoted to describing one of the most important, intrinsic, unavoidable and common quantum effect in solid-state applications: quantum tunneling. Starting from basic theory we will come to its description suitable for numerical modeling as it has been developed for the specific devices introduced in Part I. With the aim of being as much exhaustive as possible, the present work took into consideration different tunneling mechanisms, within different devices and under different modeling techniques. Thus, the subject is organized as follows. In Section 4.1 we propose a physical review about quantum origins of tunneling through the semiclassical formalism and under the envelope wavefunctions approximation. Here the simplest direct band-to-band tunneling (BTBT) and trap-assisted tunneling (TAT) mechanisms are described. Furthermore, in Section 4.2, the reader will also come across some author's original results about BTBT modeling in *p*-*i*-*n* MCT photodetectors, which found their theoretical origin in my Master's thesis [45] and which have been fully developed during the Doctorate research activity. Thanks to

a proficuous collaboration between our Department of Electronics and Telecommunications and AIM Infrarot-Module GmbH, Theresienstraße 2, D-74072 Heilbronn, Germany, a leader company in the production of IR detectors, such novel formalism has been applied for the first time in the literature to real device simulations, as we will discuss in Chapter 5.

Section 4.3 deals with a detailed description about the physical mechanism of multi-phonon emission (MPE), which is at the basis of defect-assisted tunneling transitions, a process not so relevant in IR photodetectors but absolutely crucial in wide gap GaN-based light emitters, as we will demonstrate. A compact model for trap-assisted tunneling (TAT) in these devices is derived here from solid-state principles. This description, moreover, can be found as a theoretical introduction of the invited paper [144] we published on a Special Issue of the *Journal of Computational Electronics* (see Ref. [145]), thanks to the opportunity given by its then Editor in Chief J. Piprek, and represents – to the best of our knowledge – the first successful example of such an MPE-based formalism applied to InGaN/GaN blue LEDs simulations. All this part of work has been possible especially thanks to the collaborations with OSRAM Opto Semiconductors GmbH, Leibnizstraße 4, 93055 Regensburg, Germany, and with colleagues from the Information Engineering Department of Padua University, who provided us all the experimental characterizations which have been essential to test our theoretical models on real LED structures.

Then, after analyzing BTBT and TAT through the semiclassical perspective, Section 4.4 introduces a full-quantum methodology, more and more employed in electronic and – also in some still rare cases [146–148] – optoelectronic devices modeling. The treatment concerning Non-Equilibrium Green's Functions (NEGFs) will conclude this theoretical chapter, before leaving space to the application of what developed here to real device simulations, the main topic of the next two chapters.

4.1 Tunneling and WKB Approximation

Here we explore the semiclassical approximation of quantum mechanics, which is at the basis of almost all tunneling analytical descriptions, both for direct (BTBT) and indirect (TAT) processes. In this section also some relevant examples of tunneling current and rate expressions, that are based on this theory, will be provided to the reader.

Let us start with describing the tunneling effect *per se*. We already seen that the Bloch wavevector must be purely imaginary within the forbidden band gap, so that $\mathbf{k} \equiv i\mathbf{k}$. This means that the Bloch plane-wave part of $\Psi_{\mathbf{k}}(z)$ for a particle supposed to travel in that region along z should in principle become an exponentially decaying function of the form $\sim e^{-\mathbf{k}z}$. Then its probability density $|\Psi(z)|^2$ calculated at fixed $\mathbf{k}$ within the band gap will be a vanishing function of the distance z, since the plane-wave component has the property of modulating the wavefunction amplitude in real space.

When a material offers the capability to bring two allowed states (separated by a forbidden region) closer at a distance shorter than, say, the $\Psi(z)$ damping then

the probability of having a particle at state $|\Psi(z)\rangle$ which tunnels across the gap becomes non negligible. From this statement three fundamental observations arise. The first one is that tunneling can be explained only under the quantum mechanical wave-particle dualism, so that no classical theory could reach these conclusions. The second important fact to be noted is that, unlike all the transitions we previously mention in this work, tunneling (at least in its simplest conceptual prototype) represents a transition in real space and not in the energy domain (a property often referred to as the "*non-locality*" of tunneling). Even if, as we will see in a moment, (phonon- or) defect-assisted processes can also involve ('oblique' or) vertical transitions. The last thing to stress – and probably the most important – is that tunneling represents an unavoidable effect whose origin comes from probabilistic nature of particles, an intrinsic property of matter. This means that no activation energies are required and that also off-state devices are affected by tunneling. Actually, the only important physical parameter is the spatial distance between the initial and the final state (although it might depend on other quantities as the electric field in the forbidden region).

The trend of modern electronics to boost the power density of integrated systems via the reduction of devices dimensions has the drawback to give its side to parasitic effects like tunneling. One must also say that in some particular cases tunneling is even exploited as *the* operating principle: think, as an example, to tunneling-field-effect transistors (TFETs) or resonant tunneling diodes (RTDs). However tunneling mechanisms are commonly seen as a source of unwanted leakage currents. As we briefly mentioned in Chapter 2, direct tunneling can heavily affect and compromise the reverse dark current of IR photodetectors, an essential feature to be reduced in order to keep under control the sensor noise: in this case the effect is rather subtle since band bending in the absorber region governs the process, no matter the detector size. At the same time defect-related tunneling processes are typical of LEDs in low-medium bias regime, as we said in Chapter 3. However, TAT can also occur in IRPDs, even if defects in MCT seem less relevant than in GaN-based materials. Moreover, being an intrinsic effect, tunneling can be only reduced via proper design (as band engineering) or controlling fabrication processes, but not eliminated at all. This should be the key point of each approach about tunneling-related issues in solid-state (opto-)electronics and it is just from this particular consideration that the importance of studying quantum tunneling arises.

Figure 4.1 shows two examples of direct (BTBT) – panel (a) – and defect-assisted (TAT) – panel (b) – electron tunneling in narrow and wide gap *p-n* junctions under reverse and forward bias, respectively. Both mechanisms here depicted belong to the class of inter-band transitions, being involved both bands. The first scheme represents an electron tunneling from a valence state towards a conduction state. Coordinates z_a and z_b are usually referred to as classical *turning points* of motion. Since BTBT can be also seen as a hole tunneling in the inverse direction (this process is not represented here), it finally corresponds to a net electron/hole (non-radiative) generation mechanism. If the reverse bias increases the two quasi-Fermi levels become furtherly splitted and the tunneling path $\ell_{\text{tunn}} = z_b - z_a$ progressively reduces,

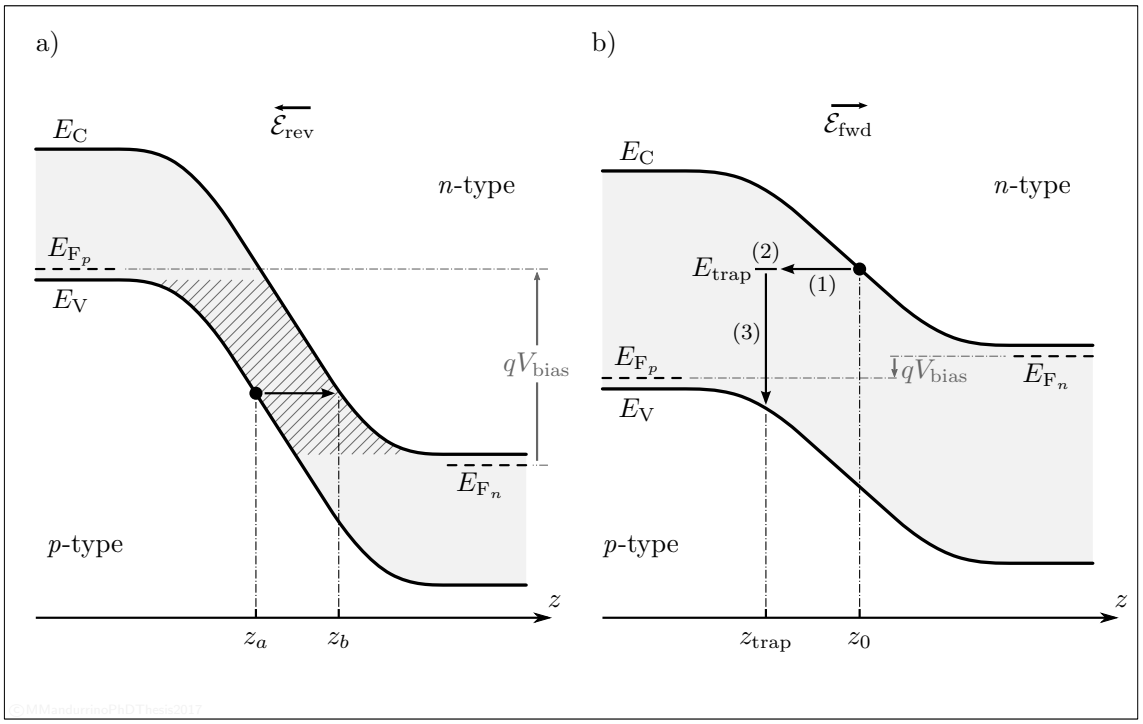

FIGURE 4.1. (a) Electron band-to-band tunneling (BTBT) in a reverse-biased narrow gap p-n junction. The transition of an electron from valence band at z_a to the conduction band at z_b (where z_a and z_b are called *turning points*) corresponds to a net generation process. The striped area represents the tunneling window. (b) Electron trap-assisted tunneling (TAT) in a forward-biased wide gap p-n junction. TAT is a three-step process: after a pure tunneling transition (1) the electron is captured by a trap of energy E_{trap} with a certain statistics (2), then it recombines with a hole in valence band (3).

because

$$\ell_{tunn} = \frac{E_g}{q\,|\mathcal{E}_{rev}(V_{bias})|}\,, \tag{4.1}$$

assuming $\mathcal{E}_{rev}$ the electric field within the gap in correspondence of the junction.

More complex is TAT represented in panel (b). Here we have only one of the possible tunneling mechanisms, where an electron leaves a conduction state and tunnels towards a defect with energy E_{trap} in the p-side. After the electron is captured by the trap, it can recombine with a hole in valence band at the same position of trap z_{trap}. Unlike previous mechanism, here we have a net recombination. This is essentially due to the opposite polarization, since wide gap emitters usually work in forward bias. Also in this case we have an inter-band transition and, again, the inverse process involving a hole with the arrows pointing in the opposite directions with respect to what shown in figure is also possible, in principle. It is straightforward that not the same statistics should be applied because the electron and hole capture are normally characterized by different coefficients $c_{n,p}$ (see Subsection 1.4.1.1), depending on their effective masses and on microscopical process details that will be analyzed in depth in Section 4.3.

Moreover, in Figure 4.2 one may find other – less relevant – tunneling processes

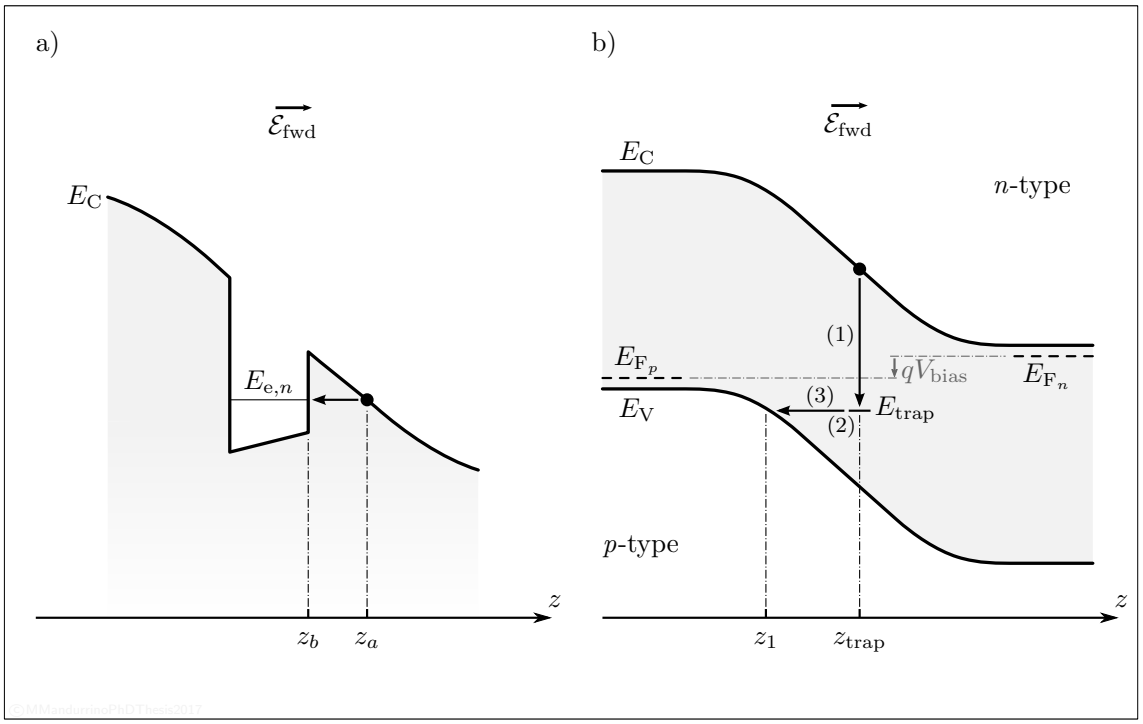

FIGURE 4.2. (a) Electron intra-band-to-band tunneling (intra-BTBT) between conduction states across the heterojunction of a single quantum well (SQW) structure under forward bias conditions. Here the final state is represented by a *quasi*-confined level in the well with energy $E_{e,n}$. (b) Alternative TAT mechanism with respect to what shown in Figure 4.1 where the electron recombination with the empty trap (1) at coordinate z_{trap} is followed by the capture process (2) and then by the pure tunneling (3).

occurring in wide gap light-emitters that we include for the sake of completeness. Panel (a) shows the intra-band-to-band tunneling (intra-BTBT) between two conduction states in the barrier and in the QW of an LED in forward bias conditions. Such process will be again useful in Section 4.4 when we will discuss the full-quantum formalism of Non Equilibrium Green's Functions (NEGF). Panel (b) instead shows different TAT paths in the same device represented in panel (b) of Figure 4.1.

At the beginning of this section we talked about spatial distance as the crucial parameter in tunneling mechanisms. In general, and more rigorously, the transition between classical and quantum mechanics occurs when the De Broglie wavelengthts $\lambda\!\!\!^{-} = \hbar/\mathbf{p}$ at play (where $\mathbf{p}$ is the particles momentum, not to be confused with the crystal momentum $\hbar\mathbf{k}$) are comparable with the characterizing dimensions of the system under investigation. So, very often the complexity of a microscopic problem can be reduced invoking the classical limit of quantum mechanics, occurring when $\hbar \to 0$, which is the only way allowing us to neglect $\lambda\!\!\!^{-}$ since the particle momentum can not be *a priori* modified. This strategy can be obviously applied also to the case of a particle tunneling through a potential barrier. If we take into account the most general (time-dependent) Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r},t) = -\frac{\hbar^2}{2m^*}\Delta_{\mathbf{r}}\Psi(\mathbf{r},t) + U(\mathbf{r})\Psi(\mathbf{r},t) \tag{4.2}$$

we notice that unfortunately the limit $\hbar \to 0$ produces a non-sense equation. This drawback has been brilliantly overcome in the work of three scientists who, in 1926, separately attained the same result [149–151]. These guys are G. Wentzel, H. A. Kramers and L. N. Brillouin, and in their honor this method has been called the *WKB approximation* (also known as *semi-* or *quasi-classical approximation* of quantum mechanics). This formalism consists in writing the wavefunction in the form

$$\Psi(\mathbf{r},t) = A\,\mathrm{e}^{\frac{i}{\hbar}S(\mathbf{r},t)}\,, \tag{4.3}$$

where A is a normalization prefactor and the (complex) function $S(\mathbf{r},t)$ has the dimensions of an action (eVs) as well as $\hbar$. By substituting Eq. (4.3) in Eq. (4.2) one obtains

$$-\frac{\partial}{\partial t}S(\mathbf{r},t) = \frac{1}{2m^*}\left(\nabla_{\mathbf{r}}S(\mathbf{r},t)\right)^2 - \frac{i\hbar}{2m^*}\Delta_{\mathbf{r}}S(\mathbf{r},t) + U(\mathbf{r})\,. \tag{4.4}$$

Now if we assume $\hbar = 0$ the simplified Schrödinger equation

$$-\frac{\partial}{\partial t}S(\mathbf{r},t) = \frac{1}{2m^*}\left(\nabla_{\mathbf{r}}S(\mathbf{r},t)\right)^2 \tag{4.5}$$

becomes the Hamilton-Jacobi equation

$$-\frac{\partial}{\partial t}S(\mathbf{r},t) = \mathbf{H}(\mathbf{r},\mathbf{p}), \tag{4.6}$$

written for the (real) classical action

$$S(\mathbf{r},t) \equiv S_{\mathrm{cl}}(\mathbf{r},t) = \int_{t_0}^{t} \mathcal{L}(\mathbf{r},\dot{\mathbf{r}},t')\,\mathrm{d}t'\,, \tag{4.7}$$

where

$$\mathbf{H}(\mathbf{r},\mathbf{p}) = \frac{\mathbf{p}^2}{2m^*} + U(\mathbf{r}) \tag{4.8}$$

is the classical Hamiltonian operator of the system (here $\mathbf{p} = \nabla_{\mathbf{r}}S_{\mathrm{cl}}(\mathbf{r},t)$ has been substituted) and $\mathcal{L}$ is the classical Lagrangian. This means that, in the limit $\hbar = 0$ Eqs. (4.3) and (4.6) describe a system following all the possible classical trajectories perpendicular to the surfaces with constant wavefunction phase $S_{\mathrm{cl}}(\mathbf{r},t)$.

As it emerges from previous calculations, the WKB approximation is legitimate only when

$$\frac{i\hbar}{2m^*}\Delta_{\mathbf{r}}S(\mathbf{r},t) \ll \frac{1}{2m^*}\left(\nabla_{\mathbf{r}}S(\mathbf{r},t)\right)^2 + U(\mathbf{r}) \tag{4.9}$$

like, for instance, in the case of particles with sufficiently high $\mathbf{p}$ travelling in a slowly variant potential $U(\mathbf{r})$, which actually satisfies the plane-wave hypothesis of Bloch theorem. For this reason we will consider hereinafter electron (and hole) $\Psi(\mathbf{r})$ including only the exponential term of the Bloch wavefunction, as in Eq. (4.3), i.e. what is commonly called *envelope wavefunction*, forgetting the other term related to the rapidly-oscillating lattice potential.

Operating now with a variable separation in the action expression we have

$$S(\mathbf{r},t) = \sigma(\mathbf{r}) - E\,t\,, \tag{4.10}$$

being E the particle energy and where we define $\sigma(\mathbf{r})$ the Hamilton's principal function, the wavefunction can be rewritten as

$$\begin{aligned}\Psi(\mathbf{r},t) &= \Psi(\mathbf{r})\,\mathrm{e}^{-\frac{i}{\hbar}Et} \\ &= A\,\mathrm{e}^{\frac{i}{\hbar}\sigma(\mathbf{r})}\,\mathrm{e}^{-\frac{i}{\hbar}Et}\,.\end{aligned} \tag{4.11}$$

Focusing only on the stationary solution, we develop the principal function according to this series expansion in $\hbar$:

$$\sigma(\mathbf{r}) = \sigma_0(\mathbf{r}) + \frac{\hbar}{i}\sigma_1(\mathbf{r}) + \left(\frac{\hbar}{i}\right)^2\sigma_2(\mathbf{r}) + \dots\,, \tag{4.12}$$

which represents the core of the whole WBK method. Now the target is to determine σ and Ψ by iterative approximations. Confining our calculations to the stationary regime, the governing equation now becomes the time-independent Schrödinger equation

$$\left(\nabla\sigma(\mathbf{r})\right)^2 - i\hbar\,\Delta\sigma(\mathbf{r}) = 2m^*\left[E - U(\mathbf{r})\right]\,. \tag{4.13}$$

By plugging Eq. (4.12) into Eq. (4.13) one obtains a system of coupled equations in the unknowns $\sigma_j(\mathbf{r})$ to be solved in sequence for $j = 0, 1, 2, \dots$:

$$\begin{cases} \left(\nabla\sigma_0(\mathbf{r})\right)^2 = 2m^*\left(E - U(\mathbf{r})\right) \\ \left(\nabla\sigma_0(\mathbf{r})\cdot\nabla\sigma_1(\mathbf{r})\right) + \dfrac{1}{2}\Delta\sigma_0(\mathbf{r}) = 0 \\ \left(\nabla\sigma_1(\mathbf{r})\right)^2 + 2\left(\nabla\sigma_0(\mathbf{r})\cdot\nabla\sigma_2(\mathbf{r})\right) + \Delta\sigma_1(\mathbf{r}) = 0 \\ \dots \end{cases}. \tag{4.14}$$

In a monodimensional problem along the z-axis, and stopping to the first order solution ($j = 0, 1$) by imposing

$$\sigma(z) \simeq \sigma_0(z) + \frac{\hbar}{i}\sigma_1(z) \tag{4.15}$$

(the second important statement of the WKB method), the system of Eq. (4.14) reduces to

$$\begin{cases} \dot{\sigma}_0^2(z) = 2m^*\left(E - U(z)\right) \equiv \mathbf{p}_z^2 \\ \dot{\sigma}_0(z)\dot{\sigma}_1(z) + \dfrac{1}{2}\ddot{\sigma}_0(z) = 0 \end{cases}, \tag{4.16}$$

in which dotted variables represent the spatial derivatives of their respecting undotted quantities. System (4.16) returns two solutions:

$$\sigma_0(z) = C \pm \int_{z_0}^{z} \mathbf{p}_{z'}\,\mathrm{d}z' \tag{4.17}$$

and

$$\sigma_1(z) \simeq -\ln\left(\sqrt{\mathbf{p}_z}\right). \tag{4.18}$$

Using these solutions and remembering Eq. (4.15) one finds the final form of the WKB wavefunction:

$$\Psi_{\mathrm{WKB}}(z) = \frac{C}{\sqrt{\mathbf{p}_z}} \mathrm{e}^{\frac{i}{\hbar} \pm \int_{z_0}^{z} \mathbf{p}_{z'}\, \mathrm{d}z'}, \tag{4.19}$$

where the lower integration limit z_0 is arbitrary unless the constant C is specified.

Written in the general form of Eq. (4.19), the WKB wavefunction is valid for each region of the $\mathbf{k}$-space. Supposing to use this formalism to describe an electron (or a hole) in a periodic lattice structure, now we can replace the free-particle momentum $\mathbf{p}$ with the crystal wavevector $\mathbf{k}$ with the common assumption $\hbar = 1$. Recalling now the first equation of system (4.16), we have

$$\mathbf{k}_z = \sqrt{2m^* \left(E - U(z)\right)}, \tag{4.20}$$

so if $E > U(z)$ then $\mathbf{k}_z$ is real, otherwise if $E < U(z)$ we have a purely imaginary $\mathbf{k}_z$. As one may understand, this exactly corresponds to the propagation conditions encountered by plane wave travelling within an allowed or a forbidden region, respectively. And this also corresponds to the physics of a particle wavefunction tunneling through a potential barrier, under the WKB hypotheses. The points where $\mathbf{k}_z$ is zero (i.e. at $E = U(z)$) coincide with the already mentioned *turning points*. In correspondence of these points the WKB approximation is typically rather weak. For this reason the particle propagation through the barrier is described by means of some mathematical manipulations which allow to join the WKB solutions calculated inside and outside the barrier with specific connection-functions at the turning points, whatever the potential shape. This procedure, however, is also useful to determine the constant of integration C we leaved undefined in previous calculations.

Before proceeding with the theoretical treatment of a tunneling WKB wavefunction is interesting to note some additional information given by Eq. (4.19). In particular, since $|\Psi| \propto 1/\sqrt{\mathbf{k}}$, then we have that

$$\begin{aligned} P[z'; z''] &= \int_{z'}^{z''} \Psi^*(z)\, \Psi(z)\, \mathrm{d}z \\ &= |\Psi(z)|^2\, \Delta z \\ &\simeq \frac{\Delta z}{\mathbf{k}_z} \sim \Delta t, \end{aligned} \tag{4.21}$$

where $P[z'; z'']$ is the probability to have the particle described by $\Psi(z)$ located within the interval $\Delta z = z'' - z'$. Hence Eq. (4.21) states that this probability is somehow proportional to the time interval Δt which is necessary to travel along Δz.

Now let us call z_a and z_b the two turning points (as also depicted in Figure 4.3). They split the (semi-)classical wavefunction trajectory in the real space into three different regions: in region I we suppose to have a forward propagating $\Psi(z)$. At

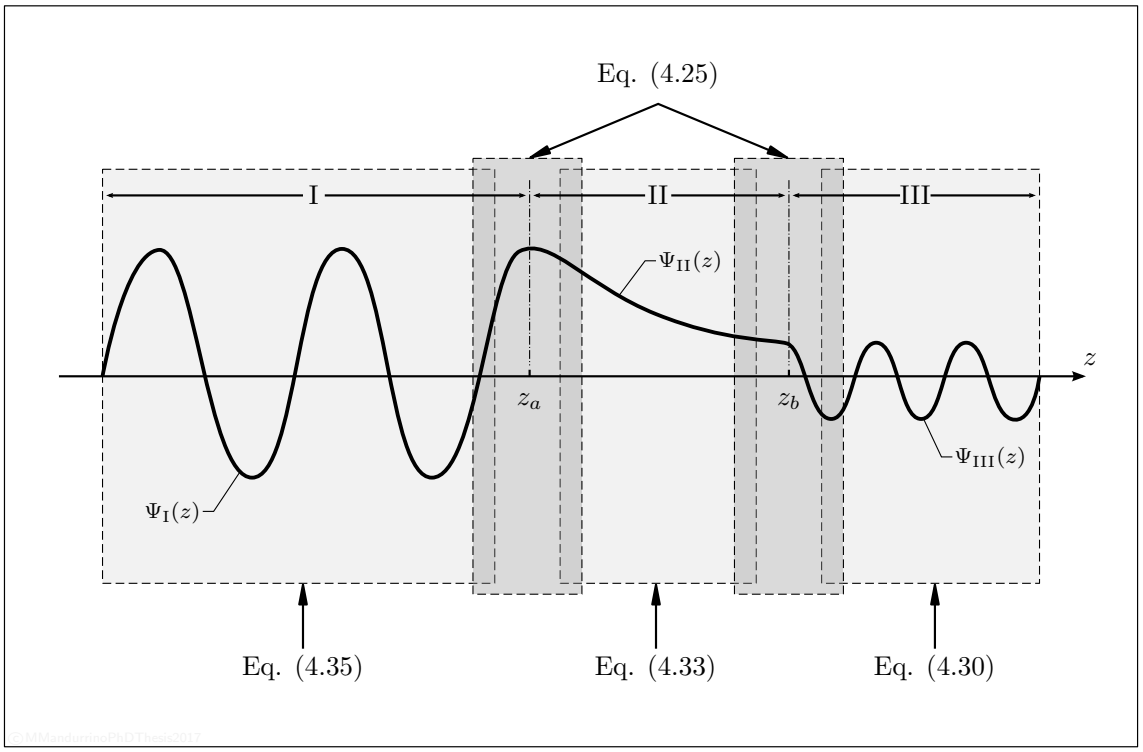

FIGURE 4.3. Solving scheme for a particle (envelope) wavefunction tunneling from left to right under WKB hypotheses. The range included between z_a and z_b is the forbidden region (II) represented by a potential barrier. It is worth noting that connection functions at turning points do overlap with asymptotic solutions written for regions I, II and III.

the point z_a the energy of the incoming particle E equals the barrier energy $U(z)$: here the wave is partially transmitted through region II (as a vanishing exponential) and in part is reflected backward into region I (as oscillating wave). Finally, at the turning point z_b the wave is transmitted into region III and, at the same time, reflected into region II (we neglect second-order counter-reflections or other resonance effects).

Let us start from region III. Operating the change of variables $\Psi(z) = \varphi(z)\sqrt[3]{u}$ with

$$u = \int_{z_b}^{u} \sqrt{\frac{2m^*\mathcal{U}}{\hbar^2}(u - z_b)}\,\mathrm{d}u\,, \tag{4.22}$$

in which $\mathcal{U}$ is a potential energy term, and substituting it into the time-independent Schrödinger equation we obtain

$$\frac{\mathrm{d}^2\varphi(z)}{\mathrm{d}u^2} + \frac{1}{u}\frac{\mathrm{d}\varphi(z)}{\mathrm{d}u} + \left(1 - \frac{1}{9u^2}\right)\varphi(z) = 0\,. \tag{4.23}$$

Eq. (4.23) is the Bessel's differential equation. This means that its solution $\varphi(z)$

assumes the role of a Bessel function $J_{\pm\frac{1}{3}}(u)$ of the order $\pm\frac{1}{3}$. Remembering that

$$J_j(u) = \sum_{n=0}^{\infty} \frac{(-1)^n \left(\frac{u}{2}\right)^{j+2n}}{n!\,\Gamma\left(j+n+1\right)}\,, \tag{4.24}$$

where Γ is the gamma function, then, after some calculations, one obtains

$$\Psi(u) = \sqrt[3]{u}\left(A\,J_{\frac{1}{3}}(u) + B\,J_{-\frac{1}{3}}(u)\right)\,. \tag{4.25}$$

Now our $\Psi(u)$, which is for the moment expressed as a function of the auxiliary variable u, must be reconducted to a suitable function of the real coordinate z. In order to do so we operate again a substitution assuming that

$$\zeta(z) = \int_{z_b}^{z} \sqrt{\frac{2m^*}{\hbar^2}(E - U(z))}\,\mathrm{d}z\,. \tag{4.26}$$

From Eq. (4.26) follows that

$$\Psi(z) \simeq \sqrt{\frac{\zeta(z)}{\mathbf{k}_z}}\left(A\,J_{\frac{1}{3}}(\zeta(z)) + B\,J_{-\frac{1}{3}}(\zeta(z))\right)\,. \tag{4.27}$$

By its construction, Eq. (4.27) satisfies all the connection requirements at the turning points. To ensure to have a propagating wave in III we must choose

$$A_{\mathrm{III}} = \frac{2}{i + \mathrm{e}^{i\frac{\pi}{6}}} \tag{4.28}$$

and

$$B_{\mathrm{III}} = \frac{2}{\mathrm{e}^{-i\frac{\pi}{6}} - \mathrm{e}^{\frac{\pi}{2}}}\,, \tag{4.29}$$

thus

$$\Psi_{\mathrm{III}}(z) \sim \sqrt{\frac{2}{\pi\mathbf{k}_z}}\exp\left(i\frac{\zeta(z) - \pi}{4}\right). \tag{4.30}$$

In the same way we can operate at the turning point z_a, where the condition expressed by Eq. (4.26) here reads

$$\zeta'(z) = \int_{z}^{z_a} \sqrt{\frac{2m^*}{\hbar^2}(E - U(z))}\,\mathrm{d}z\,. \tag{4.31}$$

By plugging Eq. (4.31) into Eq. (4.27) and using

$$A_{\mathrm{II}} = B_{\mathrm{II}} = \frac{i\,\mathrm{e}^{|\zeta(z)|+|\zeta'(z)|}}{\cos\frac{\pi}{6}} \tag{4.32}$$

we have

$$\Psi_{\mathrm{II}}(z) \sim i\sqrt{\frac{2}{\pi|\mathbf{k}_z|}}\exp\left(|\zeta(z)| + |\zeta'(z)|\right), \tag{4.33}$$

where

$$|\zeta(z)| + |\zeta'(z)| = \left|\int_{z_a}^{z_b}\sqrt{\frac{2m^*}{\hbar^2}(E - U(z))}\,\mathrm{d}z\right|, \tag{4.34}$$

and, finally,

$$\Psi_{\mathrm{I}}(z) \sim \sqrt{\frac{2}{\pi\mathbf{k}_z}}\left[\frac{A_{\mathrm{I}}}{2}\left(\mathrm{e}^{-i\frac{5\pi}{2}} + \mathrm{e}^{-i\frac{\pi}{12}}\right)\mathrm{e}^{i\zeta(z)} + \frac{A_{\mathrm{I}}}{2}\left(\mathrm{e}^{i\frac{5\pi}{2}} + \mathrm{e}^{i\frac{\pi}{12}}\right)\mathrm{e}^{-i\zeta(z)}\right], \tag{4.35}$$

with $A_{\mathrm{I}} \equiv A_{\mathrm{II}}$. Notice that Eqs. (4.35), (4.33) and (4.30) are all in the general WKB form written in Eq. (4.19). It is also worth stressing that the WKB method can be applied to any shape of potential $U(z)$, provided that proper normalization coefficients are calculated for each region. In particular, for potential discontinuities at turning points (as for the barrier surrounding a QW), the use of suitable Airy functions in place of the Bessel-like join-wavefunctions in Eq. (4.25) is often adopted.

Figure 4.3 shows the WKB scheme we proposed to solve the problem of a particle (envelope) wavefunction impinging on a potential barrier, where z_a and z_b delimit the forbidden region (II). The wavefunction is not to scale since it is demonstrated that in order to obtain a reliable connection between solutions at turning points and far from them the domain of Eq. (4.25) should include a certain number of wavelengths.

Performing the ratio

$$\begin{aligned}\mathcal{T} = \left|\frac{\Psi_{\mathrm{III}}(z)}{\Psi_{\mathrm{I}}(z)}\right|^2 &= \frac{1}{\left|\frac{A_{\mathrm{I}}}{2}\left(\mathrm{e}^{-i\frac{5\pi}{2}} + \mathrm{e}^{-i\frac{\pi}{12}}\right)\right|^2} \\ &= \exp\left[-\frac{2\sqrt{2m^*}}{\hbar^2}\left|\int_{z_a}^{z_b}\sqrt{(E - U(z))}\,\mathrm{d}z\right|\right]\end{aligned} \tag{4.36}$$

we obtain the *tunneling transmission factor*, also known as *barrier transparency*, i.e. the tunneling probability. It has to be observed that $\mathcal{T}(E)$ is a function of the energy E, a characteristic quantity of the considered particle which remains conserved during the tunneling process, even if $|\Psi_{\mathrm{I}}(z)|^2 \neq |\Psi_{\mathrm{III}}(z)|^2$.

Let us consider an interesting case, which is useful to introduce tunneling effect in semiconductor devices. Suppose to have a band diagram as in panel (a) of Figure 4.1 and a valence electron tunneling from the p-side to the n-side across the energy gap, from z_a to z_b. If the semiconductor is a direct band gap material (as in the case of HgCdTe) the wavevector perpendicular to the motion $\mathbf{k}_\perp$ is conserved (otherwise, in case of indirect band gap semiconductors the energy conservation is assured through phonon emission). Moreover, suppose also that the energy $E_\perp$ associated to the perpendicular wavevector $\mathbf{k}_\perp$ is finite and positive. As pointed out by S. M. Sze [152]

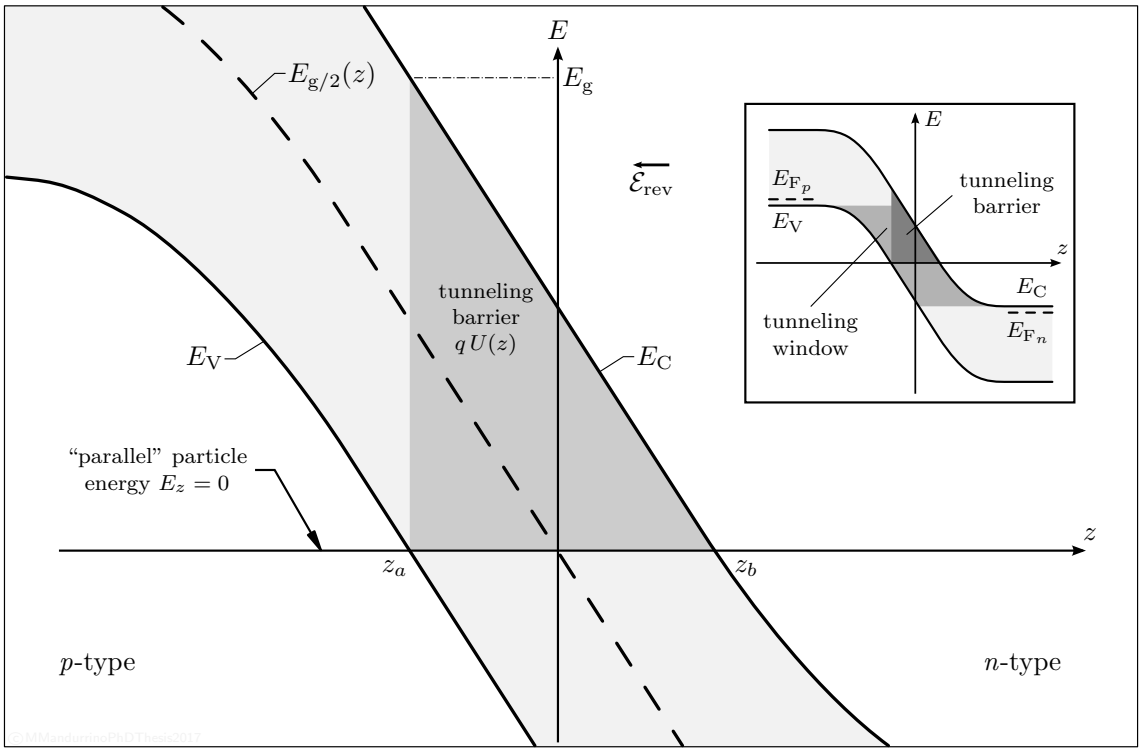

FIGURE 4.4. Tunneling barrier and characteristic energies for a direct band-to-band transition in a p-n junction under reverse bias (here the depletion region is rather extended so the barrier has been approximated by a triangle). Classical turning points of motion z_a and z_b are at coordinates defined by Eq. (4.42). In the inset: zoom-out of the diode band structure highlighting both tunneling energy window and its triangular barrier.

the potential shape within the forbidden gap does not affect so much the tunneling behavior. The most generic representation involves a triangular barrier

$$E - U_{\triangle}(z) = -\left(\frac{E_g}{2} - E_{g/2}(z)\right) , \tag{4.37}$$

where $E_{g/2}(z) = -q\mathcal{E}_{rev}z$ is the mid-gap energy. Otherwise, as claimed by J. H. Moll (see Ref. [153]), also the parabolic trend can be assumed

$$E - U_{\frown}(z) = -\frac{(E_g/2)^2 - E_{g/2}^2(z)}{E_g} , \tag{4.38}$$

which derives from band dispersion calculations performed via $\mathbf{k} \cdot \mathbf{p}$ perturbation method in the near-zero-gap InSb alloy (after E. O. Kane [154]). The narrower the band gap the stronger the quadratic correction $E_{g/2}^2(z)$ needed to describe the barrier at the forbidden gap. Performing the calculation of $\mathcal{T}(E)$ for both shapes one obtains

$$\mathcal{T}_{\triangle}(E) = \exp\left(-\frac{4\sqrt{2m^*}E_g^{3/2}}{3q\hbar\mathcal{E}_{rev}}\right) \tag{4.39}$$

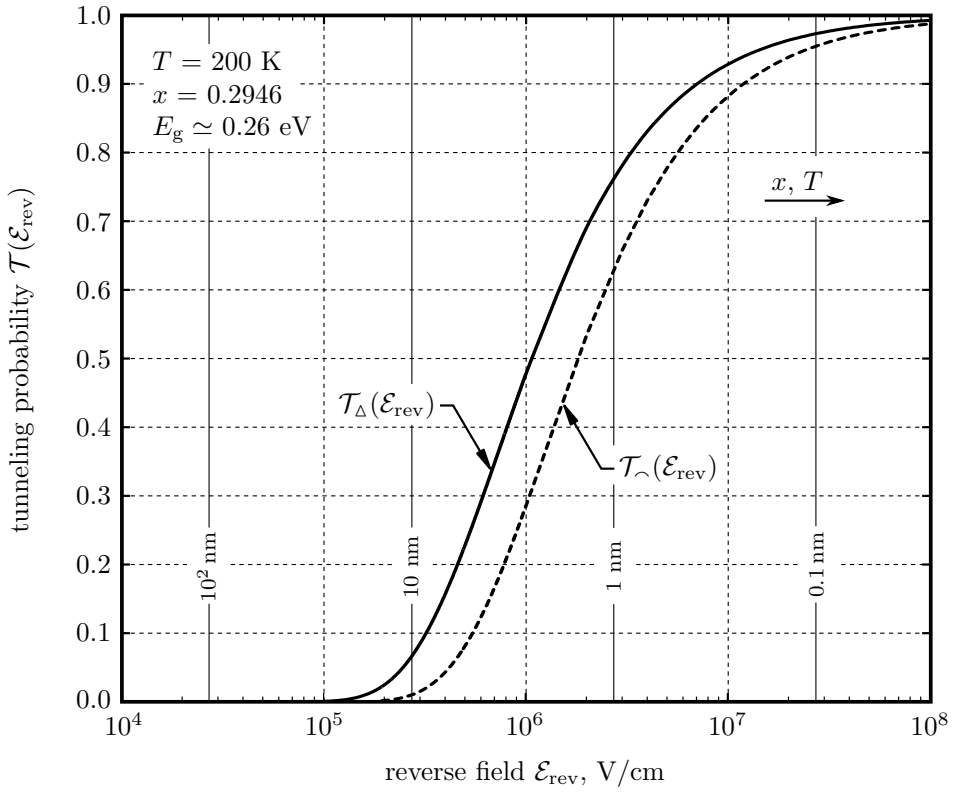

FIGURE 4.5. Tunneling probabilities $\mathcal{T}$ plotted as a function of the electric field (and tunneling path, in nm) in a reverse-biased MCT p-n diode at $T = 200$ K with cadmium fraction $x = 0.2946$. Both triangular and parabolic barrier approximations are represented where, respectively, Eqs. (4.39) and (4.40) have been used. Note: the horizontal arrow indicates that for increasing mole fraction or temperature the curves shift towards high electric fields.

and

$$\mathcal{T}_\frown(E) = \exp\left(-\frac{\pi\sqrt{m^*}E_g^{3/2}}{2\sqrt{2}q\hbar\mathcal{E}_{rev}}\right), \tag{4.40}$$

from which results to be pretty evident that, as we said, the barrier choice is rather irrelevant, since the expressions in Eqs. (4.39) and (4.40) are formally identical except for a small constant.

Let us take into account the last case, i.e. consider a parabolic barrier with the most general expression for the total energy

$$E = E_z + E_\perp\,, \tag{4.41}$$

which is the sum of energy components parallel and perpendicular to the tunneling trajectory.

In order to obtain $E - U(z) = 0$ at turning points we have to shift our reference frame in such a way that the origin is $E_{g/2}(0) = 0$ (see Figure 4.4). So, z_a and z_b

are given by

$$z_{a,b} = \pm \frac{1}{q\mathcal{E}_{\text{rev}}} \sqrt{\left(\frac{E_{\text{g}}}{2}\right)^2 + E_{\text{g}} E_{\perp}}\,, \tag{4.42}$$

where + and − signs stand for z_b and z_a, respectively. Notice that if no transversal momentum occurs then turning points are separated by the quantity $E_{\text{g}}/q\mathcal{E}_{\text{rev}}$. Introducing a component $E_{\perp}$ this distance (i.e. the tunneling path) increases by reason of Eq. (4.42). Rewriting now the tunneling probability of Eq. (4.40) we have

$$\mathcal{T}_{\frown}(E) = \overbrace{\exp\left(-\frac{4\sqrt{2m^*}E_{\text{g}}^{3/2}}{3q\hbar\mathcal{E}_{\text{rev}}}\right)}^{\mathcal{T}_{\frown}(E_z)} \overbrace{\exp\left(-\frac{2E_{\perp}}{\bar{E}}\right)}^{\mathcal{T}_{\frown}(E_{\perp})}, \tag{4.43}$$

where "parallel" and "transversal" components, again, have been highlighted, and

$$\bar{E} = \frac{\sqrt{2}q\hbar\mathcal{E}_{\text{rev}}}{\pi\sqrt{m^* E_{\text{g}}}}\,. \tag{4.44}$$

Notice that from expressions above the thermal trend of BTBT probability in HgCdTe p-n photodiodes under reverse bias can be inferred. Combining Eq. (2.17), which describes the band gap widening with increasing T, and Eq. (4.43) the inverse proportionality between $\mathcal{T}_{\frown}(E)$ and temperature is well evident. To enhance direct tunneling also low effective masses, narrower energy gap and high electric field have to be invoked: as plotted in Figure 4.5, the role of BTBT becomes relevant when the electric field in correspondence of the junction is of the order of $\sim 10^6$ V/cm, where the field is supposed to be constant within the depletion region (average field approximation). Moreover, in the case considered of MCT with cadmium fraction $x = 0.2946$ at $T = 200$ K the assumption of parabolic barrier differs from the triangular one at fixed $\mathcal{T}$ by a field factor $\sim 2\cdot 10^6$ V/cm.

For the sake of brevity, let us confuse the notation $\mathcal{T}_{\frown}$ with simply $\mathcal{T}$, since we will hereinafter make implicit the use of a parabolic barrier. In order to have a tunneling transition across a diode junction two fundamental conditions have to be met: the first is the availability of the initial state in valence band and the second is the availability of a final state in conduction band, requirements that are usually described in the Fermi-Dirac statistics by the well-known Fermi functions $f_{\text{V}}(E)$ and $(1 - f_{\text{C}}(E))$, respectively. If in non-degenerate semiconductors these conditions are automatically fulfilled (at least at 0 K) in degenerate ones the availability of states has to be carefully managed and verified. Thus, in general, the net tunneling flux per unit of differential volume ($\mathrm{d}V = A\,\mathrm{d}z$) is

$$\Phi_{\text{tunn}} = \mathcal{T}(E)\,\left(f_{\text{C}}(E) - f_{\text{V}}(E)\right)\,, \tag{4.45}$$

thus the differential current is given by

$$\begin{aligned}
\mathrm{d}I_{\text{tunn}} &= \Phi_{\text{tunn}}\, A\,\mathrm{d}z \\
&= \frac{q^2\mathcal{E}_{\text{rev}} m^*}{2\pi^2\hbar^3}\,\mathcal{T}(E)\,\left(f_{\text{C}}(E) - f_{\text{V}}(E)\right)\, A\,\mathrm{d}E_{\perp}\,\mathrm{d}z \\
&= \frac{q m^*}{2\pi^2\hbar^3}\,\mathcal{T}(E)\,\left(f_{\text{C}}(E) - f_{\text{V}}(E)\right)\, A\,\mathrm{d}E_{\perp}\,\mathrm{d}E\,,
\end{aligned} \tag{4.46}$$

where $q\mathcal{E}_{\text{rev}} = \mathrm{d}E/\mathrm{d}z$ has been used in the last passage. Hence, the tunneling net current can be obtained by integrating over the energy:

$$I_{\text{tunn}} = \frac{qm^*}{2\pi^2\hbar^3} \exp\left(-\frac{\pi\sqrt{m^*}E_{\text{g}}^{3/2}}{2\sqrt{2}q\hbar\mathcal{E}_{\text{rev}}}\right) \int \left(f_{\text{C}}(E) - f_{\text{V}}(E)\right) \mathrm{e}^{-2E_\perp/\bar{E}}\, A\, \mathrm{d}E_\perp\, \mathrm{d}E\,. \quad (4.47)$$

In order to write Eq. (4.47) in a less generic form, suitable for specific device structures, one has to account for some hypotheses. Let us assume, for instance, to apply a bias such that $V_{\text{bias}} \gg \bar{E}/q$. This is needed to have discontinuous quasi-Fermi levels at the borders of the depletion region. Suppose also to neglect the perpendicular contribution given by $E_\perp$, whose integration would be

$$\int \mathrm{e}^{-2E_\perp/\bar{E}}\, \mathrm{d}E_\perp \simeq \int \left(1 - \mathrm{e}^{-2E_{\text{min}}/\bar{E}}\right) \mathrm{d}E\,, \quad (4.48)$$

where E_{min} is the minimum between $E - E_{\text{C,min}}$ and $E_{\text{V,max}} - E$, as pointed out in Ref. [152] (with $E_{\text{C,min}}$ and $E_{\text{V,max}}$ the band edges in neutral regions). Under these hypotheses:

$$I_{\text{tunn}} = \frac{q^3\sqrt{2m^*}\mathcal{E}_{\text{rev}}V_{\text{bias}}}{4\pi^3\hbar^2\sqrt{E_{\text{g}}}} \exp\left(-\frac{\pi\sqrt{m^*}E_{\text{g}}^{3/2}}{2\sqrt{2}q\hbar\mathcal{E}_{\text{rev}}}\right). \quad (4.49)$$

Eq. (4.49) represents the most general WKB (analytical) expression of inter-band-to-band tunneling current in a direct band gap semiconductor p-n junction (under reverse bias conditions).

For the sake of completeness it has to be mentioned that WKB is not the only analytical method used to solve tunneling problems. In 1966 K. H. Gundlach [155] found an alternative procedure to solve the time-independent Schrödinger equation, rewritten as

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2}\Psi\left(\alpha(z)\right) - \alpha(z)\,\Psi\left(\alpha(z)\right) = 0\,, \quad (4.50)$$

where $\alpha(z) = \beta - z/\gamma$,

$$\beta = -\sqrt[3]{\frac{2m^*}{q^2\mathcal{E}^2\hbar^2}}\left(E - U(z_a) + z_a q\mathcal{E}\right) \quad (4.51)$$

and

$$\gamma = -\sqrt[3]{\frac{\hbar^2}{2qm^*\mathcal{E}}}\,. \quad (4.52)$$

Eq. (4.50) is the *Airy differential equation* and its solutions are the first- and second-type Airy functions $\mathrm{Ai}\left(\alpha(z)\right)$ and $\mathrm{Bi}\left(\alpha(z)\right)$. Thus the wavefunction is defined as a linear combination of Airy functions:

$$\Psi\left(\alpha(z)\right) = C_1 \cdot \mathrm{Ai}\left(\alpha(z)\right) + C_2 \cdot \mathrm{Bi}\left(\alpha(z)\right)\,, \quad (4.53)$$

where $C_{1,2}$ are normalization constants. The resulting transmission factor, that in Gundlach formalism is very compact, has been demonstrated to be almost coincident with the WKB one [156].

The last category of solving methods consists in the transfer matrix method. By approximating the potential barrier as a discrete set of N linear- or constant-piecewise potential segments and by writing the wavefunction Ψ in each sub-domain as a sum of a progressive and regressive plane wave, the total transmission is the product of N matrices each of those represents a complex transfer matrix, i.e. the matrix connecting the amplitude of the incoming and reflected wavefunctions at the boundaries of each respective domain.

Besides these important analytical methods there are also numerical strategies allowing to solve the Schrödinger equation for any kind and any complexity of potential. One of the most important procedure is based on the finite element method (FEM), which assumes again to divide the whole domain in a grid of sub-domains. With opportune approximations (details and results of this method will be presented in Section 4.4) one can obtain eigenvalues and eigenfunctions only by imposing appropriate boundary conditions.

In the next subsection we will introduce some relevant examples taken from the literature which allow us to describe the transition between purely analytical formalisms as the WKB, useful to represent small and static problems through their microscopic quantities, and the framework of device simulation, where physical complexity and non-linearity of the systems under analysis dominate and where the description focus is shifted towards average or collective variables in a quasi-stationary framework. So, we will start from theoretical principles and then we will present approaches that can be still considered belonging to the area of analytical solutions but, as it will result more clear, they are specifically concieved for the implementation into numerical solvers.

4.1.1 From pure analytical to numerical-oriented picture

Several researchers derived *ad hoc* equations in order to describe specific mechanisms of their interest, as the author also did concerning HgCdTe *p-i-n* IR photodetectors (see Subsection 4.2). However, the formalism just derived for band-to-band transitions – based on WKB wavefunction approximation and Kane's band dispersion – is subtended to almost all tunneling processes, both direct and defect-assisted.

The aim of this subsection is to present the most robust theories of band-to-band and trap-assisted tunneling mechanisms and then to move towards a device-level description of these phenomena. Being the physics behind TAT strictly dependent on the defect density, here we present only basic calculations due to G. A. M. Hurkx and valid for lightly defective materials, as narrow gap semiconductors used for IR detection. Then in Section 4.3 his formalism will be extended to other systems thanks to the multi-phonon-emission (MPE) theory, which introduces a higher order of complexity into our physical description.

4.1.1.1 Band-to-band tunneling (BTBT): Kane formalism

The study of direct tunneling processes in MCT IR photodetectors started in 70s-80s with works as those by M. A. Kinch and J. P. Rosbeck [157, 158] but the

discovery of BTBT in electronic devices has a longer history. Since 1934, when C. Zener suggested the possibility to observe electrons tunneling through the forbidden region [159], the effect of this quantum process in electronic devices was known. Some years later, in 1957, A. G. Chynoweth and K. G. McKay [160] were able to associate some new experimental features occurring in p-n junctions under reverse bias to Zener tunneling mechanism. Then, late 50s have been characterized by a series of interesting theoretical investigations on direct tunneling (see, for instance, the works by G. H. Wannier, E. N. Adams, W. Franz and L. V. Keldysh in Refs. [161–164]) which essentially focused their attention on perturbation effects to the system Hamiltonian and their impact on the integral within the tunneling probability.

Following Kane's 1960 paper [154], a true milestone in this area of research, we account for a 1D time-independent Schrödinger equation written in the lattice momentum representation in the form

$$\left[E_n(\mathbf{k}) - i\mathcal{E}\frac{\partial}{\partial \mathbf{k}_z} - E_{\text{tot}}\right]\phi_n(\mathbf{k}) - \sum_m \mathcal{E}\,\mathbf{X}_{n,m}\,\phi_m(\mathbf{k}) = 0\,, \tag{4.54}$$

where n, m are band indices, $\mathcal{E}$ is the (reverse) electric field, $\phi_n(\mathbf{k})$ are wavefunction such that, in the real space representation,

$$\psi(\mathbf{r}) = \sum_{n,\mathbf{k}} \phi_n(\mathbf{k})\,\Psi_{n,\mathbf{k}}(\mathbf{r}) \tag{4.55}$$

(with $\Psi_{n,\mathbf{k}}(\mathbf{r})$ a Bloch wavefunction), E_n are electron energies, E_{tot} is an overall system energy term and where $i\partial/\partial\mathbf{k}_z$ and $\mathbf{X}_{n,m}$ are the so-called "intra-band" and "inter-band" operators, respectively. Being Eq. (4.54) written in the wavevector space, also diagonal elements $\mathbf{X}_{n,n}$ have an intrinsic intraband nature, which depends on the electron energy via

$$E_n(\mathbf{k},\mathcal{E}) = E_n(\mathbf{k}) - \mathcal{E}\,\mathbf{X}_{n,n}\,, \tag{4.56}$$

in which only a first-order dependence on the electric field has been accounted for. At the same time off-diagonal elements are defined

$$\mathbf{X}_{n,m} = i\int u^*_{n,\mathbf{k}}(z)\frac{\partial}{\partial\mathbf{k}}\,u_{m,\mathbf{k}}(z)\,\mathrm{d}z\,, \tag{4.57}$$

with $u_{n,\mathbf{k}}(z)$ the rapidly oscillating part of Bloch wavefunctions $\Psi_{n,\mathbf{k}}(\mathbf{r})$.

The interband transition matrix elements, which Kane demonstrated to be valid both for first- and second-order field dependence, are given by

$$M_{n,m}(\mathbf{k}_z) = -\frac{\mathcal{E}}{\mathbf{K}}\int \mathbf{X}_{n,m}\,\exp\left(\frac{i}{\mathcal{E}}\int_0^{\mathbf{k}_z} E_n(\mathbf{k}',\mathcal{E}) - E_m(\mathbf{k}',\mathcal{E})\,\mathrm{d}\mathbf{k}'\right)\mathrm{d}\mathbf{k}_z\,, \tag{4.58}$$

where $\mathbf{K}$ is the principal vector of the reciprocal lattice. Supposing to have a two-band dispersion relation like

$$E_{\text{C,V}}(\mathbf{k}) = \frac{E_\text{g}}{2} + \frac{\hbar^2\mathbf{k}^2}{2m_0} \pm \frac{\sqrt{E_\text{g}^2 + \frac{E_\text{g}\hbar^2\mathbf{k}^2}{m_\text{r}^*}}}{2}\,, \tag{4.59}$$

where

$$m_{\mathrm{r}}^{*} = \frac{m_{\mathrm{C}}^{*} + m_{\mathrm{V}}^{*}}{m_{\mathrm{C}}^{*}\, m_{\mathrm{V}}^{*}} \tag{4.60}$$

is the reduced effective mass, the interband operator reads

$$\mathbf{X}_{\mathrm{C,V}} = i \frac{\hbar E_{\mathrm{g}}^{2/3}}{2\sqrt{m_{\mathrm{r}}^{*}}\,(E_{\mathrm{C}} - E_{\mathrm{V}})^{2}}\,. \tag{4.61}$$

Assuming a sufficiently small and constant electric field $\mathcal{E}$ and approximating the integral in Eq. (4.58) by expanding its integrating in the vicinity of the point giving the maximum contribution to the integral and then extending the domain from $-\infty$ to $+\infty$, one can rewrite Eq. (4.58) as

$$M_{n,m}(\mathbf{k}_z) = \frac{\pi\mathcal{E}}{3\mathbf{K}} \exp\left(-\frac{\pi E_{\mathrm{g}}^{3/2}}{4\hbar\mathcal{E}}\right). \tag{4.62}$$

The tunneling rate $\mathcal{R}_{\mathrm{tunn}}$ (in $\mathrm{cm}^{-3}\mathrm{s}^{-1}$) can be found multiplying the transition probability per electron per unit time

$$W_{\mathrm{V}\to\mathrm{C}} = \frac{2\pi}{\hbar}\left|M_{n,m}(\mathbf{k}_z)\right|^{2} \rho(E) \tag{4.63}$$

(being $\rho(E) \simeq \mathbf{K}/2\pi\mathcal{E}$ the DOS under constant field approximation) by the number of electrons per volume and integrating for all occupied states. The result is

$$\mathcal{R}_{\mathrm{tunn}} = \frac{q^{2}\sqrt{m_{\mathrm{r}}^{*}}}{18\pi\hbar^{2}\sqrt{E_{\mathrm{g}}}}\,\mathcal{E}^{2} \exp\left(-\frac{\pi\sqrt{m_{\mathrm{r}}^{*}}E_{\mathrm{g}}^{3/2}}{2q\hbar\mathcal{E}}\right). \tag{4.64}$$

Some considerations follow:

1. Eq. (4.64) holds for direct band gap semiconductors, where momentum is conserved during tunneling transitions.

2. Since we are modeling direct transitions from valence to conduction band in reverse biased junctions, the rate $\mathcal{R}_{\mathrm{tunn}}$ actually represents a BTBT generation rate G_{BTBT}.

3. Notice that Eq. (4.64) can be written in the more general form

$$\mathcal{R}_{\mathrm{tunn}} = A\,\mathcal{E}^{2} \cdot \exp\left(-B/\mathcal{E}\right), \tag{4.65}$$

where Kane expressions for coefficients A and B are

$$A_{\mathrm{Kane}} = \frac{q^{2}\sqrt{m_{\mathrm{r}}^{*}}}{18\pi\hbar^{2}\sqrt{E_{\mathrm{g}}}}\,, \tag{4.66}$$

which is adimensional, and

$$B_{\mathrm{Kane}} = \frac{\pi\sqrt{m_{\mathrm{r}}^{*}E_{\mathrm{g}}^{3}}}{2q\hbar} \tag{4.67}$$

in V/cm units as an electric field. As a matter of fact, B is the *critical field* governing the exponential growth of $\mathcal{R}_{\text{tunn}}$, as will become more clear in the following.

4. In semiconductors where $m_{\text{e}}^* \gg m_{\text{h}}^*$ we can safely assume that $m_{\text{r}}^* \simeq m_{\text{e}}^*$ (for electron tunneling).

5. Here it has not been demonstrated but it is worth noting that in the forbidden region wavefunctions in Kane formalism have the same asymptotic trend of WKB wavefunctions.

By taking inspiration from Kane theory, almost three decades later G. A. M. Hurkx developed a tunneling model suitable for direct band gap reverse biased p-n junctions under the hypothesis of constant electric field [165, 166]. According to Hurkx the BTBT rate can be expressed by

$$\mathcal{R}_{\text{tunn}} = A\mathcal{E}^2 D\left(\varphi(z), E_{\text{F}_n}, E_{\text{F}_p}\right) \cdot \exp\left(-B/\mathcal{E}\right), \tag{4.68}$$

where

$$D\left(z, \varphi, E_{\text{F}_n}, E_{\text{F}_p}\right) = \frac{1}{\exp\left(-\frac{E_{\text{F}_p}(z) - q\varphi(z)}{k_{\text{B}}T}\right) + 1} - \frac{1}{\exp\left(-\frac{E_{\text{F}_n}(z) - q\varphi(z)}{k_{\text{B}}T}\right) + 1}, \tag{4.69}$$

where $\varphi(z)$ is the electrostatic potential and $E_{\text{F}_{n,p}}$ the electron/hole quasi-Fermi levels. In the Hurkx model the "perpendicular energy" has not been neglected, indeed the function D is an approximation of the integral in Eq. (4.48), thus:

$$D\left(z, \varphi, E_{\text{F}_n}, E_{\text{F}_p}\right) \simeq \int \left(1 - \mathrm{e}^{-2E_{\min}/\bar{E}}\right) \mathrm{d}E\,. \tag{4.70}$$

To obtain the Hurkx rate for indirect semiconductors it is sufficient to use, in Eq. (4.68), the exponent 5/2 in place of 2.

Once the material-dependent parameters A and B have been fixed, Eq. (4.68) allows to model BTBT in each device (also under forward bias), thanks to the function D which depends on macroscopic quantities that can be easily computed by numerical simulators. It is also interesting to highlight the effect of temperature on tunneling current pointed out by Hurkx in Ref. [165]. As reported there,

$$\frac{\mathrm{d}\left(\ln I_{\text{tunn}}\right)}{\mathrm{d}T} = -\frac{3}{2E_{\text{g}}(T)} \frac{B\left(E_{\text{g}}(T)\right)}{\mathcal{E}} \frac{\mathrm{d}E_{\text{g}}(T)}{\mathrm{d}T}, \tag{4.71}$$

which gives an almost linear trend of $I_{\text{tunn}}(T)$, whose slope essentially varies by reason of the material band gap widening with T.

Thanks to more recent works by R. Adar and K. Jóźwikowski *et al.* [167, 168] it has been demonstrated that the most suitable choice of A and B for MCT material system is

$$A_{\text{MCT}}(x, T) = \frac{q^2\sqrt{2\,m^*(x, T)}}{4\pi^3\hbar^2\sqrt{E_{\text{g}}(x, T)}} \tag{4.72}$$

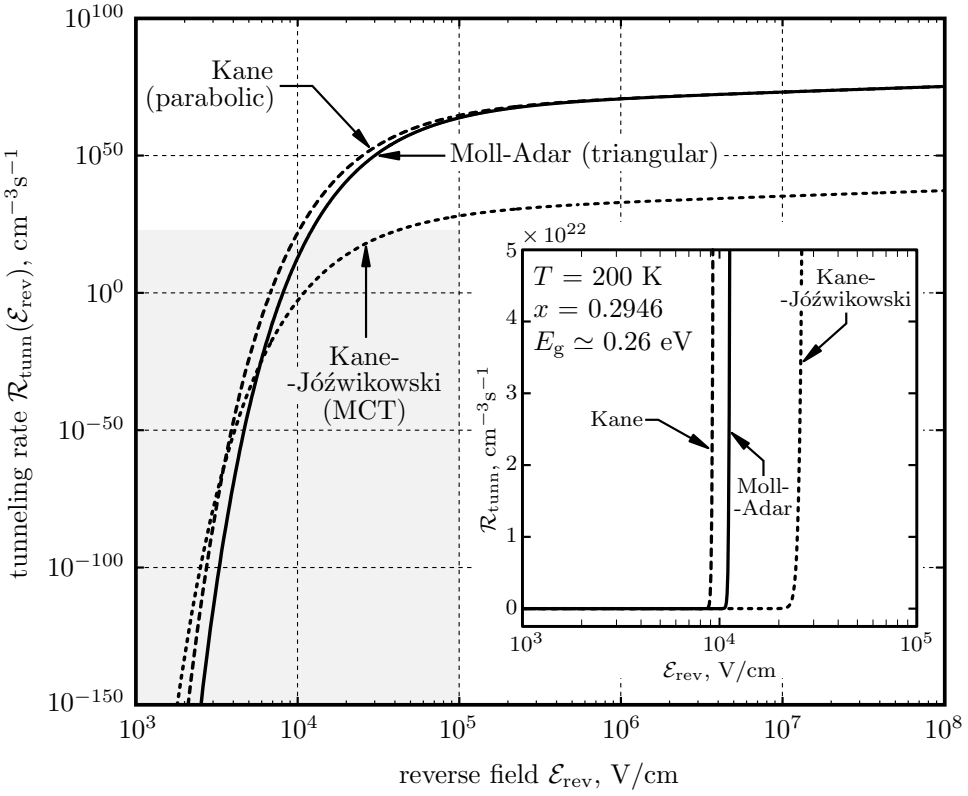

FIGURE 4.6. Comparison between different tunneling rate formulations computed for an MCT p-n junction under reverse bias at $T = 200$ K and with $x = 0.2946$: Kane curve (dashed) is represented by Eq. (4.65), with parameters of Eqs. (4.66) and (4.67). The so-called Moll-Adar curve (solid), instead, consists in the Adar rate (see Eq. (4.74)) which exploits the triangular tunneling barrier presented by Moll. Finally, the Kane-Jóźwikowski model (dotted) is computed with Eq. (4.65) and parameters expressed by Eqs. (4.72) and (4.73). The grey area represents the domain of inset plot. Inset: same plot in log-linear scale.

and

$$B_{\rm MCT}(x,T) = \frac{\pi\sqrt{m^*(x,T)\,E_{\rm g}^3(x,300\,{\rm K})}}{2\sqrt{2}q\hbar}\,, \tag{4.73}$$

where $m^*(x,T)$ and $E_{\rm g}(x,T)$ are given, respectively, by Eqs. (2.14) and (2.17), and where $A_{\rm MCT}(x,T)$ and $B_{\rm MCT}(x,T)$ have to be inserted into Eq. (4.65).

Figure 4.6 shows the number of electron/hole pairs generated by tunneling per unit volume per second (tunneling rate $\mathcal{R}_{\rm tunn}$) calculated as a function of the reverse electric field in an MCT p-n junction at $T = 200$ K and with $x = 0.2946$ via different models. In the inset a log-linear version of the same plot has been included to appreciate the different "threshold" of each curve. The Moll-Adar model refers to the WKB-based Moll formalism for a triangular barrier (see Eq. (4.39)) from which

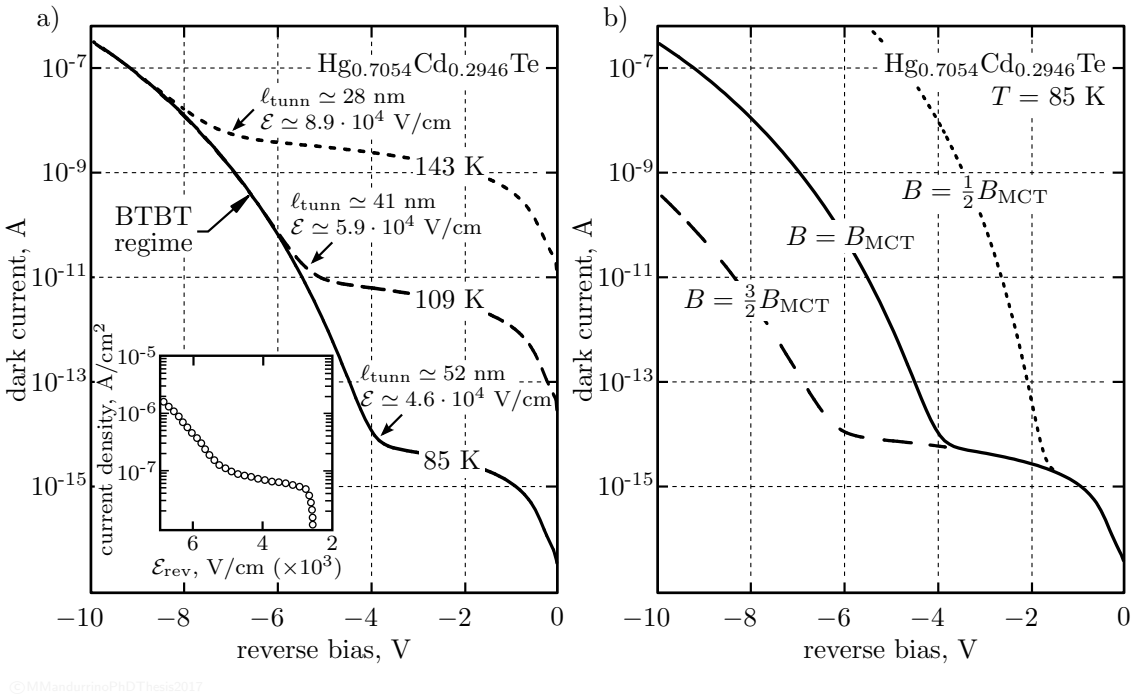

FIGURE 4.7. (a) Calculated electrical behavior of an MCT *p-n* (abrupt) junction under reverse bias in dark conditions at $T = 85$ K, 109 K and 143 K including Kane-Jóźwikowski model. Notice the BTBT onset occurring at different bias (field). Inset: measured dark characteristics of an $Hg_{0.757}Cd_{0.243}Te$ MIS detector at $T = 55$ K (adapted and re-edited from Ref. [19]). (b) For the same structure of (a), calculated (dark) electrical behavior at fixed $T = 85$ K and different critical fields B.

R. Adar [167] deduced the rate expression

$$\mathcal{R}_{\text{Moll-Adar}} = \frac{q^2 \sqrt{m^*} \mathcal{E}^2}{4\sqrt{2}\pi^2 \hbar^2 \sqrt{E_g}} \exp\left(-\frac{4\sqrt{2m^* E_g^3}}{3q\hbar\mathcal{E}} \right). \tag{4.74}$$

The Kane-Jóźwikowski curve, instead, is mixing Kane model (see Eq. (4.65)) with MCT parameters reported in Eqs. (4.72) and (4.73) [168]. Notice that by tuning the critical field B also the threshold rate (i.e. the point where $\mathcal{R}_{tunn}$ starts to grow) changes significantly (inset of Figure 4.6). While for standard parabolic and triangular barriers the rate onset lies approximately at $\mathcal{E}_{rev} \simeq 10^4$ V/cm, the Kane formula with $A_{MCT}(x,T)$ and $B_{MCT}(x,T)$ MCT parameters has a critical field at least two times greater.

In Figure 4.7 we represented some examples of calculated $I(V)$ characteristics for an MCT-based abrupt *p-n* diode in dark conditions via the simulation procedure that has been explained in Section 1.5 and through the Kane-Jóźwikowski model. In panel (a) device temperature is tuned and the BTBT onset changes consequently while in (b) T has been fixed and the critical field is varied in order to highlight the dependence of BTBT current raising with respect to B. Moreover, in the inset of panel (a) we included also an example of measured characteristics of an MIS

photodetector (taken from Ref. [19]) showing the same qualitative trend plotted both in (a) and (b).

4.1.1.2 Trap-assisted tunneling (TAT): Hurkx formalism

Trap-assisted tunneling (TAT) in electronic devices is known since 1958, when T. Yajima and L. Esaki [169] were investigating Ge p-n junctions in the energy regime where direct transitions are not allowed. They supposed that the diode excess current was in part due to transport mechanisms assisted by defects in the crystal. Few years later, in 1961, A. G. Chynoweth *et al.* [16] developed a first TAT model to explain the behavior of Si-based Esaki diodes.

Here we are interested in inter-band TAT processes occurring in GaN-based LEDs. As we mentioned in Section 4.1, TAT is a three-step mechanism composed by tunneling, capture (or trapping) and recombination (see panel (b) of Figure 4.1). However, Hurkx pointed out that there are no quantum criticisms in describing the whole process as a unique SRH event [165,166,170]. From this perspective TAT can be seen as a recombination mechanism mediated by SRH centers, that in III-nitrides can be lattice (deep) defects or dopant (shallow) impurities. The recombination rate is weighted by the carrier probability density at the trap site $|\Psi(z_{\text{trap}})|^2$ and, at the same time, by the probability of scattering occurring between the tunneling particle and the trap itself. This scattering event can be globally described by the capture cross section

$$\begin{aligned} \sigma_{n,p} &= \frac{c_{n,p}}{v_{\text{th}_{n,p}}} \\ &= \left(\tau_{\text{SRH}_{n,p}} N_{\text{trap}} v_{\text{th}_{n,p}}\right)^{-1} , \end{aligned} \tag{4.75}$$

where N_{trap} is the trap density, $c_{n,p}$ are the carrier capture rates, $\tau_{\text{SRH}_{n,p}}$ and

$$v_{\text{th}_{n,p}} = \sqrt{\frac{2E_{\text{kin}}}{m^*_{n,p}}} = \sqrt{\frac{3k_{\text{B}}T}{m^*_{n,p}}} \tag{4.76}$$

their SRH lifetimes and thermal velocities, given by Eqs. (1.47) and (1.34).

The formalism developed by Hurkx in 1989 considerably simplifies the physics behind TAT in a very smart and intuitive way: to describe tunneling via defects a modified form of the usual SRH rate is used, where the new trap cross section is multiplied by a field-effect factor (following the approach of G. Vincent *et al* [171]) which, in turn, accounts for the wavefunction probability density calculated under WKB approximation. This strategy returns a very simple formulation which offers a model highly suitable for the implementation into numerical solvers. For this reason we propose Hurkx theory at this point of the work, leaving more complicated treatments of TAT in the next sections, were additional quantum effects and nonlocality properties will be introduced in our description.

Regarding Figure 4.8, the mechanism considered by Hurkx seems to differ from the TAT we already described. However, as we will see talking about the MPE theory, after some hypotheses we can safely bring the case of Hurkx to our case by

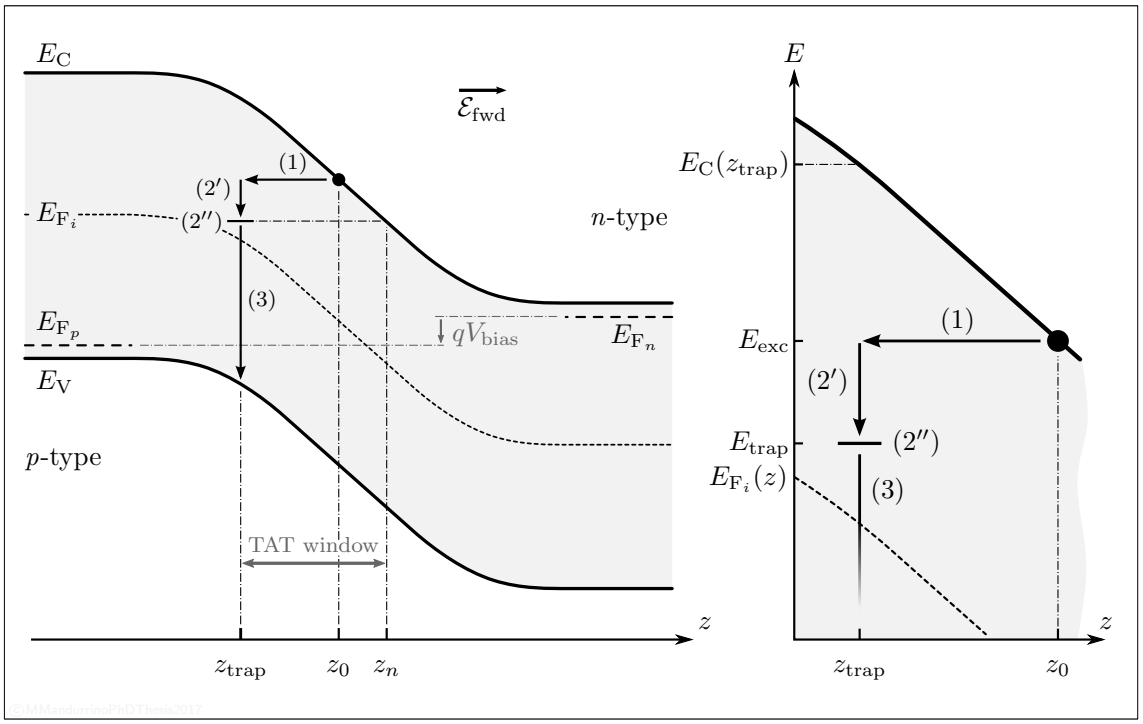

FIGURE 4.8. Left: scheme of Hurkx trap-assisted tunneling (TAT) in a forward biased p-n junction. The process is composed as follows: (1) pure electron tunneling from CB to a state with energy E_{exc} at position z_{trap}; (2′) energy relaxation down into the trap with energy E_{trap}; (2″) trapping; (3) recombination with VB hole. Right: zoom of the process scheme around the trap.

simply introducing the role of phonons in the capturing step. Now let us suppose to have an electron in conduction band with a finite probability of tunneling towards the p-side of a diode under forward bias. Its energy non necessarily coincides with the trap energy E_{trap} so we can assume that, in general, the electron energy lies in the range $\Delta E_n = E_C(z_{trap}) - E_{trap}$. Calling this energy E_{exc}, then a relaxation process from the excited state to the trap may occur. So, the electron is captured by the trap with a certain rate c_n and then it recombines with a hole in valence band. The symmetrical process occurs for holes, where $\Delta E_p = E_{trap} - E_V(z_{trap})$. Notice that, since LEDs normally operate at high temperature (> 300 K) there are always allowed valence states in the p-side for a tunneling electron, even in case of non-degenerate semiconductors.

The idea of Hurkx was to introduce proper *field-effect enhancement functions* $\Gamma_{n,p}(\mathcal{E})$ into Eq. (4.75) in order to account for the higher carrier concentration in the depletion region due to such kind of tunneling. Thus, in the Hurkx formalism we have

$$\begin{aligned} \sigma_{n,p}(\mathcal{E}) &= (\Gamma_{n,p}(\mathcal{E}) + 1)\,\sigma_{n,p} \\ &= \frac{\Gamma_{n,p}(\mathcal{E}) + 1}{\tau_{\mathrm{SRH}_{n,p}}\, N_{trap}\, v_{\mathrm{th}_{n,p}}} \\ &= \left(\tau_{n,p}(\mathcal{E})\, N_{trap}\, v_{\mathrm{th}_{n,p}}\right)^{-1} , \end{aligned} \tag{4.77}$$

with

$$\tau_{n,p}(\mathcal{E}) = \frac{\tau_{\mathrm{SRH}_{n,p}}}{\Gamma_{n,p}(\mathcal{E}) + 1}\,, \tag{4.78}$$

under the hypothesis of constant electric field and supposing to apply a local correction only at the trap site. In fact

$$\begin{cases} \Gamma_n(\mathcal{E}) + 1 = \dfrac{e_n}{e_{0n}} = \dfrac{n(z_{\mathrm{trap}})}{n_0(z_{\mathrm{trap}})} \\ \Gamma_p(\mathcal{E}) + 1 = \dfrac{e_p}{e_{0p}} = \dfrac{p(z_{\mathrm{trap}})}{p_0(z_{\mathrm{trap}})}\,, \end{cases} \tag{4.79}$$

being $e_{n,p}$ and $e_{0,n,p}$ the emission rates in standard conditions and under zero-field conditions and similarly for carrier concentrations $n(z_{\mathrm{trap}})$, $p(z_{\mathrm{trap}})$ and $n_0(z_{\mathrm{trap}})$, $p_0(z_{\mathrm{trap}})$. Eqs. (4.79) can be also expressed in terms of capture rates, which is more useful in our case, by recalling the *balance equations* [172]

$$\begin{cases} c_n = \dfrac{e_n}{n(z_{\mathrm{trap}})}\,\dfrac{f_{\mathrm{trap}}}{1 - f_{\mathrm{trap}}} \\ c_p = \dfrac{e_p}{p(z_{\mathrm{trap}})}\,\dfrac{1 - f_{\mathrm{trap}}}{f_{\mathrm{trap}}}\,, \end{cases} \tag{4.80}$$

where f_{trap} represents the trap occupation function.

As already mentioned, for the TAT rate we can exploit the SRH-like expression in Boltzmann statistics (see Eq. (1.46))

$$\mathcal{R}_{\mathrm{TAT}} = \frac{np - n_i^2}{\tau_p(\mathcal{E})\left(n + n_i\,\mathrm{e}^{\frac{E_{\mathrm{trap}} - E_{\mathrm{F}_i}}{k_{\mathrm{B}}T}}\right) + \tau_n(\mathcal{E})\left(p + n_i\,\mathrm{e}^{\frac{E_{\mathrm{F}_i} - E_{\mathrm{trap}}}{k_{\mathrm{B}}T}}\right)}\,, \tag{4.81}$$

where n_i is the intrinsic density and E_{F_i} the intrinsic Fermi level. From the definition of the field-effect functions, with the use of balance equations (4.80), it follows that TAT rate reduces to [166]

$$\mathcal{R}_{\mathrm{TAT}} = N_{\mathrm{trap}}\frac{c_n\,c_p\,n(z_{\mathrm{trap}})\,p(z_{\mathrm{trap}}) - e_n\,e_p}{c_n\,n(z_{\mathrm{trap}}) + c_p\,p(z_{\mathrm{trap}}) + e_n + e_p}\,. \tag{4.82}$$

To compute the field-effect functions, since the trap position is *a priori* unknown, let us to rewrite $n(z_{\mathrm{trap}})$ as $n_{\mathrm{trap}}(z)$ (and, similarly, $p(z_{\mathrm{trap}}) \equiv p_{\mathrm{trap}}(z)$ for holes). This simple change in notation is useful to appreciate the local carrier concentration variation with respect to spatial coordinate. In the same way we indicate the standard concentrations $n_0(z_{\mathrm{trap}})$ and $p_0(z_{\mathrm{trap}})$ as $n_{0,\mathrm{trap}}(z)$ and $p_{0,\mathrm{trap}}(z)$ respectively. By using the Airy asymptotic trend of WKB wavefunctions under triangular barrier approximation we have, for electrons,

$$n_{\mathrm{trap}}(z) = n_{0,\mathrm{trap}}(z) - \int_{z_n}^{z_0} \frac{\mathrm{d}n(z)}{\mathrm{d}z}\bigg|_{z_0} P(z_0)\,\mathrm{d}z_0\,, \tag{4.83}$$

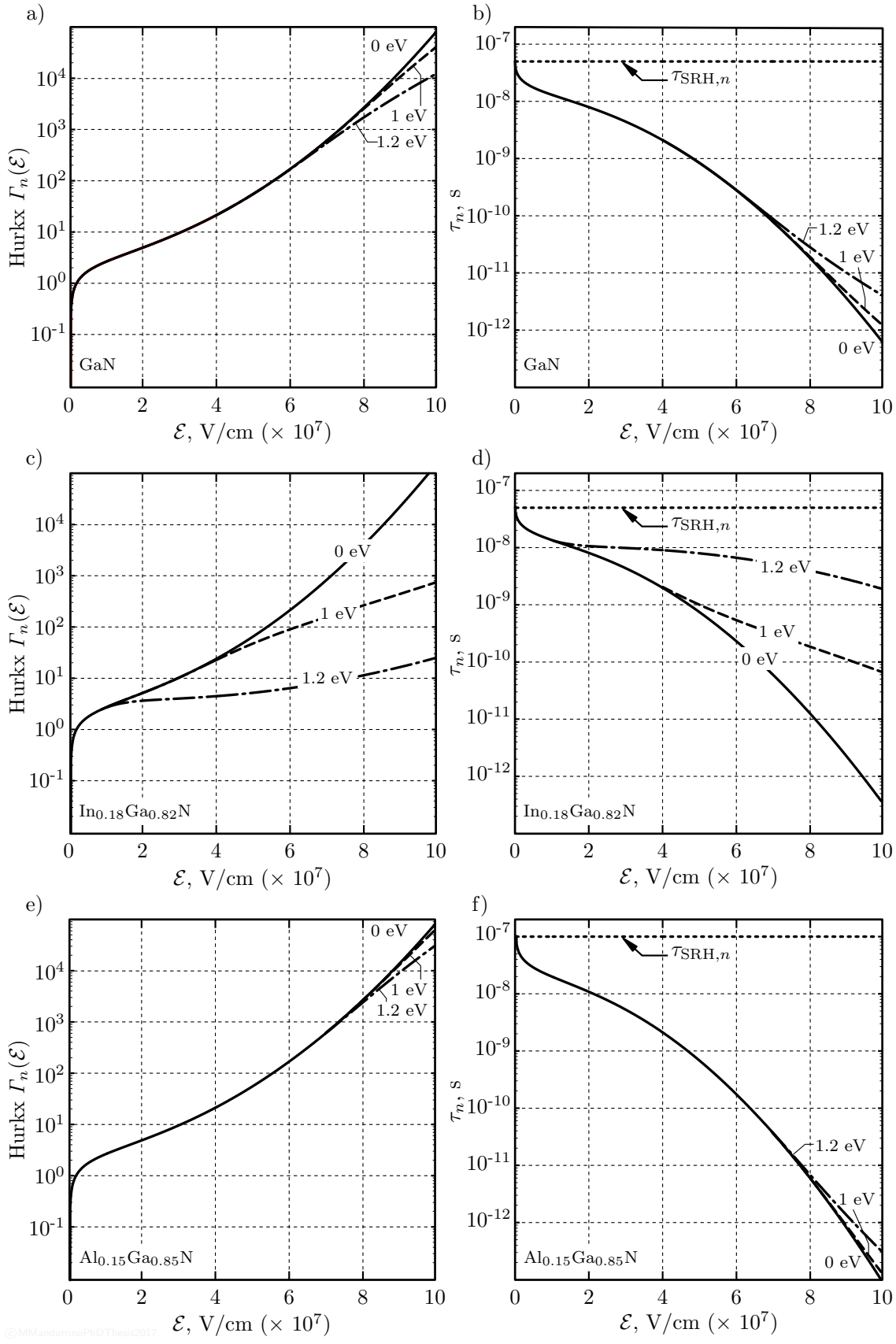

FIGURE 4.9. Hurkx field-effect function $\Gamma_n(\mathcal{E})$ and electron lifetime $\tau_n(\mathcal{E})$ in GaN, InGaN and AlGaN calculated via Eq. (4.86) at 300 K and for different energies $E_{\mathrm{trap}} - E_{\mathrm{F}_i}$.

where the integration is performed along the range of all the possible initial states from z_n to z_0 (see labels in Figure 4.8). The function

$$P(z_0) = \frac{\mathrm{Ai}^2\left(\sqrt[3]{\frac{2q\mathcal{E}m_n^*}{\hbar^2}}\,(z_{\mathrm{trap}} - z_0)\right)}{\mathrm{Ai}^2(0)} \tag{4.84}$$

represents the probability to have electrons tunneling from z_0 to z_{trap}, where the denominator holds for $z_0 = z_{\mathrm{trap}}$. It is worth noting that Hurkx formalism accounts only for electron energies $E_{\mathrm{exc}} \geq E_{\mathrm{trap}}$, neglecting electrons tunneling with energy lower than the trap and their subsequent capture via excitation. This requirement will be partially relaxed within the MPE theory which includes lattice oscillations at the trap site.

By using Eq. (4.83) and the second equalities in Eq. (4.79) one may write

$$e_n = e_{0n}\left(1 + \frac{1}{k_{\mathrm{B}}T}\int_0^{\Delta E_n} \mathrm{e}^{\frac{E}{k_{\mathrm{B}}T}}\,\frac{\mathrm{Ai}^2\left(\frac{2m_n^*E}{\hbar^2}\sqrt[3]{\frac{\hbar^4}{4q^2\mathcal{E}^2m_n^{*\,2}}}\right)}{\mathrm{Ai}^2\,(0)}\,\mathrm{d}E\right)\,, \tag{4.85}$$

which transforms into c_n according to Eq. (4.80). Now comparing the first equalities in Eq. (4.79) with Eq. (4.85) we have

$$\Gamma_{\mathrm{Hurkx}_{n,p}}(\mathcal{E}) = \frac{\Delta E_{n,p}}{k_{\mathrm{B}}\,T}\int_0^1 \exp\left[\frac{\Delta E_{n,p}}{k_{\mathrm{B}}\,T}u + \Lambda_{n,p}(\mathcal{E})\,u^{\frac{3}{2}}\right]\mathrm{d}u\,, \tag{4.86}$$

where u is an energy auxiliary integration variable normalized in the range 0 to 1 and with

$$\Lambda_{n,p}(\mathcal{E}) = -\frac{4\sqrt{2m_{n,p}^*\Delta E_{n,p}^3}}{3q\hbar\mathcal{E}} \tag{4.87}$$

the exponential argument of the tunneling probability through a triangular barrier $\mathcal{T}_{\triangle}(E)$ (see Eq. (4.39)) where E_{g} here is replaced by the energy differences $\Delta E_{n,p}$. Notice also that at high forward electrostatic fields the band bending is such that E_{trap} can lie below the CB minimum $E_{\mathrm{C,min}}$ in neutral n-type region. In this case we define $\Delta E_n = E_{\mathrm{C}} - E_{\mathrm{C,min}}$. Symmetrically, for holes, $\Delta E_p = E_{\mathrm{V,max}} - E_{\mathrm{V}}$.

In Figure 4.9 we plotted room-temperature Hurkx field-effect function – panels (a), (c) and (e) – and TAT lifetime – panels (b), (d) and (f) – for electrons in GaN, InGaN and AlGaN as functions of the local average electrostatic field $\mathcal{E}$ and for different trap energies, defined from the intrinsic Fermi energy via $E_{\mathrm{trap}} - E_{\mathrm{F}_i}$. As occurs in case of SRH recombination, while the defect level shifts towards E_{F_i} (high ΔE_n) the field-enhancement function increases and, in turn, the tunneling lifetime decreases (i.e. the TAT rate is maximized). Notice also that under low-field conditions $\Gamma_n(\mathcal{E}) \to 0$ and both $\mathcal{R}_{\mathrm{TAT}}$ and $\tau_n(\mathcal{E})$ tend to the SRH regime.

4.2 A Novel BTBT Formulation for MCT IRPDs

As anticipated in the introduction, the present section deals with a particular model of direct band-to-band tunneling conceived by the author to model BTBT in

reverse bias *p-i-n* HgCdTe photodetectors. The model, originally presented in my Master's thesis [45] and for the first time applied to real MCT IRPD structures in a 2015 paper we published in the *Journal of Electronic Materials* (see Ref. [173]), takes inspiration by Kane formalism and merges some recent results coming from different areas of electronics applications.

All the models we discussed about direct tunneling are based on physical assumptions valid for *p-n* junctions. However, since most of modern MCT IR photodetectors are essentially constituted by *p-i-n* structures the need of a specific description arises. With this motivation the author proposed a novel formulation strictly suitable for this kind of devices. We focused our reasoning on the electric field, because this is the critical quantity which makes the major difference between the depletion region of a *p-n* diode with respect to that of a *p-i-n* diode. Moreover, the Kane approximation of constant field allows us to make some physical assumptions.

As already pointed out by W. Vandenberghe *et al.* in their theoretical work [174], the electric field in the depletion region can be splitted into an *effective* and an *average* component. Soon after A. S. Verhulst *et al.* [175] and K. Ahmed *et al.* [176] have exploited this hypothesys attaining the same result: the dependence of Kane BTBT rate by $\mathcal{E}^2$ can be rewritten in *p-i-n* diodes as

$$\mathcal{E}^2 \simeq \mathcal{E} \cdot \mathcal{E}_{\mathrm{av}}^{(D-1)} \,, \tag{4.88}$$

i.e. as the product of the effective field times a certain power of the average field, where D is an adimensional empirical parameter to be determined *ad hoc* for each simulated structure. This allows to rewrite the BTBT generation rate in the form

$$G_{pin} = A \cdot \mathcal{E} \cdot \mathcal{E}_{\mathrm{av}}^{(D-1)} \cdot \exp\left(-\frac{B}{\mathcal{E}_{\mathrm{av}}}\right) . \tag{4.89}$$

Let us now derive the above mentioned novel formulation starting from the most general expression of the average field depending on the energy gap and tunneling path that can be found in Eq. (4.1) and which reads

$$\mathcal{E}_{\mathrm{av}} = \frac{E_{\mathrm{g}}}{q\,\ell_{\mathrm{tunn}}} \,. \tag{4.90}$$

For what concerns instead the effective field, and referring to the *p-i-n* junction in Figure 4.10, we know that its theoretical maximum, due to space charge regions, reaches

$$|\mathcal{E}_{\mathrm{max}}| = \frac{qN_{\mathrm{A}}}{\epsilon} z_p \,, \tag{4.91}$$

where z_p is the depletion region extension in the *p*-side. Supposing now to have a constant effective field $\mathcal{E}$ across the depletion region, pinched at $|\mathcal{E}_{\mathrm{max}}|$ such that

$$\mathcal{E}(z)|_{\mathrm{dep}} = |\mathcal{E}_{\mathrm{max}}| = \mathcal{E}_{\mathrm{av}} = \text{const.} \,, \tag{4.92}$$

then it straightforwardly follows

$$\ell_{\mathrm{tunn}} = \frac{\epsilon E_{\mathrm{g}}}{q^2 N_{\mathrm{A}} z_p} \,, \tag{4.93}$$

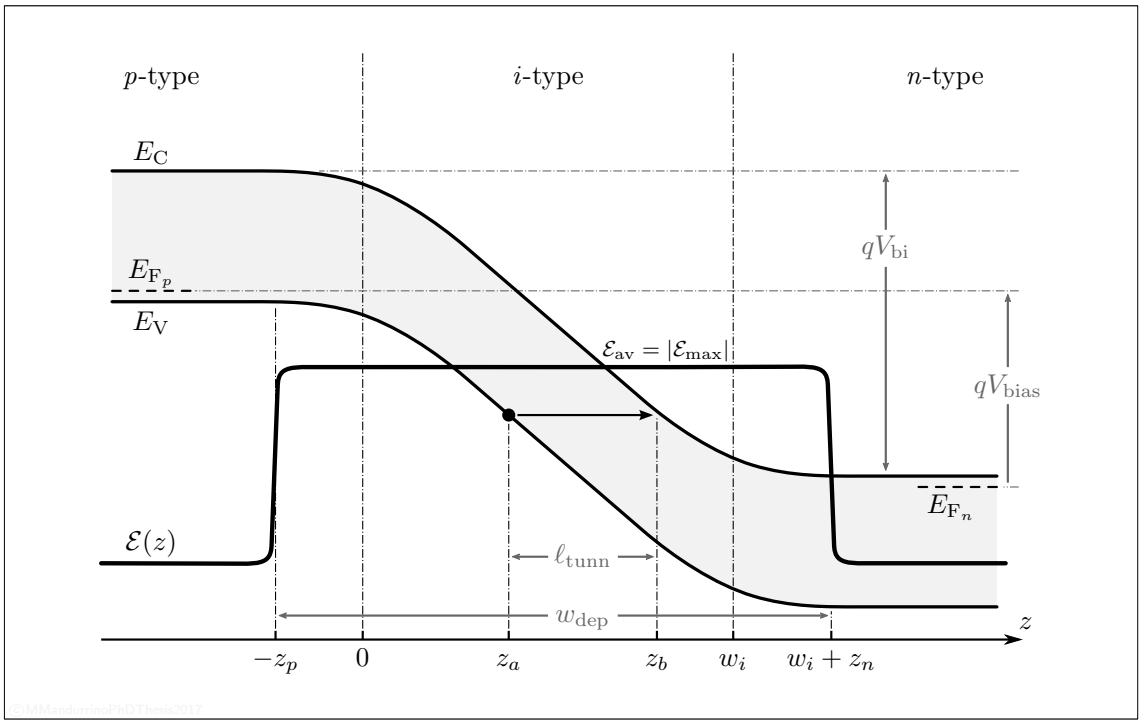

FIGURE 4.10. Band diagram of a narrow gap p-i-n homojunction in reverse bias conditions. The p- and n-type regions are doped by N_A and N_D, respectively, while the intrinsic region is non-intentionally (or lightly) doped by N_i. The scheme also represents, superimposed, the assumed trend of the electric field across the depletion region (energies and distances are not to scale).

where z_p has to be determined.

To obtain a close-form expression of z_p we recall the *charge neutrality law*

$$\begin{aligned}\frac{2\epsilon\left(V_{bi}+V_{bias}\right)}{q}+N_i w_i^2-\frac{N_i^2 w_i^2}{N_D} &= x_p^2\left(\frac{N_A^2}{N_D}+N_A\right)\\ &\quad +x_p\left(-\frac{2N_A N_i w_i}{N_D}+2N_A w_i\right)\end{aligned} \tag{4.94}$$

from which z_p stems as

$$\begin{aligned}z_p &= \frac{\sqrt{qN_AN_D\left[qw_i^2\left(N_D-N_i\right)\left(N_A+N_i\right)+2\mathcal{E}w_{dep}\epsilon\left(N_A+N_D\right)\right]}}{qN_A\left(N_A+N_D\right)}\\ &\quad +\frac{qw_iN_A\left(N_i-N_D\right)}{qN_A\left(N_A+N_D\right)}\\ &\overset{N_A\gg N_D}{\simeq}\sqrt{\frac{2\epsilon N_i\mathcal{E}w_{dep}}{qN_A^2}},\end{aligned} \tag{4.95}$$

where the relation $V_{bi}+V_{bias}=\mathcal{E}w_{dep}$ has been used. Notice that z_p is (obviously) field-dependent. Thus, also the tunneling path becomes even more explicitly a function of the electric field, since

$$\ell_{tunn}(\mathcal{E})=\frac{\epsilon E_g}{q^2N_Az_p(\mathcal{E})} \tag{4.96}$$

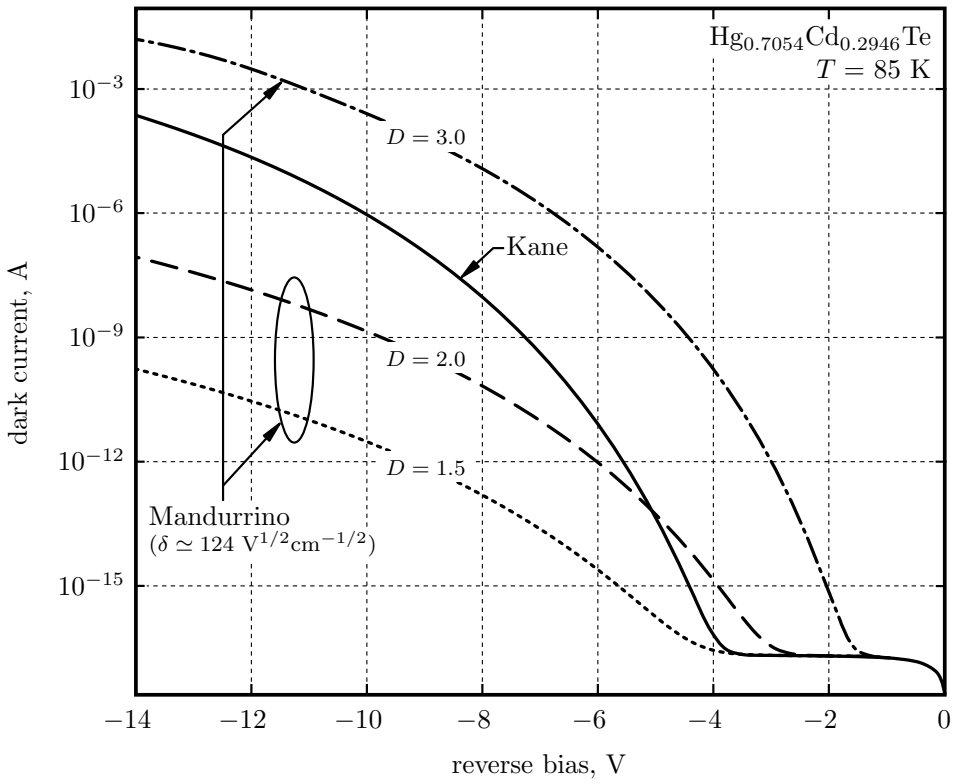

FIGURE 4.11. Comparison between simulated reverse dark currents in an $Hg_{0.7054}Cd_{0.2946}Te$ *p-i-n* photodiode at 85 K obtained implementing both Kane and Mandurrino model for different values of D and where rate parameters have always been set as A_{MCT} and B_{MCT}.

and, in turn,

$$\mathcal{E}_{\mathrm{av}}(\mathcal{E}) = \frac{E_{\mathrm{g}}}{q\,\ell_{\mathrm{tunn}}(\mathcal{E})} = \frac{qN_{\mathrm{A}}}{\epsilon} z_p(\mathcal{E}) \tag{4.97}$$

also the average field results to be a function of the constant and bias-dependent effective field $\mathcal{E}$. By plugging Eq. (4.95) into Eq. (4.97) one may find

$$\mathcal{E}_{\mathrm{av}}(\mathcal{E}) \simeq \delta\sqrt{\mathcal{E}}\,, \tag{4.98}$$

where

$$\delta \simeq \sqrt{\frac{2qN_i w_{\mathrm{dep}}}{\epsilon}}\,. \tag{4.99}$$

Then, according to the present new *p-i-n* formulation, the BTBT rate in Eq. (4.89) becomes

$$G_{\mathrm{Mandurrino}} = A \cdot \delta^{(D-1)} \cdot \mathcal{E}^{\left(\frac{D+1}{2}\right)} \cdot \exp\left(-\frac{B}{\delta\sqrt{\mathcal{E}}}\right), \tag{4.100}$$

which results in a more useful form under the numerical simulation perspective respecting the previous ones obtained by E. O. Kane, first, and K. Ahmed, subsequently. In fact, Eq. (4.100) depends on macroscopic quantities that should be

known in advance or, otherwise, that can be easily computed by the solver. Also notice that in the formalism N_i is indicated in place of the usual n_i notation to distinguish eventual non-intentional doping from native intrinsic concentration.

Figure 4.11 shows the results of a numerical test in which different reverse dark current trends are obtained implementing the Kane-Jóźwikowski model and the novel formulation of Eq. (4.100) with different values of the parameter D. The test is performed at 85 K on an $Hg_{0.7054}Cd_{0.2946}Te$ *p-i-n* structure composed by 2 µm-thick *p*-type region doped by $N_A = 10^{17}$ cm^{-3}, followed by a 3 µm-thick intrinsic layer with $N_i = 10^{14}$ cm^{-3} and then by a 2 µm-thick *n*-type region doped by $N_D = 10^{17}$ cm^{-3}. For the structure we chosen the parameter δ results to be equal to 124.172 V$^{1/2}$cm$^{-1/2}$ (this value has been obtained through the first equality in Eq. (4.95), i.e. without any approximation, being $N_A = N_D$).

As a final remark, it has to be said that the constant field approximation we invoked results well satisfied in *p-i-n* homojunctions, in general, and especially in HgCdTe diodes, so this assumption can be safely adopted for simulation purposes. Indeed, as already said, satisfactory agreements with experimental data taken on MCT pixels has been obtained in our paper of Ref. [173] (see Chapter 5).

4.3 MPE Theory for Defect-Assisted Tunneling

The concepts here exposed, as well as the brief description about Hurkx TAT model, constitute part of the theoretical introduction to the invited paper published in the 2015 special issue of the *Journal of Computational Electronics*, vol. 14, no. 2 [145], on the topic "simulation of GaN-based light-emitting diodes" (see the editorial presentation by J. Piprek [144]).

In Subsection 4.1.1.2 we have seen that Hurkx TAT model is a "local" description by definition, since all the quantities concurring to the field-effect functions are computed at the trap site. However a rigorous treatment would not disregard the non-local nature of tunneling, which is due to the different physical characteristics experienced by the particle at its initial state, during the transition (trapping included) and, finally, at the final state. Moreover, the Hurkx treatment does not account for some implications involved in the capture process such as lattice vibrations and phonon emission, thus its SRH-like rate may result too much generic until specific capture processes are included. Nevertheless, Hurkx approach has the great advantage to be so compact that its low computational cost has represented, in the last 30 years, one of the most advantageous solutions for many device designers.

An important step towards the description of proper trapping dynamics was settled in early 90s by A. Shenk, a researcher from Swiss Institute of Technology who was working on microscopic modeling of trap-assisted [177] and direct [178] tunneling in silicon. Based on the same principle of the previous model, according to which TAT can be modeled through a field-dependent enhancement of the SRH lifetime, he described the trapping mechanism in terms of lattice vibrations. Indeed, via non-radiative multiphonon emission (MPE) mechanisms part of the energy carried by an electron tunneling towards a defect is released to the trap and then dissi-

pated by the lattice. As we will see in detail, this explanation allows also tunneling transitions performed by electrons with energy lower than the trap energy. This is possible because no carrier excitation has to be invoked since lattice oscillations can justify, via energetic overlaps of the involved states, the trapping of such electrons. A symmetric reasoning can be inferred for holes tunneling from the p-type side of the junction towards defects in the n-side (under forward bias conditions). The last – but not less important – consideration that may arise about Schenk model before entering in the details concerns its applicability: although the formalism was conceived for silicon, MPE theory can be safely applied also to the GaN-based material system since its most crucial requirement, that is the condition about energies $\Delta E_{n,p}$ greater than the highest possible phonon energy, is still perfectly fulfilled in case of typical III-N (deep) impurities.

Starting from basic MPE theory (see, e.g., some famous works in Refs. [179,180]) the probability to have phonon emission can be expressed in the form

$$P_{\mathrm{MPE}_{n,p}}(\mathcal{E}, l) = \frac{(l \mp S)^2}{S} \exp\big(- S\,(2 f_{\mathrm{BE}} + 1)\big) \exp\left(-\frac{l\hbar\omega_0}{2k_{\mathrm{B}}T}\right) \mathcal{B}_l(\gamma)\,, \tag{4.101}$$

where the signs $\mp$ refer, respectively, to electrons and holes, l is the number of phonons emitted,

$$f_{\mathrm{BE}} = \frac{1}{\exp\left(\hbar\omega_0 / k_{\mathrm{B}}\,T\right) - 1} \tag{4.102}$$

and $\hbar\omega_0$ their Bose-Einstein occupation function and energy, respectively (where a monomodal emission is assumed),

$$\gamma = 2\,S\,\sqrt{f_{\mathrm{BE}}\,\left(f_{\mathrm{BE}} + 1\right)}\,, \tag{4.103}$$

$\mathcal{B}_l(\gamma)$ a l-th order Bessel function and, finally, S is the so-called Huang-Rhys factor [179]. We will return in more depth on this concept, however, at present it is enough to state that S gives a measure of how carrier and phonon are coupled. By simplifying the physics one can say that [181]

$$S \simeq \frac{|\,\langle \Psi_{\mathrm{b}}\,|\,U\,|\,\Psi_{\mathrm{b}}\rangle\,|^2}{2\,\left(\hbar\,\omega_0\right)^2}\,, \tag{4.104}$$

where U is the interaction potential and $|\Psi_{\mathrm{b}}\rangle$ is the bound state at z_{trap}, composed by the carrier-trap coupling. In other words S describes the strength of diagonal carrier-phonon coupling or, in turn, the average number of emitted phonons [182].

The hypotheses whereby Eq. (4.101) has been written are essentially two: the first one has already been mentioned and refers to the monomodal emission (also known as Einstein approximation). The second assumption concerns the bound state dynamics and goes under the name of Born-Oppenheimer approximation. Assuming that each lattice site experiences a proper thermal-induced oscillation, according to this approximation we are simply supposing that, in particular, the state $|\Psi_{\mathrm{b}}\rangle$ behaves like an harmonic oscillator described by the Hamiltonian

$$H_{\mathrm{b}} = -z\,\sqrt{2S}\,\hbar\omega_0 + \mathrm{const.} \tag{4.105}$$

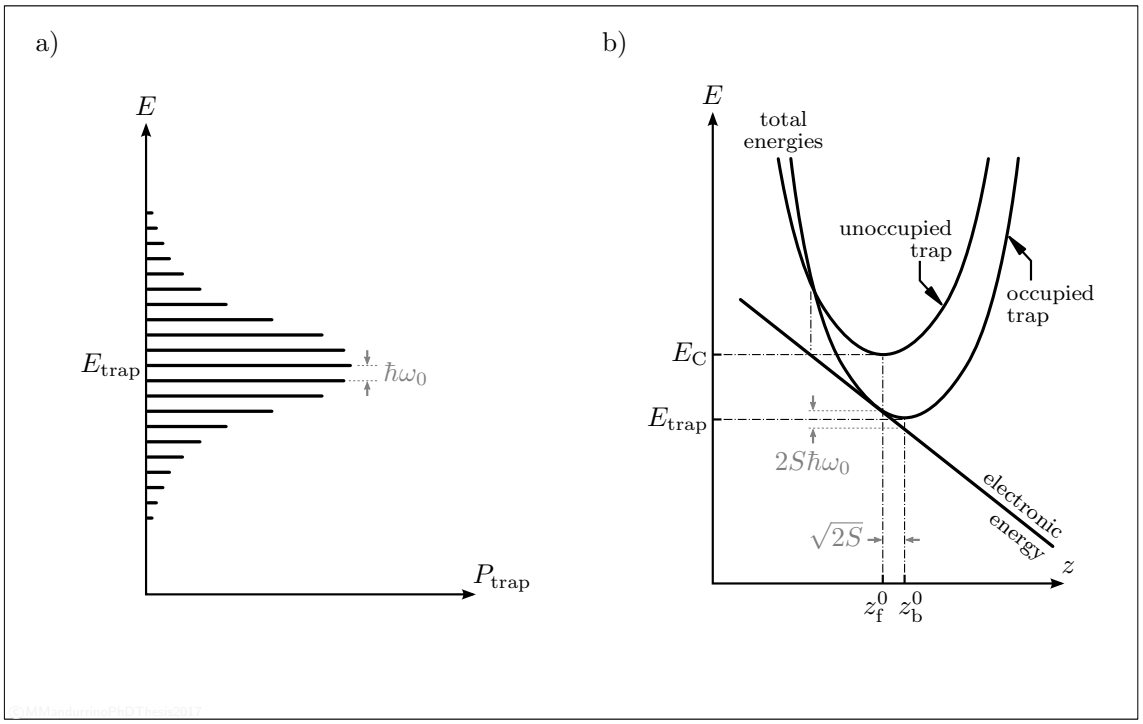

FIGURE 4.12. (a) Occupation probability $P_{\text{trap}}(E)$ of electronic quasi-levels in a trap with energy E_{trap}. (b) Configuration diagram showing electronic and total (electronic + elastic) parabolic energy in the non-radiative multiphonon emission (MPE) process, where z_{f}^0 and z_{b}^0 represent the equilibrium coordinates of the oscillating states before and after the capture, respectively.

Such state is oscillating around a point, z_{b}^0, which results spatially shifted by $\sqrt{2S}$ with respect to z_{f}^0, the center of oscillation of the unbound state (the system constituted by the "free" carrier plus the unoccupied trap). The Hamiltonian written in Eq. (4.105) determines a discrete spectrum of occupation probabilities versus energy

$$P_{\text{trap}}(\hbar\omega) = \left(1 - \mathrm{e}^{-\frac{\hbar\omega_0}{k_{\text{B}}T}}\right) \sum_{n=0}^{+\infty} \mathrm{e}^{-\frac{n\hbar\omega_0}{k_{\text{B}}T}} \, \mathcal{B}_l^2 \left(2\sqrt{S\left(n+\tfrac{1}{2}\right)}\right) , \tag{4.106}$$

where n labels the quasi-levels, which is maximum at E_{trap} and distributes symmetrically around it according to a bell-shaped function. Thus, each of the possible discrete states, separated by the quantized phonon energy $\hbar\omega_0$, defines a quasi-level in which the tunneling electron may fall. This is the reason why electrons with energies symmetrically distributed around E_{trap} have the possibility to be captured, even if with probabilities lower than at the center of the distribution.

This kind of trapping mechanism not only shifts the oscillation center along the lattice coordinate, but produces a shift also in the (electronic) energy domain. The amount of this difference equals twice the so-called *Frank-Condon* shift or relaxation energy $\epsilon_{\text{r}} = S\hbar\omega_0$. See Figure 4.12, panel (b), for a schematic description of the capture process and its characteristic physical parameters.

In order to obtain the rate of a process which includes the tunneling of an electron with energy $E_{\text{trap}} + l\hbar\omega_0$ from conduction band towards a trap with energy E_{trap} and

its MPE capture (see Figure 4.13) one has to account for the electronic wavefunction. In part this has already been done in Eq. (4.106) since the Bessel function $\mathcal{B}_l$ derives from a Bessel-Fourier series expansion of the wavefunction phase factor [183]. On the other hand the pure tunneling probability reads

$$P_{\mathrm{tunn}_{n,p}}(\mathcal{E},l) = \sqrt{\hbar\Theta_{\mathrm{C,V}}}\, K_{n,p}\left(\frac{\Delta E_{n,p} - l\hbar\omega_0}{\hbar\Theta_{\mathrm{C,V}}(\mathcal{E})}\right), \tag{4.107}$$

where l is the number of phonons that are emitted, $\Delta E_{n,p}$ has the usual meaning, $\Theta_{\mathrm{C,V}}$ is the electrooptical frequency such that

$$\Theta_{\mathrm{C,V}}(\mathcal{E}) = \sqrt[3]{\frac{q^2\,\mathcal{E}^2}{2\,\hbar\,m^*_{\mathrm{tunn}_{n,p}}}}, \tag{4.108}$$

with $m^*_{\mathrm{tunn}_{n,p}}$ the electron/hole tunneling effective masses, and

$$K(y) = \mathrm{Ai}'^2(y) - y\,\mathrm{Ai}^2(y) \tag{4.109}$$

is a WKB-like electrooptical function, written as dependent from first type Airy function and its derivative Ai and Ai′.

Thus, the total TAT+MPE capture rate is given by

$$c_{\mathrm{TAT+MPE}} = \sum_l P_{\mathrm{MPE}}(\mathcal{E},l)\, P_{\mathrm{tunn}}(\mathcal{E},l)\, f(E_{\mathrm{trap}} + l\hbar\omega_0), \tag{4.110}$$

where $f(E_{\mathrm{trap}} + l\hbar\omega_0)$ is the Fermi-Dirac probability of finding an occupied (initial) state in conduction band with energy $E_{\mathrm{trap}} + l\hbar\omega_0$. The Schenk field-effect function can be written as

$$\Gamma_{\mathrm{Schenk}_{n,p}}(\mathcal{E},l) = \frac{\sum\limits_{l\geq 0} \rho_{0_{\mathrm{C,V}}}(l)\, f_{\mathrm{trap}}(l)\, P_{\mathrm{tunn}_{n,p}}(\mathcal{E},l)\, P_{\mathrm{MPE}_{n,p}}(\mathcal{E},l)}{\sum\limits_{l\geq 0} \rho_{0_{\mathrm{C,V}}}(l)\, f_{\mathrm{trap}}(l)\, P_{\mathrm{MPE}_{n,p}}(\mathcal{E},l)}, \tag{4.111}$$

where $\rho_{0_{C,V}}$ are the zero-field conduction/valence density of states. By inserting Eqs. (4.101) and (4.107) into Eq. (4.111) the following expression can be obtained both for electrons and holes

$$\Gamma_{\mathrm{Schenk}_{n,p}}(\mathcal{E},l) = \frac{\pi \sum\limits_{l\geq 0} (l-S)^2\; \mathrm{e}^{-\frac{l\hbar\omega_0}{2k_{\mathrm{B}}T}}\, \mathcal{B}_l(\gamma)\sqrt{\hbar\Theta_{\mathrm{C,V}}(\mathcal{E})}\, K_{n,p}}{\beta_{n,p} \sum\limits_{l\geq \frac{\Delta E_{n,p}}{\hbar\omega_0}} (l-S)^2\; \mathrm{e}^{-\frac{l\hbar\omega_0}{2k_{\mathrm{B}}T}}\, \mathcal{B}_l(\gamma)\sqrt{l\hbar\omega_0 - \Delta E_{n,p}}}, \tag{4.112}$$

where $\beta_n = 3$ and $\beta_p = 1$.

In order to make Eq. (4.112) more adapt to the numerical implementation, Schenk provided in Ref. [177] two kinds of approximated formulas valid in the high-

and low-temperature regime and derived by supposing a high number of emitted phonons l such that

$$\frac{\Delta E_{n,p} - l\hbar\omega_0}{\hbar\Theta_{\mathrm{C,V}}(\mathcal{E})} \gg 1 \tag{4.113}$$

(this apparently removes the dependency of Γ on l). Such assumption allows to perform a two-step procedure: at the first stage one can replace summations with integrals and then develop functions $K_{n,p}$ and $\mathcal{B}_\ell$ in their series expansions, preserving only the asymptotic trend. Secondly, the arguments of resulting energy integrals, constituted by some exponential terms times opportune prefactors, undergo a second-order expansion. In this way, for electrons, one obtains

$$\begin{aligned}
\Gamma_{\mathrm{Schenk}_n}(\mathcal{E}) \overset{\text{low-}T}{\simeq} \frac{1}{3}&\left(1 + \frac{\left(\hbar\Theta_{\mathrm{C}}(\mathcal{E})\right)^{3/2}\sqrt{\Delta E_n - E_{\mathrm{opt}}(\mathcal{E})}}{E_{\mathrm{opt}}(\mathcal{E})\hbar\omega_0}\right)^{-1/2} \\
&\times \frac{\left(\hbar\Theta_{\mathrm{C}}(\mathcal{E})\right)^{3/4}\left(\Delta E_n - E_{\mathrm{opt}}(\mathcal{E})\right)^{1/4}}{2\sqrt{\Delta E_n E_{\mathrm{opt}}(\mathcal{E})}}\left(\frac{\hbar\Theta_{\mathrm{C}}(\mathcal{E})}{k_{\mathrm{B}}T}\right)^{3/2} \\
&\times \exp\left[\frac{-2\left(\Delta E_n - E_{\mathrm{opt}}(\mathcal{E})\right) + \hbar\omega_0 - k_{\mathrm{B}}T}{2\hbar\omega_0} + \frac{\Delta E_n + \frac{k_{\mathrm{B}}T}{2}}{\hbar\omega_0}\right. \\
&\qquad \left.\times \ln\left(\frac{\Delta E_n}{\epsilon_R}\right) - \frac{E_{\mathrm{opt}}(\mathcal{E})}{\hbar\omega_0}\cdot\ln\left(\frac{E_{\mathrm{opt}}(\mathcal{E})}{\epsilon_R}\right)\right] \\
&\times \exp\left[-\frac{4}{3}\left(\frac{\Delta E_n - E_{\mathrm{opt}}(\mathcal{E})}{\hbar\Theta_{\mathrm{C}}(\mathcal{E})}\right)^{3/2}\right] \\
&\times \exp\left(\frac{\Delta E_n - E_{\mathrm{opt}}(\mathcal{E})}{k_{\mathrm{B}}T}\right)
\end{aligned} \tag{4.114}$$

and

$$\begin{aligned}
\Gamma_{\mathrm{Schenk}_n}(\mathcal{E}) \overset{\text{high-}T}{\simeq} \frac{1}{3}&\left(1 + \frac{2\epsilon_{\mathrm{r}}k_{\mathrm{B}}T}{\left(\hbar\Theta_{\mathrm{C}}(\mathcal{E})\right)^{3/2}\sqrt{\Delta E_n - E_{\mathrm{opt}}(\mathcal{E})}}\right)^{-1/2} \\
&\times \frac{E_{0,\mathrm{act}} + \Delta E_n}{k_{\mathrm{B}}T}\left(\frac{\hbar\Theta_{\mathrm{C}}(\mathcal{E})}{\Delta E_n + \epsilon_{\mathrm{r}}}\right)^{3/2} \\
&\times \exp\left(\frac{E_{0,\mathrm{act}} - E_{\mathrm{act}}(\mathcal{E})}{k_{\mathrm{B}}T}\right) \\
&\times \exp\left[-\frac{4}{3}\left(\frac{\Delta E_n - E_{\mathrm{opt}}(\mathcal{E})}{\hbar\Theta_{\mathrm{C}}(\mathcal{E})}\right)^{3/2}\right] \\
&\times \exp\left(\frac{\Delta E_n - E_{\mathrm{opt}}(\mathcal{E})}{k_{\mathrm{B}}T}\right),
\end{aligned} \tag{4.115}$$

where the only new terms are the energies $E_{\mathrm{opt}}(\mathcal{E})$, $E_{0,\mathrm{act}}$ and $E_{\mathrm{act}}(\mathcal{E})$. These parameters are defined as

$$E_{\mathrm{opt}}(\mathcal{E}) = 2\frac{2\epsilon_{\mathrm{r}}k_{\mathrm{B}}T}{\left(\hbar\Theta_{\mathrm{C}}(\mathcal{E})\right)^{3/2}}\left(\sqrt{\frac{\left(2\epsilon_{\mathrm{r}}k_{\mathrm{B}}T\right)^2}{\left(\hbar\Theta_{\mathrm{C}}(\mathcal{E})\right)^3} + \Delta E_n + \epsilon_{\mathrm{r}}} - \frac{2\epsilon_{\mathrm{r}}k_{\mathrm{B}}T}{\left(\hbar\Theta_{\mathrm{C}}(\mathcal{E})\right)^{3/2}}\right) - \epsilon_{\mathrm{r}}\,, \tag{4.116}$$

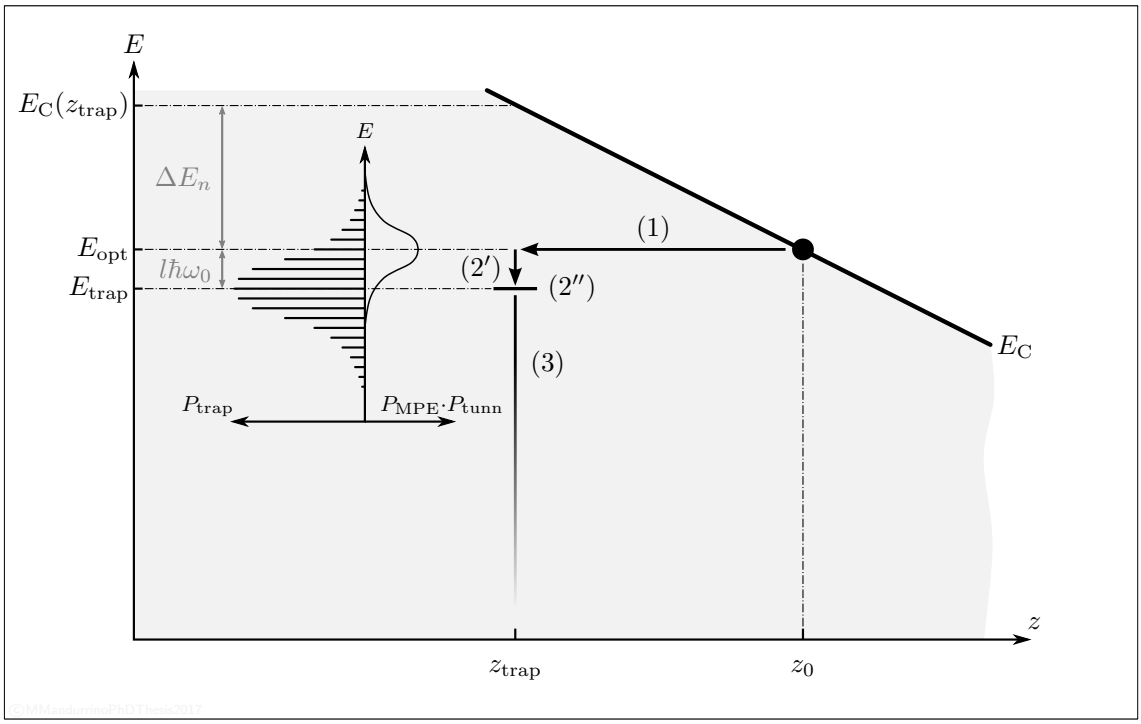

FIGURE 4.13. Schematic representation (not to scale) of electron trap-assisted tunneling (TAT) including multiphonon emission (MPE) as considered in the present work (and also by Schenk formalism [177]): TAT (1), MPE of l phonons (2′), trapping (2″) and recombination (3). The meaning of labels is explained in the text. Notice also that the overlap between $P_{\mathrm{MPE}} \cdot P_{\mathrm{tunn}}$ and P_{trap} determines a non-null probability of TAT even for electron energies below E_{trap}.

$$E_{0,\mathrm{act}} = \frac{(\Delta E_n - \epsilon_\mathrm{r})^2}{4\epsilon_\mathrm{r}}\,, \tag{4.117}$$

valid for $\mathcal{E} = 0$, and

$$E_{\mathrm{act}}(\mathcal{E}) = \frac{(E_{\mathrm{opt}}(\mathcal{E}) - \epsilon_\mathrm{r})^2}{4\epsilon_\mathrm{r}}\,. \tag{4.118}$$

Figure 4.14 shows in panel (a) the plot of the approximated Schenk field-effect formulas as reported in Eqs. (4.114) and (4.115). Panel (b), instead, represents some examples of related electron lifetimes, calculated at different temperature or relaxation energy $\epsilon_\mathrm{r} = S\hbar\omega_0$. Both panels refer to a sample of silicon where the electric field is directed along the $\langle 111 \rangle$ axis.

Apart from parameters $E_{0,\mathrm{act}}$ and $E_{\mathrm{act}}(\mathcal{E})$, which depend on some other already known quantities, what deserves particular attention is the *optimum transition energy* $E_{\mathrm{opt}}(\mathcal{E})$. Stemming from the series expansion we discussed just above, it derives from the product of probabilities $P_{\mathrm{MPE}_{n,p}}(\mathcal{E}, l)$ and $P_{\mathrm{tunn}_{n,p}}(\mathcal{E}, l)$ appearing in the numerator of Eq. (4.111). So, it results from a combination of an MPE term, whose probability is roughly represented by a decaying exponential of argument $-l\hbar\omega_0$ (see Eq. (4.101)) with a maximum when $l = 0$, i.e. at E_{trap} (the most probable transition, as already observed, should occur towards the state which implies the lowest number of emitted phonons), and a pure tunneling term, whose probability instead

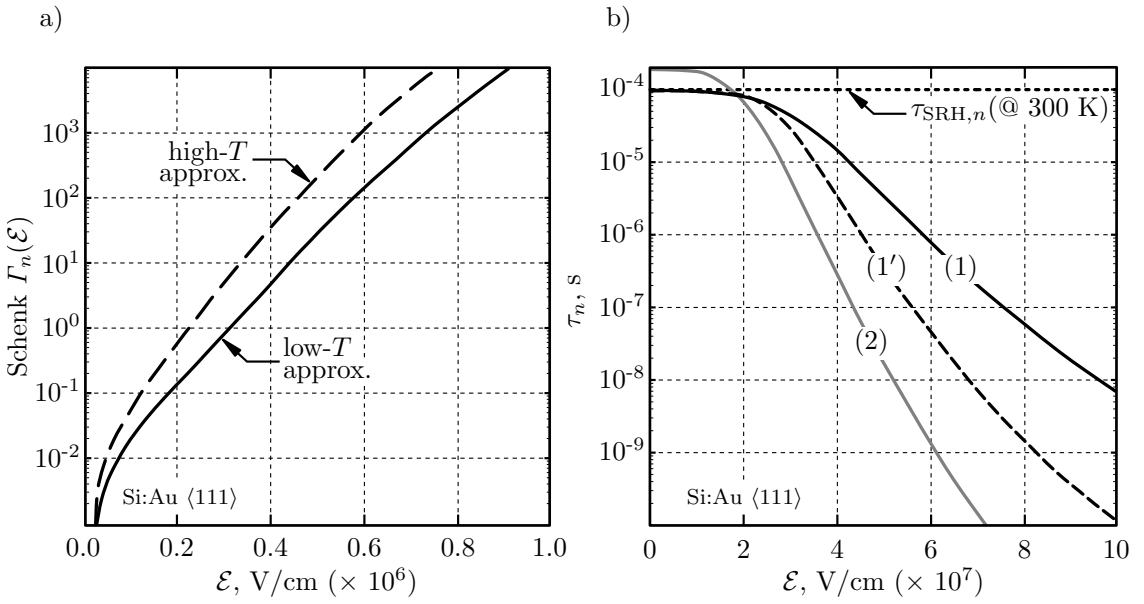

FIGURE 4.14. (a) Schenk field-effect function $\Gamma_n(\mathcal{E})$ in Si ⟨111⟩ calculated both under high- and low-T approximation. (b) Electron lifetime for the same material with $\Delta E_n = 0.55$ eV, $S\hbar\omega_0 = 238$ meV, $T = 300$ K (1) or $T = 200$ K (2) and with $\Delta E_n = 0.55$ eV, $S\hbar\omega_0 = 130$ meV, $T = 300$ K (1′). A similar trend of (1′) can be obtained by keeping constant $S\hbar\omega_0$ and increasing S at the expense of ω_0. Adapted and re-edited from Ref. [177].

is maximum for $\Delta E_{n,p} = 0$ (see Eq. (4.107)), since tunneling is maximized when the trap state lies close to the band edge or, in other words, when the tunneling path is minimized. This probability overlap translates into a bell-shaped function of the energy (not to be confused with the trend of P_{trap}) which is maximum when $E = E_{\mathrm{opt}}(\mathcal{E})$. Thus, Schenk $\Gamma_{n,p}$ functions follow the most probable tunneling path, occurring at the most probable transition energy $E_{\mathrm{opt}}(\mathcal{E})$, where $E_{\mathrm{trap}} < E_{\mathrm{opt}} < E_{\mathrm{C}}$ for electrons and $E_{\mathrm{V}} < E_{\mathrm{opt}} < E_{\mathrm{trap}}$ for holes. Typically this value is located close to the trap energy for high fields or low temperatures and migrates towards the band edges otherwise.

Despite Eqs. (4.114) and (4.115) are quite complicated due to the fact that they are accounting also for phonon emission, Schenk formalism overcomes the difficulty in computing the energy integral within Hurkx field-factor, resulting to be easily implementable into numerical solvers as well as the Hurkx one. However, as stated many times in this work, one of the most important features of TAT consists in the non-local nature of tunneling carriers wavefunction. But, unfortunately, not even MPE can be considered a non-local formalism. Let us see why. First of all one has to note that Schenk model is based on a constant-field assumption, essential to ensure phonon emission: in fact, quantum uncertainty imposes to MPE the requirement to have a bound state $|\Psi_{\mathrm{b}}\rangle$ spreading in $\mathbf{k}$-space in order to maximize carrier-phonon coupling in real space. This is automatically ensured by accounting for a constant field over several lattice constants around z_{trap}. Thus, all field-dependent quantities,

as the electrooptical function and frequency $K_{n,p}$ and $\Theta_{n,p}$ respectively, are locally referred to the trap site by definition.

Moreover, field-effect functions are not only functions of the electric field, but they also strongly depend on the characteristic energies of the process. Phonon energy $\hbar\omega_0$, that can be obtained both via experimental measurements and theoretical calculations, in fact, is an important parameter. Then we have the dependency on the trap energy E_{trap} which, as well as $\hbar\omega_0$, determines part of the process statistics via the optimum energy $E_{\text{opt}}(\mathcal{E})$. Finally we have the conduction and valence band energies E_{C} and E_{V} that are trivially local since they refer just to the trap site.

The most important physical material parameters which are in common among all the above mentioned quantities are essentially four: the trap energy E_{trap}, the Huang-Rhys factor S, the monomodal phonon frequency ω_0 and the electrooptical frequency Θ. E_{trap}, S and ω_0 can be obtained via Deep-Level Transient Spectroscopy (DLTS) experiments [171,177,183–185] and are local or average quantities by definition since the transient does not allow to discriminate spatial variations. Thus they are only mean values within the space charge region (SCR) and by reason of that the model is unavoidably local. Furthermore, Θ is field-dependent, so all previous discussions about locality of $\mathcal{E}$ hold again. It is also worth noting that the electrooptical frequency depends on the effective tunneling mass of carriers $m^*_{\text{tunn}_{n,p}}$, which not necessarily coincides with the standard effective mass $m^*_{n,p}$, as deeply argued in Refs. [132,183,186–189] and essentially due to some inadequacies of standard solid-state framework in describing trap states, as in case of wavefunction formalism and $E(\mathbf{k})$ dispersion relation picture.

To capture into MPE treatment some important non-local features related to the pure tunneling stage of TAT process – still remaining confined within the semi-classical approximation – now we follow in part the theory coming from the works by J. H. Zheng *et al.* [190] and S. Makram-Ebeid *et al.* [185]. Moreover, we will distinguish between the case of elastic and inelastic carrier-trap interaction through a formalism that allows us to explicit the overall TAT capture rate expression as the sum of two components coming from such scattering regimes. For this reason our perspective will hereinafter shift from lifetime/cross-section to capture rate modeling framework.

Again under both Einstein and Born-Oppenheimer approximations, the transition rate occurring between the initial *free* state (i.e. decoupled quasi-free carrier plus unoccupied trap) and the final *bound* state (carrier-trap coupling) can be written through the well-known Fermi golden rule as a sum computed over all the possible n_{f} free states:

$$W_{\text{f}\to\text{b}} = \frac{2\pi}{\hbar} \sum_{n_{\text{f}}} \left| \langle \Psi_{\text{f}} \left| \, H_{\text{f}\to\text{b}} \, \right| \Psi_{\text{b}} \rangle \right|^2 \delta \left(E_{\text{b}} - E_{\text{f}} \right) , \tag{4.119}$$

where indices "f" and "b" respectively refer to the free and bound states and where $\langle \Psi_{\text{f}} \left| \, H_{\text{f}\to\text{b}} \, \right| \Psi_{\text{b}} \rangle$ is the transition matrix element governed by the perturbation Hamiltonian $H_{\text{f}\to\text{b}}$ between the state $|\Psi_{\text{f}}\rangle$ with energy E_{f} and $|\Psi_{\text{b}}\rangle$ with energy E_{b}. Assuming a linear carrier-phonon interaction in the energy domain and by splitting

the overall state wavefunction Ψ in its electronic ψ and vibrational ϕ part we have

$$\begin{aligned}\langle \Psi_{\mathrm{f}} \,|\, H_{\mathrm{f}\to\mathrm{b}} \,|\, \Psi_{\mathrm{b}} \rangle &= \langle \psi_{\mathrm{f}} \,|\, G(z) \,|\, \psi_{\mathrm{b}} \rangle \left\langle \phi_{\mathrm{f}} \,\middle|\, z - z_{\mathrm{f}}^{0} \,\middle|\, \phi_{\mathrm{b}} \right\rangle \\ &= V \langle \phi_{\mathrm{f}} \,|\, \Delta z_{\mathrm{f}} \,|\, \phi_{\mathrm{b}} \rangle \ , \end{aligned} \tag{4.120}$$

where $G(z)$ is the optical deformation potential, V the electronic part of the transition matrix element and $\Delta z_{\mathrm{f}} = z - z_{\mathrm{f}}^{0}$, being z_{f}^{0} the equilibrium position of the oscillating free state, as already stated in this work, such that

$$z_{\mathrm{f}}^{0} = -\frac{\langle \psi_{\mathrm{f}} \,|\, G(z) \,|\, \psi_{\mathrm{f}} \rangle}{\hbar\omega_0} \, . \tag{4.121}$$

By means of modified Bessel functions Eq. (4.119) can be rewritten as

$$\begin{aligned} W_{\mathrm{f}\to\mathrm{b}} &\simeq \frac{\pi}{2\hbar} \left| V \cdot \Delta z_{\mathrm{f}\to\mathrm{b}}^{0} \right|^2 \left(1 - \frac{\epsilon}{S\hbar\omega_0} \right)^2 F\left(\epsilon\right) \\ &= \frac{\pi}{\hbar} |V|^2 S \left(1 - \frac{l}{S} \right)^2 F\left(\epsilon\right) \, , \end{aligned} \tag{4.122}$$

in which $|\Delta z_{\mathrm{f}\to\mathrm{b}}^{0}| = |z_{\mathrm{b}}^{0} - z_{\mathrm{f}}^{0}| = \sqrt{2S}$ is an already known MPE parameter (see, e.g., panel (b) of Figure 4.12) and where $F(E)$ is an opportune *auxiliary function* calculated for the energy

$$\epsilon = \frac{|E_{\mathrm{b}} - E_{\mathrm{f}}|}{2} = l\hbar\omega_0 \, . \tag{4.123}$$

Different researchers obtained approximated forms of the function $F(E)$, however we follow in particular the work published by K. Huang and A. Rhys [179] which exploits a Bessel function expansion:

$$F(E) = \frac{1}{\hbar\omega_0} \exp\left[-S\left(2f_{\mathrm{BE}} + 1\right) + \frac{E}{2k_{\mathrm{B}}T} \right] \sum_{l} \mathcal{B}_l(\gamma) \cdot \delta(l\hbar\omega_0 - E) \, , \tag{4.124}$$

that gives

$$\begin{aligned} F\left(\epsilon\right) \ \simeq \ & \frac{1}{\hbar\omega_0\sqrt{2\pi}} \exp\left[-S\left(2f_{\mathrm{BE}} + 1\right) + \frac{\Delta E_{n,p} + \frac{3}{2}k_{\mathrm{B}}T}{2k_{\mathrm{B}}T} + \chi \right] \\ & \times \frac{1}{\sqrt{\chi}} \left(\frac{\gamma}{l + \chi} \right)^{l} \, , \end{aligned} \tag{4.125}$$

where Eq. (4.125) stems from the polynomial approximation derived in Refs. [181, 190] and where γ is the same of Eq. (4.103) and $\chi = \sqrt{l^2 + \gamma^2}$.

Now we have to compute the electronic transition matrix element V that, as well as $\langle \phi_{\mathrm{f}} \,|\, \Delta z_{\mathrm{f}} \,|\, \phi_{\mathrm{b}} \rangle$, is intrinsically non-local by its nature, since it couples two states that are related also to quantities spatially (and temporally) separated by the tunneling path. Following A. Palma *et al.* [181]:

$$V \simeq 2\pi S \left(\hbar\omega_0\right)^2 \frac{\Omega_{\mathrm{trap}}}{\zeta_{\mathrm{trap}}} \int_{\zeta_{\mathrm{trap}}} |\psi_{\mathrm{f}}\rangle|^2 \, \mathrm{d}z \, , \tag{4.126}$$

assuming that the trap can be treated as a solid-cube entity [191] whose side is

$$\zeta_{\text{trap}} = \frac{\hbar}{2\sqrt{2\, m^*_{\text{tunn}_{n,p}}\, \Delta E_{n,p}}} \tag{4.127}$$

and $\Omega_{\text{trap}} = \zeta^3_{\text{trap}}$ is the trap volume. It also has to be mentioned that tunneling effective masses $m^*_{\text{tunn}_{n,p}}$ refer to the trap site.

Once we have the transition rate of Eq. (4.122), written by using the auxiliary function $F(\epsilon_{\text{r}})$ of Eq. (4.125) and the (electronic) transition matrix element V of Eq. (4.126), we are able to obtain the TAT+MPE capture rate from the transition rate $W_{\text{f}\to\text{b}}$ as a sum over conduction/valence free state energies weighted by statistical distributions of states such that:

$$\begin{aligned} c_{\text{MPE}_{n,p}} &= \Omega \int\limits_{\text{CB,VB}} W_{\text{f}\to\text{b}}\, f_{\text{C,V}}(E)\, \rho_{\text{C,V}}(E)\, \text{d}E \\ &= \frac{2\Omega_{\text{trap}} S \omega_0 \sqrt{m^*_{\text{tunn}_{n,p}} \left(m_0 k_{\text{B}} T\right)^3}}{\hbar^3 \sqrt{\chi}} \frac{(S-l)^2}{S} \left(\frac{\gamma}{l+\chi}\right)^l \\ &\quad \times \exp\left(-S(2f_{\text{BE}}+1) + \frac{\Delta E_{n,p} + \frac{3}{2} k_{\text{B}} T}{2 k_{\text{B}} T} + \chi\right) \\ &\quad \times \mathcal{F}_{1/2}\left(\frac{E_{\text{F}_{n,p}}(z_0) - E_{\text{C,V}}(z_0)}{k_{\text{B}} T}\right) \frac{|\psi(z_{\text{trap}})|^2}{|\psi(z_0)|^2}\,, \end{aligned} \tag{4.128}$$

where Ω is the device volume, $f_{\text{C,V}}(E)$ are the Fermi-Dirac distributions of free states in both bands and $\rho_{\text{C,V}}(E)$ their DOS.

It is worth highlighting that some quantities now result to be referred to the trap site, as the energy difference $\Delta E_{n,p}$, the tunneling effective masses $m^*_{\text{tunn}_{n,p}}$ or the carrier wavefunction ψ at the numerator of the last term (which is a measure of the pure tunneling probability, calculated under the WKB approximation). Some others, instead, have to be calculated at the turning point z_0, like quasi-Fermi levels $E_{\text{F}_{n,p}}$ (that, contrarily to previous models, here are no more forced to be constant along the SCR), band edges $E_{\text{C,V}}$ and the carrier wavefunction at the denominator of the WKB term. This complex framework defines the non-locality we were searching for since the beginning.

As anticipated, the present formulation can also account for the particular case of elastic carrier-trap scattering, i.e. when the optimum energy coincides with the trap energy, determining the absence of phonon emission: $E_{\text{ops}} \equiv E_{\text{trap}}$. In this case we have

$$\begin{aligned} c_{\text{el}_{n,p}} &= \frac{2\Omega_{\text{trap}} \Theta_{\text{C,V}} \sqrt{8 m^*_{\text{tunn}_{n,p}} m_0^3}}{\hbar^4 \pi} \\ &\quad \times \left(E_{\text{C,V}}(z_{\text{trap}}) - E_{\text{trap}}\right)^2 \left(E_{\text{trap}} - E_{\text{C,V}}(z_0)\right)^{\frac{3}{2}} \\ &\quad \times \mathcal{F}_{1/2}\left(\frac{E_{\text{F}_{n,p}}(z_0) - E_{\text{trap}}}{k_{\text{B}} T}\right) \frac{|\psi(z_{\text{trap}})|^2}{|\psi(z_0)|^2}\,, \end{aligned} \tag{4.129}$$

where E_{trap} is measured from E_{F_i} and the total capture rate is obtained through

$$c_{\text{TAT}_{n,p}} = c_{\text{MPE}_{n,p}} + c_{\text{el}_{n,p}} \,. \tag{4.130}$$

Notice that in Eq. (4.129) we find band edges $E_{\text{C,V}}$ defined both at the trap site and at the tunring point.

Now one can observe that Eqs. (4.128) and (4.129) have a common form. Thus they can be expressed by the generic formula

$$c_{n,p} = A\,\frac{2\Omega_{\text{trap}}\sqrt{m^*_{\text{tunn}_{n,p}} m_0^3}}{\hbar^4}\,\bar{F}(E)\,\mathcal{F}_{1/2}\!\left(\frac{E^*}{k_{\text{B}}T}\right)\frac{|\psi(z_{\text{trap}})|^2}{|\psi(z_0)|^2}\,, \tag{4.131}$$

where only some terms differentiates in case of phonon-assisted or elastic process. In particular, prefactors A are

$$\begin{cases} A_{\text{MPE}} = S\hbar\omega_0\sqrt{(k_{\text{B}}T)^3}\dfrac{(S-l)^2}{S} \\ A_{\text{el}} = \dfrac{\sqrt{8}\Theta_{\text{C,V}}}{\pi} \end{cases} . \tag{4.132}$$

Then $\bar{F}(E)$, partly coming from the auxiliary function $F(E)$, are described by

$$\begin{cases} \bar{F}_{\text{MPE}}(E) = \exp\left(-S(2f_{\text{BE}}+1) + \dfrac{\Delta E_{n,p} + \frac{3}{2}k_{\text{B}}T}{2k_{\text{B}}T} + \chi\right)\dfrac{1}{\sqrt{\chi}}\left(\dfrac{\gamma}{l+\chi}\right)^l \\ \bar{F}_{\text{el}}(E) = \Big(E_{\text{C,V}}(z_{\text{trap}}) - E_{\text{trap}}\Big)^2\Big(E_{\text{trap}} - E_{\text{C,V}}\,(z_0)\Big)^{\frac{3}{2}} \end{cases} . \tag{4.133}$$

Finally, the energy parameter E^*, deriving from integrand functions $f_{\text{C,V}}(E)$ and $\rho_{\text{C,V}}(E)$ in Eq. (4.128), reads

$$\begin{cases} E^*_{\text{MPE}} = E_{\text{F}_{n,p}}\,(z_0) - E_{\text{C,V}}\,(z_0) \\ E^*_{\text{el}} = E_{\text{F}_{n,p}}\,(z_0) - E_{\text{trap}} \end{cases} . \tag{4.134}$$

Summarizing our results about TAT, we started in Subsection 4.1.1.2 with the formalism developed by G. A. Hurkx, who introduced a new way of simulating TAT through specific field-effect enhancement functions $\Gamma_{n,p}(\mathcal{E})$ acting on carrier lifetimes $\tau_{\text{TAT}_{n,p}}(\mathcal{E}) = \tau_{n,p}\,(\Gamma_{n,p}(\mathcal{E}) + 1)$, where $\tau_{n,p}$ are the standard Shockley-Read-Hall lifetimes that have to be substituted by $\tau_{\text{TAT}_{n,p}}(\mathcal{E})$ within the SRH rate of Eq. (1.46) in order to obtain the TAT rate written in Eq. (4.81). In the Hurkx formalism all the relevant quantities playing within the model are "local", in the sense that band edges $E_{\text{C,V}}$, quasi-Fermi levels $E_{\text{F}_{n,p}}$ and electric field $\mathcal{E}$ are computed at the trap site z_{trap}: thus also lifetimes (and, in turn, capture cross sections) are local. Moreover, both $E_{\text{F}_{n,p}}$ and $\mathcal{E}$ are assumed to be constant along the SRC region.

The process considered by Hurkx accounts for a carrier – firstly tunneling with a certain energy E_{exc} greater than E_{trap} – that relaxes and then is capture by a trap, with a subsequent recombination mechanism. No specifications about the

nature of relaxation and trapping processes are made explicit. By introducing the multiphonon emission (MPE), A. Schenk preserved the procedure passing through field-functions but he also included lattice oscillations in his TAT formalism. This important further theoretical step allows to model the relaxation/trapping stage as a consequence of the Born-Oppenheimer approximation, where the unoccupied trap state interacts with the incoming carrier and with band edges through the emission of a certain number l of phonons. This mechanism is able to explain and justify the presence of an optimum TAT path, which follows suitable energy minimization rules. Also Schenk formalism assumes constant field and quasi-Fermi levels in the SCR and, again, it is a local tunneling model.

By accounting for a microscopic formulation of transition rates (based on the well-known Fermi golden rule) we finally included some opportune features to such MPE picture conferring non-local properties to the TAT model. We derived a formalism – already discussed in literature [181, 185, 190] and also implemented in commercial numerical solvers [192] – which rewrites capture rates as a sum of a phonon-assisted and elastic component, and where non-locality is given by accounting for characteristic energies of the process calculated both at the trap site and at the turning point. The model also introduces a WKB-based tunneling probability term which depends on carrier wavefunctions at the initial and final state. So the formalism is actually non-local, and no constant fields nor energies have to be invoked.

4.4 Full-Quantum Tunneling Simulation

All tunneling models we discussed in previous sections of this chapter were based on the semiclassical approximation of quantum mechanics and, for this reason, they are explicitly suitable for DD-based numerical simulations. Such kind of formalism provides the precious advantage of having a device-level description by means of relatively low-computational cost algorithms. However, this advantage is paid in terms of number and nature of physical effects that can be captured with a unique rate equation. If we are interested in specific quantum side effects, not included in a semiclassical approximation, or if we want to shift the focus towards more microscopic systems, then full-quantum formalisms are needed. Tunneling represents the most important mechanism that can be automatically described through such microscopic framework, being a pure quantum effect.

This kind of description results to be more computationally expensive than semiclassical one. For this reason many authors take advantage of a mesoscopic simulation framework, in which both semiclassical and quantum models coexist. Nonetheless, this procedure can lead to some information leakage when different scales of description have to be put in reciprocal cross-talk since modeling *always* implies a certain level of approximation. So, in the author's opinion, mesoscopic modeling should be carefully exploited only in particular cases, where the physics is such that the lost information do not compromise the final solution of the system: thus, only in specific *ad hoc* situations.

After studying and developing semiclassical theories for BTBT and TAT, which absorbed the main part of my first two Ph.D. years, the third year has been (also) devoted to the research in the field of full-quantum modeling.

Several techniques have been developed so far in the area of computational quantum physics. They can be divided into two main families, according to their most relevant quantity: so we find "density-matrix" models and "Green's function" models. In the next subsection these two modeling strategies will be discussed. Starting with a brief description of the matrix operator in quantum mechanics we will see the case of tunneling simulations in the Wigner-function approach. Then, a particular method for solving the Schrödinger equation will be introduced in order to explain in detail the theory about non-equilibrium Green's functions (NEGFs).

4.4.1 Density matrix and Wigner functions formalism

A quantum system consists of a collection of states, each of which can be described by a wavefunction $\Psi(\mathbf{r}, t)$ or, in Dirac formalism, by the ket $|\Psi\rangle$. The most generic representation of a state is not unique because typically it is not a *pure* state but a *mixed* state, i.e. a (linear) superposition of normalized pure states $|\psi_j\rangle$ as

$$|\Psi\rangle = \sum_{j=1}^{N} c_j \, |\psi_j\rangle \, , \tag{4.135}$$

weighted by opportune coefficients c_j, where the total number N can be also infinity. The density operator related to Eq. (4.135) is

$$\hat{\rho} = \sum_{j=1}^{N} a_j \, |\psi_j\rangle \, \langle\psi_j| \, , \tag{4.136}$$

where $\{a_j\}$ ($a_j \in \mathbb{R}$, $\forall j$) represent the probabilities to find the system in the state $|\psi_j\rangle$ such that $\sum_{j=1}^{N} a_j = 1$. From this property and from Eq. (4.136) stems the hermiticity of $\hat{\rho}$ ($\hat{\rho} \equiv \hat{\rho}^{\dagger}$) and also that the trace is unitary. In fact, choosing an arbitrary orthonormal basis $\{b_k\}$ such that

$$\sum_{k=1}^{N} |b_k\rangle \, \langle b_k| = 1 \, , \tag{4.137}$$

then

$$\begin{aligned} \mathrm{Tr}\,(\hat{\rho}) &= \sum_{j=1}^{N} a_j \sum_{k=1}^{N} \langle b_k \,|\, \Psi_j\rangle \, \langle \Psi_j \,|\, b_k\rangle \\ &= \sum_{j=1}^{N} a_j \sum_{k=1}^{N} \langle \Psi_j \,|\, b_k\rangle \, \langle b_k \,|\, \Psi_j\rangle \\ &= \sum_{j=1}^{N} a_j \, \langle \Psi_j \,|\, \Psi_j\rangle = 1 \end{aligned} \tag{4.138}$$

and, moreover,

$$\hat{\rho} \geq 0\,, \tag{4.139}$$

such that all of its eigenvalues are non-negative.

In the case of a pure state, instead, the system has a unique representation $|\Psi\rangle$. Consequently in this case the density operator simply reduces to

$$\hat{\rho} = |\Psi\rangle \langle\Psi| \tag{4.140}$$

because the probability of finding the state $|\Psi\rangle$ is actually always 1. This means that $\hat{\rho}^2 = \hat{\rho}$, in fact

$$\hat{\rho}^2 = |\Psi\rangle \langle\Psi \,|\, \Psi\rangle \langle\Psi| = |\Psi\rangle \langle\Psi| = \hat{\rho}\,, \tag{4.141}$$

and then

$$\mathrm{Tr}\left(\hat{\rho}^2\right) = \mathrm{Tr}\left(\hat{\rho}\right) = 1\,. \tag{4.142}$$

Now let us consider again a mixed state with the same base $\{b_k\}$ of before. Then the elements

$$\rho_{\alpha,\beta} = \sum_{j=1}^{N} a_j \,\langle b_\alpha \,|\, \psi_j\rangle \,\langle \psi_j \,|\, b_\beta\rangle \tag{4.143}$$

give the density matrix ρ. Density operator $\hat{\rho}$ (and its corresponding matrix ρ) is an important quantum mechanical operator since it allows to define the expectation value $\langle\hat{A}\rangle$ of any operator $\hat{A}$. In fact:

$$\begin{aligned}
\langle\hat{A}\rangle &= \sum_{j=1}^{N} a_j \langle\psi_j|\hat{A}|\psi_j\rangle \\
&= \sum_{j=1}^{N} a_j \sum_{\alpha=1}^{N}\sum_{\beta=1}^{N} \langle\psi_j \,|\, b_\beta\rangle \,\langle b_\beta|\hat{A}|b_\alpha\rangle \,\langle b_\alpha \,|\, \psi_j\rangle \\
&= \sum_{\alpha=1}^{N}\sum_{\beta=1}^{N}\sum_{j=1}^{N} a_j \,\langle b_\alpha \,|\, \psi_j\rangle \,\langle \psi_j \,|\, b_\beta\rangle \,\langle b_\beta|\hat{A}|b_\alpha\rangle \\
&= \sum_{\alpha=1}^{N}\sum_{\beta=1}^{N} \langle b_\alpha \,|\, \hat{\rho} \,|\, b_\beta\rangle \,\langle b_\beta|\hat{A}|b_\alpha\rangle \\
&= \sum_{\alpha=1}^{N}\sum_{\beta=1}^{N} \langle b_\beta|\hat{\rho}^\dagger|b_\alpha\rangle\langle b_\beta|\hat{A}|b_\alpha\rangle \\
&= \sum_{\alpha=1}^{N}\sum_{\beta=1}^{N} \rho^*_{\beta,\alpha}\, A_{\beta,\alpha}\, \delta_{\alpha,\beta} \\
&= \mathrm{Tr}\left(\hat{\rho}\hat{A}\right),
\end{aligned} \tag{4.144}$$

where δ is the Kronecker delta such that

$$\delta_{\alpha,\beta} = \begin{cases} 1 & \text{if} \quad \alpha = \beta \\ 0 & \text{if} \quad \alpha \neq \beta \end{cases}. \tag{4.145}$$

Suppose now to have a closed system described by Eq. (4.136) composed only by a single electron with two possible energy states E_α and E_β. Suppose also to consider a Hamiltonian which is the sum of a coherent single-particle $\mathbf{H}_0$ and a incoherent many-body Hamiltonian $\mathbf{H}'$. Then the time-evolution of $\hat{\rho}$ is given by the so-called semiconductor Bloch equation (SBE) given by [193, 194]

$$\begin{aligned}\frac{\mathrm{d}\rho_{\alpha\beta}}{\mathrm{d}t} &= \left.\frac{\mathrm{d}\rho_{\alpha,\beta}}{\mathrm{d}t}\right|_{\mathbf{H}_0} + \left.\frac{\mathrm{d}\rho_{\alpha,\beta}}{\mathrm{d}t}\right|_{\mathbf{H}'} \\ &= \frac{(E_\alpha - E_\beta)}{i\hbar}\,\delta_{\alpha,\beta,\alpha',\beta'} + \sum_{\alpha',\beta'} S_{\alpha,\beta,\alpha',\beta'} \\ &= \sum_{\alpha',\beta'} \mathcal{L}_{\alpha,\beta,\alpha',\beta'}\,\rho_{\alpha',\beta'}\,, \end{aligned} \tag{4.146}$$

where $\mathcal{L}$ is a Liouville super-operator. Eq. (4.146) actually represents a quantum equation of motion that, in the semiclassical limit – i.e. neglecting non-diagonal terms of $\hat{\rho}$ – reduces to the BTE. The single-particle SBE for $\rho_{\alpha,\beta}$ gives, for each density matrix element, the solution

$$\rho_{\alpha,\beta}(t) = \rho_{\alpha,\beta}(t_0)\,\exp\left[(t-t_0)\,\frac{E_\alpha - E_\beta}{i\hbar}\right], \tag{4.147}$$

while the many-body component follows a more complicated time-evolution which involves phonon-assisted and higher-order density matrices and must be treated separately.

If we apply a Weyl-Wigner transform

$$u_{\alpha,\beta}(\mathbf{r},\mathbf{k}) = \int \psi_\alpha\left(\mathbf{r}+\mathbf{r}'/2\right)\,\frac{\mathrm{e}^{-i\mathbf{k}\cdot\mathbf{r}'}}{(2\pi)^{3/2}}\,\psi^*_\beta\left(\mathbf{r}-\mathbf{r}'/2\right)\,\mathrm{d}\mathbf{r}' \tag{4.148}$$

to Eq. (4.136) we obtain the so-called Wigner function [195]

$$f^{\mathrm{W}}(\mathbf{r},\mathbf{k}) = \sum_{\alpha,\beta} u_{\alpha,\beta}(\mathbf{r},\mathbf{k})\,\rho_{\alpha,\beta}\,, \tag{4.149}$$

that can be generalized for a generic set basis $\{|\gamma\rangle\}$

$$\begin{aligned} f^{\mathrm{W}}_{\gamma_1,\gamma_2}(\mathbf{r}) &= \sum_{\alpha,\beta} u^{\alpha,\beta}_{\gamma_1,\gamma_2}(\mathbf{r})\,\rho_{\alpha,\beta} \\ &= \int K_{\gamma_1,\gamma_2}(\mathbf{r};\mathbf{r}',\mathbf{k}')\,f^{\mathrm{W}}(\mathbf{r}',\mathbf{k}')\,\mathrm{d}\mathbf{r}'\,\mathrm{d}\mathbf{k}'\,, \end{aligned} \tag{4.150}$$

where

$$K_{\gamma_1,\gamma_2}(\mathbf{r};\mathbf{r}',\mathbf{k}') = \sum_{\alpha\beta} u_{\alpha\beta}(\mathbf{r},\mathbf{k})\,u^*_{\alpha,\beta}(\mathbf{r}',\mathbf{k}')\,. \tag{4.151}$$

In the same manner we can obtain the expectation value of an operator $\hat{A}$ as

$$\langle\hat{A}\rangle = \frac{1}{\hbar^3}\int \mathrm{d}\mathbf{r} \int A^{\mathrm{W}}(\mathbf{r},\mathbf{k})\,f^{\mathrm{W}}(\mathbf{r},\mathbf{k})\,\mathrm{d}\mathbf{k}\,, \tag{4.152}$$

where

$$A^{\mathrm{W}}(\mathbf{r},\mathbf{k}) = \sum_{\alpha,\beta} u_{\alpha,\beta}(\mathbf{r},\mathbf{k})\, A_{\alpha,\beta}\,. \tag{4.153}$$

In order to extend the validity of Eq. (4.149) to the case of open systems, we have to restrict its domain to a volume Ω by introducing proper boundary conditions, where f^{W} assumes the value $f_{\mathrm{b}}^{\mathrm{W}}$. Thus, we get

$$\begin{aligned} f^{\mathrm{W}}(\mathbf{r},\mathbf{k};t) = &\int_{\Omega} \mathrm{d}\mathbf{r}' \int \hat{T}(\mathbf{r},\mathbf{k};\mathbf{r}',\mathbf{k}';t-t_0)\, f^{\mathrm{W}}(\mathbf{r}',\mathbf{k}';t_0)\,\mathrm{d}\mathbf{k}' \\ &+ \int \mathrm{d}\mathbf{r}_{\mathrm{b}} \int \mathrm{d}\mathbf{k}' \int_{t_0}^{t} \hat{T}(\mathbf{r},\mathbf{k};\mathbf{r}_{\mathrm{b}},\mathbf{k}';t-t')\, f_{\mathrm{b}}^{\mathrm{W}}(\mathbf{r}_{\mathrm{b}},\mathbf{k}';t')\, v(\mathbf{k}')\,\mathrm{d}t'\,, \end{aligned} \tag{4.154}$$

where $\hat{T}$ is an evolution operator and $v(\mathbf{k}')$ is the group velocity normal to the boundary surface at $\mathbf{r}_{\mathrm{b}}$. As one may see in Eq. (4.154), the general function $f^{\mathrm{W}}(\mathbf{r},\mathbf{k};t)$ is the evolution of two components: the first one evaluated at initial time t_0 within the volume Ω and the second one evaluated at boundaries for times $t' > t_0$.

Contrarily to what may appear from the parallelism with semiclassical BTE, the function we derived is not a true distribution function. It is a real function and it can assume also negative values. For these reasons we call f^{W} a *pseudo-distribution function*.

The strategies to model quantum transport in electronic devices which include also tunneling mechanisms are different. For instance N. C. Kluksdahl *et al.* [196,197] tested as f^{W} the Gaussian wave packet (GWP)

$$f_{\mathrm{GWP}}(\mathbf{r},\mathbf{k}) = \frac{1}{a\hbar\,(2\pi)^{3/2}}\, \mathrm{e}^{-\frac{\mathbf{r}^2}{4a^2}}\, \mathrm{e}^{-\frac{4(\mathbf{k}-\mathbf{k}_0)^2}{a^2}} \tag{4.155}$$

to describe a particle impinging on a potential barrier. Through a numerical treatment, i.e. solving Eq. (4.155) by means of a discretization procedure both for space and momentum, they were able to find the evidence of several quantum effects in a AlGaAs/GaAs resonant-tunneling diode (RTD), like the bistability in the negative-differential conductivity (NDC) regime, the zero-bias anomaly and the nonlocality of the potential term in the equation of motion respecting both position and momentum.

Another possible methodology is the implementation of a Monte Carlo (MC) approach based on the WF pseudo-distribution. Before entering in details, we want to develop the dynamical equations of the WF, following the approach presented by C. Jacoboni *et al.* [198], and to this purpose we rewrite Eq. (4.149) as

$$\begin{aligned} f^{\mathrm{W}}(\mathbf{r},\mathbf{k}) &= \int \mathrm{e}^{-i\mathbf{k}\cdot\mathbf{r}'/\hbar}\, \rho\left(\mathbf{r}+\mathbf{r}'/2, \mathbf{r}-\mathbf{r}'/2\right)\,\mathrm{d}\mathbf{r}' \\ &= \int \mathrm{e}^{-i\mathbf{k}\cdot\mathbf{r}'/\hbar}\, \langle \mathbf{r}+\mathbf{r}'/2\,|\,\Psi\rangle\,\langle\Psi\,|\,\mathbf{r}-\mathbf{r}'/2\rangle\,\mathrm{d}\mathbf{r}' \\ &= \int \mathrm{e}^{-i\mathbf{k}\cdot\mathbf{r}'/\hbar}\, \Psi\left(\mathbf{r}+\mathbf{r}'/2\right)\Psi^*\left(\mathbf{r}-\mathbf{r}'/2\right)\,\mathrm{d}\mathbf{r}'\,, \end{aligned} \tag{4.156}$$

where $\mathbf{r}' = 2X$, with X the "center of mass" vector of the system, are the Wigner coordinates and where $\Psi(\mathbf{r})$ are ensemble-state wavefunctions. Then, for a generic noninteracting Hamiltonian

$$H = \frac{\hat{p}^2}{2m} + V(\mathbf{r}) \tag{4.157}$$

we have the dynamical equation

$$\frac{\partial}{\partial t} f^{\mathrm{W}}(\mathbf{r}, \mathbf{k}; t) + \frac{\mathbf{k}}{m} \nabla f^{\mathrm{W}}(\mathbf{r}, \mathbf{k}; t) = \frac{1}{\hbar^3} \int \mathcal{V}^{\mathrm{W}}(\mathbf{r}, \mathbf{k} - \mathbf{k}')\, f^{\mathrm{W}}(\mathbf{r}, \mathbf{k}'; t)\, \mathrm{d}\mathbf{k}' \,, \tag{4.158}$$

where

$$\mathcal{V}^{\mathrm{W}}(\mathbf{r}, \mathbf{k} - \mathbf{k}') = \frac{1}{i\hbar} \int \mathrm{e}^{-i(\mathbf{k}-\mathbf{k}')\mathbf{r}'/\hbar} \left[V\left(\mathbf{r} + \mathbf{r}'/2\right) V\left(\mathbf{r} - \mathbf{r}'/2\right) \right] \mathrm{d}\mathbf{r}' \tag{4.159}$$

is a real function representing the potential term, which behaves like both "scattering in" and "scattering out" contributions in the semiclassical BTE. Moreover, $\mathcal{V}^{\mathrm{W}}$ is a non-local function of space, since potentials V are evaluated at a point different from $\mathbf{r}$. Now we extend the analysis by taking into account electron-phonon interactions through the reduced electron density matrix

$$\begin{aligned} \tilde{\rho}(\mathbf{r}, \mathbf{r}') &= \mathrm{Tr}_{\mathrm{ph}} \left[\rho(\mathbf{r}, \{n_{\mathbf{q}}\}; \mathbf{r}', \{n'_{\mathbf{q}}\}) \right] \\ &= \sum_{\{n_{\mathbf{q}}\}} \rho(\mathbf{r}, \{n_{\mathbf{q}}\}; \mathbf{r}', \{n'_{\mathbf{q}}\}) \end{aligned} \tag{4.160}$$

evaluated over the basis $|\mathbf{r}, \{n_{\mathbf{q}}\}\rangle$ of the electron and lattice vibrational state, where

$$\rho(\mathbf{r}, \{n_{\mathbf{q}}\}; \mathbf{r}', \{n'_{\mathbf{q}}\}) = \overline{\langle \mathbf{r}, \{n_{\mathbf{q}}\} | \Psi(t) \rangle \langle \Psi(t) | \mathbf{r}', \{n'_{\mathbf{q}}\} \rangle} \tag{4.161}$$

and $\{n_{\mathbf{q}}\}$ represents the occupation number of the phonon mode $\mathbf{q}$ having frequency $\omega_{\mathbf{q}}$. The generalized $f^{\mathrm{W}}(\mathbf{r}, \{n_{\mathbf{q}}\}; \mathbf{r}', \{n'_{\mathbf{q}}\})$ can be derived with respect to time and than applied to the Schrödinger equation, in which the electron-phonon Hamiltonian part is described by

$$H_{\text{e-p}} = \sum_{\mathbf{q}} i\hbar F(\mathbf{q}) \left[a_{\mathbf{q}}\, \mathrm{e}^{i\mathbf{q}\cdot\mathbf{r}} - a^{\dagger}_{\mathbf{q}}\, \mathrm{e}^{-i\mathbf{q}\cdot\mathbf{r}} \right] , \tag{4.162}$$

where $F(\mathbf{q})$ is a real function describing the electron-phonon interaction. This procedure yields a coupling term of the following form:

$$\begin{aligned} \left.\frac{\partial}{\partial t} f^{\mathrm{W}}\right|_{\text{e-p}} = & \sum_{\mathbf{q}'} i\hbar F(\mathbf{q}') \\ & \times \Big\{ \mathrm{e}^{i\left(\mathbf{q}'\mathbf{r} - \omega_{\mathbf{q}'}(t-t_0)\right)}\, f^{\mathrm{W}}_{-} \sqrt{n_{\mathbf{q}'} + 1} \\ & - \mathrm{e}^{-i\left(\mathbf{q}'\mathbf{r} - \omega_{\mathbf{q}'}(t-t_0)\right)}\, f^{\mathrm{W}}_{+} \sqrt{n_{\mathbf{q}'}} \\ & + \mathrm{e}^{-i\left(\mathbf{q}'\mathbf{r} - \omega_{\mathbf{q}'}(t-t_0)\right)}\, f'^{\mathrm{W}}_{-} \sqrt{n'_{\mathbf{q}'} + 1} \\ & - \mathrm{e}^{i\left(\mathbf{q}'\mathbf{r} - \omega_{\mathbf{q}'}(t-t_0)\right)}\, f'^{\mathrm{W}}_{+} \sqrt{n'_{\mathbf{q}'}} \Big\} \,, \end{aligned} \tag{4.163}$$

where

$$\begin{cases} f_{-}^{\mathrm{W}} = f^{\mathrm{W}}\left(\mathbf{r}, \mathbf{k} - \dfrac{\hbar\mathbf{q}'}{2}, \{n_1, \ldots, n_{\mathbf{q}'} + 1, \ldots\}, \{n'_{\mathbf{q}}\}; t\right) \\ f_{+}^{\mathrm{W}} = f^{\mathrm{W}}\left(\mathbf{r}, \mathbf{k} + \dfrac{\hbar\mathbf{q}'}{2}, \{n_1, \ldots, n_{\mathbf{q}'} - 1, \ldots\}, \{n'_{\mathbf{q}}\}; t\right) \\ f'^{\mathrm{W}}_{-} = f^{\mathrm{W}}\left(\mathbf{r}, \mathbf{k} - \dfrac{\hbar\mathbf{q}'}{2}, \{n_{\mathbf{q}}\}, \{n'_1, \ldots, n'_{\mathbf{q}'} + 1, \ldots\}; t\right) \\ f'^{\mathrm{W}}_{+} = f^{\mathrm{W}}\left(\mathbf{r}, \mathbf{k} + \dfrac{\hbar\mathbf{q}'}{2}, \{n_{\mathbf{q}}\}, \{n'_1, \ldots, n'_{\mathbf{q}'} + 1, \ldots\}; t\right) \end{cases} . \tag{4.164}$$

So the total dynamical equation for the Wigner function $f^{\mathrm{W}}(\mathbf{r}, \{n_{\mathbf{q}}\}; \mathbf{r}', \{n'_{\mathbf{q}}\})$ reads:

$$\frac{\partial}{\partial t} f^{\mathrm{W}} + \frac{\mathbf{k}}{m} \nabla f^{\mathrm{W}} + \mathbf{F} \nabla_{\mathbf{k}} f^{\mathrm{W}} = \frac{1}{\hbar^3} \int \mathcal{V}^{\mathrm{W}}(\mathbf{r}, \mathbf{k} - \mathbf{k}')\, f^{\mathrm{W}} + \frac{\partial}{\partial t} f^{\mathrm{W}} \bigg|_{\text{e-p}} \mathrm{d}\mathbf{k}' , \tag{4.165}$$

in which $\mathbf{F}$ is a constant or harmonic force. The r.h.s. terms in Eq. (4.163) represent the possible interactions (*vertex*) in which phonon absorption/emission induces a change of the occupation number $\{n_{\mathbf{q}}\}$ by unity and of the electron momentum by $\Delta\mathbf{k} = \hbar\mathbf{q}/2$. If a second vertex occurs after the first interaction, such that the momentum is completely transferred, the transition is *real*. Otherwise, if the half momentum is returned, the transition is *virtual*.

The solution of Eq. (4.163) is a nontrivial problem due to non-diagonal terms involved in the interaction and since the number of variables is very high. The MC method helps us in the sense that a numerical solution which passes through the sampling of all the possible phonon modes allows to reduce the complexity of this problem. The strategy that can be adopted consists in introducing new temporary variables $\bar{\mathbf{r}}(t') = \mathbf{r}(\mathbf{r}^*, \mathbf{k}^*; t')$ and $\bar{\mathbf{k}}(t') = \mathbf{k}(\mathbf{r}^*, \mathbf{k}^*; t')$, where

$$\begin{cases} \mathbf{r}^*(\mathbf{r}, \mathbf{k}; t) = \mathbf{r} - \int_{t_0}^{t} \dot{\mathbf{r}}(\tau)\, \mathrm{d}\tau \\ \mathbf{k}^*(\mathbf{r}, \mathbf{k}; t) = \mathbf{k} - \int_{t_0}^{t} \dot{\mathbf{k}}(\tau)\, \mathrm{d}\tau \\ t^*(\mathbf{r}, \mathbf{k}; t) = t \end{cases} , \tag{4.166}$$

in order to finally obtain the following integral form of Eq. (4.165)

$$\begin{aligned} f^{\mathrm{W}}\left(\mathbf{r}, \mathbf{k}, \{n_{\mathbf{q}}\}, \{n'_{\mathbf{q}}\}; t\right) = & f^{\mathrm{W}}\left(\bar{\mathbf{r}}(t_0), \bar{\mathbf{k}}(t_0), \{n_{\mathbf{q}}\}, \{n'_{\mathbf{q}}\}; t\right) \\ & + \int_{t_0}^{t} \mathrm{d}t' \frac{1}{3} \int \mathcal{V}^{\mathrm{W}}\left(\bar{\mathbf{r}}(t'), \bar{\mathbf{k}}(t') - \mathbf{k}'\right) \\ & \qquad \times f^{\mathrm{W}}\left(\bar{\mathbf{r}}(t'), \mathbf{k}', \{n_{\mathbf{q}}\}, \{n'_{\mathbf{q}}\}; t'\right) \mathrm{d}\mathbf{k}' \\ & + \int_{t_0}^{t} \frac{\partial f^{\mathrm{W}}(t')}{\partial t} \bigg|_{\text{e-p}} \mathrm{d}t' . \end{aligned} \tag{4.167}$$

The first term of Eq. (4.167) is a ballistic contribution at initial time t_0 whereas the following addenda, instead, represent phonon and potential interactions under the integral operator that we call $\mathbf{S}$. Thus, inserting Eq. (4.167) into itself, an iterative form

$$f^{\mathrm{W}} = f_0^{\mathrm{W}} + \mathbf{S}f_0^{\mathrm{W}} + \mathbf{SS}f_0^{\mathrm{W}} + \cdots \tag{4.168}$$

can be obtained. This kind of convergent Neumann expansion series represents the MC solution of the problem. The meaning of the zeroth-order term f_0^{W} is trivial; the first-order one, instead, gives all the WF contributions coming from particles with any position and momentum, scattering at time t' and, after a proper free flight, arriving to the state $(\mathbf{r}, \mathbf{k})$. During the free flights, trajectories are classical, whereas at the vertex momentum changes abruptly as discussed above.

Assigning an arbitrary probability P_n (with $n = 0, 1, \ldots$) to each phase space point, the WF at time t evaluated in one of these points, times a proper weight term $W = 1/P_n$, represents an estimated value of the WF itself. By sampling all the terms in Eq. (4.168) through a MC procedure one is able, in principle, to describe the system in terms of scattering events and free flight paths: since n denotes the number of interaction vertices, when $n = 0$ only the ballistic term is considered, such that no random selection is required and the WF at $(\mathbf{r}, \mathbf{k}; t)$ is evaluated starting from the initial state $(\bar{\mathbf{r}}(t_i), \bar{\mathbf{k}}(t_i); t_i)$. When $n > 0$, all the n potential vertices have to be accounted for. In these cases the initial time and momentum transfer probabilities $P_{\Delta\mathbf{k}_i}$ and P_{t_i} can be selected arbitrarily. Thus, the particle momentum just before the vertex is updated to $\mathbf{k}_i = \bar{\mathbf{k}}(t_i) - \Delta\mathbf{k}_i$, then a new free flight is evaluated. The phonon interactions, on the other hand, are included by considering random selection of the vertices and of the infinite possible Wigner paths contributing to the WF, each of which is weighted by a proper factor. Moreover, the phonon occupation numbers are allowed to assume all possible values in order to account for a phonon bath always in equilibrium conditions.

The framework just outlined can be easily extended to the case of many-particle systems. The only warning to keep in mind is the possibility of particle-particle interactions. In this case the resulting vertex is considered non-null only if the momentum transfer satisfies the relation $\Delta\mathbf{k}_1 = -\Delta\mathbf{k}_2$, where $\mathbf{k}_1$ and $\mathbf{k}_2$ are the momenta of two generic interacting particles.

An interesting example of such MC scheme here discussed for WF formalism is constituted by the resonant tunneling diode (RTD), a device commonly based on AlGaAs/GaAs heterojunctions. The typical barrier height – of the order of some tenths of meV – is such that resonant states, due to the tunneling transparency, are present inside the well. As the increasing applied bias reaches the resonant energy, the diode $I(V)$ characteristic experiences a current peak then, for higher voltages, the current becomes smaller giving rise to a NDC regime.

In Ref. [199] P. Bordone *et al.* tested such MC-WF formalism on a AlGaAs/GaAs double barrier RTD, in which the two AlGaAs barriers are 0.28 eV high and 2.8 nm wide and are separated by a GaAs layer of 5 nm. By computing the electron-phonon term of the dynamical equation via MC sampling of the phonon modes, the authors have been able to study the resonant tunneling phenomenon, correctly reproducing

the NDC regime. Quantum nature of NDC behavior was put in clear evidence also in Ref. [198] for a very similar RTD structure.

4.4.2 Non-Equilibrium Green's Functions (NEGFs)

Besides Wigner function (WF) formalism another rigorous method providing full-quantum simulations in micro and mesoscopic systems – whose basic theory was introduced by L. V. Keldysh, L. P. Kadanoff, and G. Baym in 60s [200, 201], but recently improved mainly thanks to S. Datta [202] – is becoming increasingly important. No more based on density matrix and on functions defined in the phase space $(\mathbf{r}, \mathbf{k})$, non-equilibrium Green's functions (NEGFs) method derives from Schrödinger equation, although it can not be considered a wavefunction-based formalism. As we shall see, WF and NEGF are linked by some opportune mathematical transforms.

Let us start from the monodimensional (1D) time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial z^2}\Psi_n(z) + U(z)\Psi_n(z) = E_n\Psi_n(z) \tag{4.169}$$

where, as usual in the present dissertation, z represents the spatial dimension along which we are projecting our physical system, U is the potential and m^* the effective mass referred to the particles (fermions) which populate the system itself. We know that Eq. (4.169) is an eigenvalue equation whose solutions, labeled with the index n, include a certain set of coupled energy eigenvalues E_n and wavefunction eigenvectors Ψ_n. These solutions define a discrete set of allowed states that change according to the shape of the potential U.

The analytical solution of Eq. (4.169) is trivial but quite cumbersome in the case of symmetrical problems, i.e. where the potential shape is simple (e.g., think about the case of a finite square well described and solved in Subsection 3.3.1). If we want to solve the eigenvalue problem for *any* kind of potential then we have to resort to appropriate numerical strategies. Here we present a simplification that somehow resembles the discretization procedures seen talking about classical transport models in semiconductors (see Subsection 1.5.3): to this purpose now we introduce a finite difference (FD) approximation scheme of Eq. (4.169) according to which the second partial derivative over z is written as

$$\frac{\partial^2}{\partial z^2}\Psi_n(z) \overset{\text{FD}}{\approx} \frac{\Psi_n(z_i + \Delta z) - 2\,\Psi_n(z_i) + \Psi_n(z_i - \Delta z)}{(\Delta z)^2}\,. \tag{4.170}$$

Eq. (4.170) represents the second-orded *central difference* scheme applied to the Schrödinger equation where, if we divide our z domain into a number $N+1$ of points, then Δz becomes the spacing between two consecutive mesh nodes z_i and z_{i+1}, where $i = 1, \ldots, N$. In order to close the system we have to impose opportune boundary conditions (BC) in the initial z_1 and final z_{N+1} points. A common choice is represented by homogeneous Dirichlet BCs (see also Subsection 1.5.2) for the wavefunctions, that give

$$\begin{cases} \Psi_n(z_1) = 0 \\ \Psi_n(z_{N+1}) = 0 \end{cases}. \tag{4.171}$$

Now let us rewrite Eq. (4.169) in a more suitable form:

$$\mathbf{H}\Psi_n(z_i) = E_n\Psi_n(z_i)\,, \tag{4.172}$$

where $\mathbf{H}$ represents the Hamiltonian operator

$$\mathbf{H} = -\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial z^2} + U(z)\,. \tag{4.173}$$

In order to make Eq. (4.172) numerically implementable we have to discretize also $\mathbf{H}$. This can be achieved by rewriting it in a matrix form:

$$[\mathbf{H}] = [\mathbf{K}] + [\mathbf{U}]\,, \tag{4.174}$$

where $[\mathbf{K}]$ is the kinetic energy matrix and $[\mathbf{U}]$ the potential energy matrix. If the potential shape U is *a priori* known, the latter can be easily constructed as a diagonal matrix, whose elements are the potential values calculated at the mesh nodes $U(z_i)$. Thus:

$$[\mathbf{U}] = \begin{pmatrix} U(z_1) & 0 & 0 & 0 & \cdots & 0 \\ 0 & U(z_2) & 0 & 0 & \cdots & 0 \\ 0 & 0 & U(z_3) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & U(z_{N+1}) \end{pmatrix} \tag{4.175}$$

or, in a more compact form:

$$[\mathbf{U}]_{i,j} = U(z_i)\,\delta_{i,j} \tag{4.176}$$

where, again, we use the Kronecker delta.

The kinetic energy matrix, instead, deserves more attention. Exploiting the central difference FD scheme discussed above, it is possible to write $[\mathbf{K}]$ as a tri-diagonal matrix, where the main diagonal includes the terms $-2\,\Psi_n\left(z_i\right)/(\Delta z)^2$:

$$[\mathbf{K}]_{i,j} = -\frac{\hbar^2}{2m^*}\frac{-2}{(\Delta z)^2}\,\delta_{i,j}\,, \quad \text{for} \quad i,j = 1,\cdots,N+1 \tag{4.177}$$

and the two secondary upper/lower diagonals are $\Psi_n\left(z_i \pm \Delta z\right)/\left(\Delta z\right)^2$, so:

$$[\mathbf{K}]_{i,j+1} = [\mathbf{K}]_{i+1,j} = -\frac{\hbar^2}{2m^*}\frac{1}{(\Delta z)^2}\,\delta_{i,j}\,, \quad \text{for} \quad i,j = 1,\cdots,N\,. \tag{4.178}$$

Thus:

$$[\mathbf{K}] = -\frac{\hbar^2}{2m^*}\frac{1}{(\Delta z)^2}\begin{pmatrix} 2 & -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}. \tag{4.179}$$

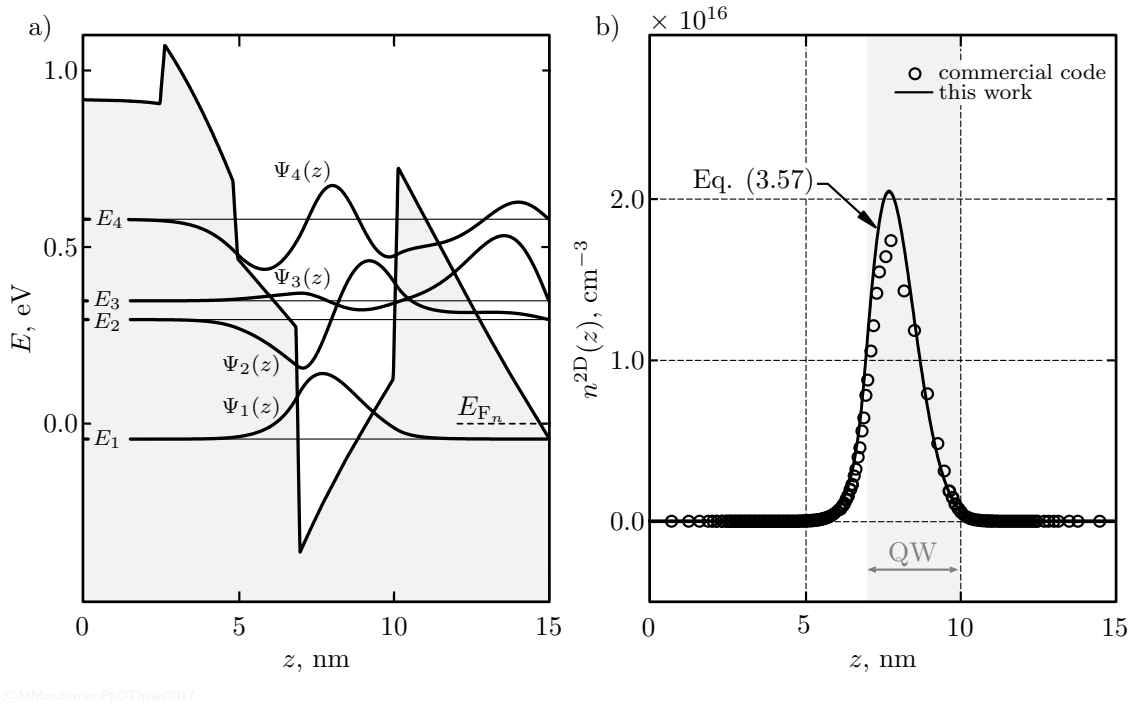

FIGURE 4.15. Monodimensional (1D) AlGaN/InGaN/GaN heteroscructure with quantum well (QW) simulated through the finite-central-difference Schrödinger solver here described: (a) conduction band diagram reporting electron eigenenergies E_n and eigenfunctions $\Psi_n(z)$. Note the wavefunction broadening into the barriers and some tunneling leakage occurring on the right side of the device. (b) Bidimensional (2D) electron density calculated via author's routine and also through a commercial TCAD tool.

Once we discretized both energies and space, we can solve the FD Schrödinger equation written in Eq. (4.172) through direct or iterative methods. At the end one should obtain a set of coupled E_n and $\Psi_n(z_i)$ solutions – both for electrons and holes – starting from any potential U. By knowing the system eigenfunctions it is also possible to compute some interesting physical quantities, as the microscopic expectation values of the energy and momentum operators. Moreover, in confining systems as quantum wells, one can even compute the 2D charge density according to formulas written in Eqs. (3.57) and (3.58).

In Figure 4.15 we can find some results from the 1D Schrödinger solver implemented by the author and based on the finite-central-difference scheme we discussed. Panel (a) represents the electron energy profile (conduction band) for a test AlGaN/InGaN/GaN heterostructure under forward bias conditions (2.5 V) where $x_{\mathrm{Al}} = 15\%$ and $x_{\mathrm{In}} = 18\%$. The device simulated here is very compact and includes: 2.5 nm-thick p-GaN and 2.5 nm-thick p-AlGaN EBL, both doped by $N_{\mathrm{A}} = 2 \cdot 10^{19}$ cm^{-3}, followed by non-intentionally doped 2 nm-thick GaN, 3 nm-thick InGaN quantum well and 2.5 nm-thick GaN barrier. The structure concludes with a 2.5 nm-thick n-GaN layer doped by $N_{\mathrm{D}} = 5 \cdot 10^{18}$ cm^{-3}. Besides the potential, we represented also the electron quasi-Fermi level and the outcoming energy levels and (normalized) wavefunctions. A similar approach can be applied for holes in valence

band (here not represented).

The potential profile is extracted from the DD solution of a commercial TCAD simulator [192] and then used as input file for the 1D Schrödinger solver. The same commercial suite has also been exploited in panel (b), were the electron density $n^{2D}(z)$ calculated both with the TCAD tool and the method described in this work through Eq. (3.57) have been compared.

From panel (a) of Figure 4.15 one may clearly appreciate the occurrence of wavefunctions penetration into both barriers surrounding the QW. Moreover, some tunneling leakage coming from high-order solutions can be seen overcoming the barrier on the n-type side. Since these quasi-bound states (QBS) may induce intra-band tunneling through such triangular barrier also in real devices, we further investigate the phenomenon developing a strategy to compute the related intra-band tunneling current density. Let us start considering again the continuity equation for electrons written without accounting for any generation/recombination process. So we have

$$\nabla_{\mathbf{r}} J_n(z) = q \frac{\partial n(z,t)}{\partial t} \,. \tag{4.180}$$

At the same time the density associated to the wavefunction, a concept somehow linked to the electron density but referred to the Ψ, can be set up as

$$n(z,t) \simeq |\Psi(z,t)|^2 \,. \tag{4.181}$$

By plugging Eq. (4.181) into Eq. (4.180) we obtain

$$\nabla_{\mathbf{r}} J_n(z) \simeq q \frac{\partial}{\partial t} |\Psi(z)|^2 \,. \tag{4.182}$$

This means that, if we are interested in the time-variation of charge density, we have to deal with time-variations of wavefunction probability density. Developing the calculations:

$$\begin{aligned} \frac{\partial}{\partial t} |\Psi(z,t)|^2 &= \frac{\partial}{\partial t} \Psi^*(z,t)\Psi(z,t) \\ &= \Psi(z,t)\frac{\partial}{\partial t}\Psi^*(z,t) + \Psi^*(z,t)\frac{\partial}{\partial t}\Psi(z,t) \\ &= -\frac{\Psi(z,t)}{i\hbar}\left(\mathbf{H}^*\Psi^*(z,t)\right) + \frac{\Psi^*(z,t)}{i\hbar}\left(\mathbf{H}\Psi(z,t)\right) \,, \end{aligned} \tag{4.183}$$

where $\mathbf{H}$ is the Hamiltonian and were we used the relations

$$\begin{cases} i\hbar\dfrac{\partial}{\partial t}\Psi(z,t) = \mathbf{H}\Psi(z,t) \\ -i\hbar\dfrac{\partial}{\partial t}\Psi^*(z,t) = \mathbf{H}^*\Psi^*(z,t) \end{cases} . \tag{4.184}$$

Assume again to have

$$\mathbf{H} = \frac{\hat{\mathbf{p}}^2}{2m^*} = -\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial z^2} + U(z) \tag{4.185}$$

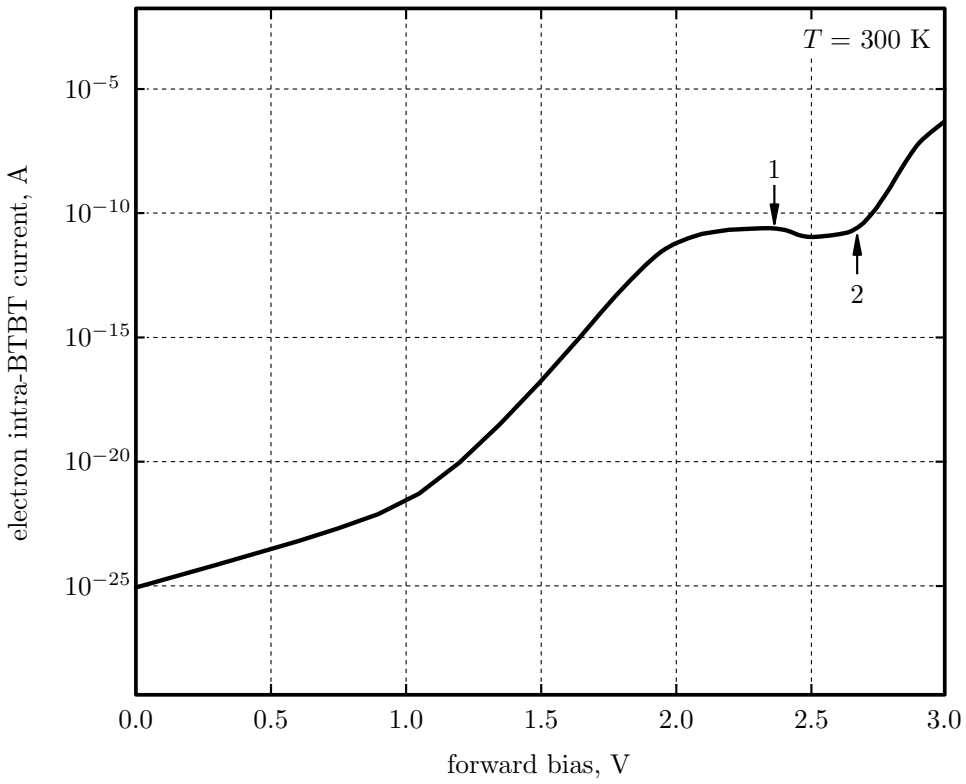

FIGURE 4.16. Electron intra-band-to-band tunneling (intra-BTBT) current simulated through the 1D Schrödinger solver by integrating the current density computed with the momentum operator formalism. Note that the zero-bias current is non-null. Moreover at point "1" there is a local maximum, due to the fact that the fundamental level E_1 is fully occupied (it crosses the quasi-Fermi energy), then at point "2" the current rises again since the second level E_2 starts to be gradually occupied.

where

$$\hat{\mathbf{p}} = -i\hbar\frac{\partial^2}{\partial z^2} \tag{4.186}$$

is the momentum operator. Putting together Eqs. (4.182) and (4.183) we obtain

$$\begin{aligned} J_n(z) &= -q\frac{i\hbar}{2m^*}\Big(\Psi^*(z,t)\frac{\partial}{\partial t}\Psi(z,t) - \Psi(z,t)\frac{\partial}{\partial t}\Psi^*(z,t)\Big) \\ &= \frac{q}{2m^*}\Big(\Psi^*(z,t)\left(\hat{\mathbf{p}}\Psi(z,t)\right) + \left(\hat{\mathbf{p}}\Psi(z,t)\right)^*\Psi(z,t)\Big) \\ &= \frac{q}{m^*}\,\mathrm{Re}\Big\{\Psi^*(z)\left(\hat{\mathbf{p}}\Psi(z)\right)\Big\} \end{aligned} \tag{4.187}$$

where $\Psi(z,t)$ represents the injected wave.

If we multiply Eq. (4.187), calculated at different bias step, by opportune occupation functions f_{FD} in the QW and in CB, and then performing an integration over the device volume, it is possible to extract the intra-band-to-band tunneling (intra-BTBT) current as a function of the applied bias. Figure 4.16 shows the result

of this calculation, referred to the same device previously described in the text. An interesting behavior can be observed: in correspondence of the point labeled with "1" the current experiences a local maximum since at that applied bias the electron quasi-Fermi energy is crossing the first confined level in QW. Then the curve rises again by virtue of a progressive occupation of the second level. Also note the finite current when the device is at equilibrium.

The problem we solved so far represents a considerable simplification of what is the behavior of electrons in real devices in the sense that our BC refer to contacts that are supposed to be very far from the active region, such that only a small subdomain just around the well contributed to the BTBT current. This means that travelling waves (and, thus, their current density) have been computed within a sort of pseudo-open BC where quantum particles does not really exit from device contacts, i.e. without accounting for the environment in which the device operates. Now we want to overcome this mathematical trick and introduce in our formalism the features of an open system, by directly acting on its microscopic description.

Let us start, so, with introducing what we call "*leads*", namely two segments of device connecting the core to the extremal contacts. Left and right contacts plus their respecting leads are modeled by introducing two nodes (say, z_0 and z_{N+2} respectively) in our discretization scheme. They represent semi-infinite prolongations of the device, each of which is folded in one single point per side, that connect it to the external environment. Suppose also that we are injecting on the left contact a quantum particle (i.e. an electron) described by the progressive wavefunction

$$\overrightarrow{\Psi}_{\mathrm{L}}(z) = 1 \cdot \mathrm{e}^{i\kappa_{\mathrm{L}} z} \tag{4.188}$$

whose amplitude, for simplicity, is chosen to be unitary and where κ_{L} is its wavevector. To make the formalism complete we have to account for a reflecting component on the left contact given by the regressive wave

$$\overleftarrow{\Psi}_{\mathrm{L}}(z) = r \cdot \mathrm{e}^{-i\kappa_{\mathrm{L}} z}\,, \tag{4.189}$$

where r is the normalized reflection coefficient. This determines, at the end of the device (i.e. at the right contact), a transmitted wavefunction

$$\overrightarrow{\Psi}_{\mathrm{R}}(z) = t \cdot \mathrm{e}^{i\kappa_{\mathrm{R}} z}\,, \tag{4.190}$$

again a progressive wave, where $t = 1 - r$ is the normalized transmission coefficient and κ_{R} its wavevector at the contact on the right side. The set of equations which includes Eqs. (4.188), (4.189) and (4.190), is a new set of open-system boundary conditions which connect our device to the outside world.

Note that the overall expression of the wavefunction on the left side becomes

$$\begin{aligned} \Psi(z) &= \overrightarrow{\Psi}_{\mathrm{L}}(z) + \overleftarrow{\Psi}_{\mathrm{L}}(z) \\ &= 1 \cdot \mathrm{e}^{i\kappa_{\mathrm{L}} z} + r \cdot \mathrm{e}^{-i\kappa_{\mathrm{L}} z}\,. \end{aligned} \tag{4.191}$$

According to the BC just derived we know that at node z_0 – the point which connects the left lead to the device – the wavefunction reads

$$\Psi(z_0) = 1 + r\,, \tag{4.192}$$

while, at z_{-1},

$$\begin{aligned}\Psi(z_{-1}) &= \mathrm{e}^{-i\kappa_{\mathrm{L}}\Delta z} + r\cdot\mathrm{e}^{i\kappa_{\mathrm{L}}\Delta z} \\ &= \mathrm{e}^{-i\kappa_{\mathrm{L}}\Delta z} + \Psi(z_0)\mathrm{e}^{i\kappa_{\mathrm{L}}\Delta z} - \mathrm{e}^{i\kappa_{\mathrm{L}}\Delta z} \\ &= \Psi(z_0)\mathrm{e}^{i\kappa_{\mathrm{L}}\Delta z} - \left(\mathrm{e}^{i\kappa_{\mathrm{L}}\Delta z} - \mathrm{e}^{-i\kappa_{\mathrm{L}}\Delta z}\right)\,, \end{aligned} \tag{4.193}$$

where in the first line we used the relation $r = \Psi(z_0) - 1$ and where we also recalled that $z_{-1} - z_0 = -\Delta z$.

Now we can rewrite the eigenvalue equation written in Eq. (4.172) as

$$\Big(E_n[\mathbf{I}] - [\mathbf{H}]\Big)\Psi_n(z_i) = 0\,, \tag{4.194}$$

where $\mathbf{I}$ is the $(N+1)\times(N+1)$ identity diagonal matrix. Since Eq. (4.193) we know that, in order to account for our open-BC, all the diagonal terms of $[\mathbf{H}]$ have to be added to

$$\mathrm{H_L} = -\frac{\hbar^2}{2m^*\left(\Delta z\right)^2}\,\mathrm{e}^{i\kappa_{\mathrm{L}}\Delta z} \tag{4.195}$$

while the r.h.s. is no more null, and becomes the $(N+1)\times 1$ vector $[\mathbf{S}]$, where

$$[\mathbf{S}]_1 = \mathrm{S_L} = \frac{\hbar^2}{2m^*\left(\Delta z\right)^2}\left(\mathrm{e}^{i\kappa_{\mathrm{L}}\Delta z} - \mathrm{e}^{-i\kappa_{\mathrm{L}}\Delta z}\right)\,, \tag{4.196}$$

and

$$[\mathbf{S}]_i = 0\,, \quad \text{for} \quad i = 2,\ldots,N+1\,. \tag{4.197}$$

Eq. (4.195) introduces the contribution coming from the injected wavefunction and is the so-called *source term*. Fixing the energy of our system, the Schrödinger equation may be rewritten in the form

$$\Big(E[\mathbf{I}] - [\mathbf{H}] - [\Sigma_{\mathrm{L}}] - [\Sigma_{\mathrm{R}}] - [\Sigma_{\mathrm{scat}}]\Big)\Psi_n(z_i) = [\mathbf{S}] = i[\mathbf{g}]\,, \tag{4.198}$$

where Σ_{L} and Σ_{R} are called *self energies*, and are due to the influence of leads (and, in turn, of the environment) on the device. They are defined as

$$[\Sigma_{\mathrm{L}}] = \frac{\hbar^2}{2m^*\left(\Delta z\right)^2}\begin{pmatrix} \mathrm{e}^{i\kappa_{\mathrm{L}}\Delta z} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \tag{4.199}$$

and

$$[\Sigma_{\mathrm{R}}] = \frac{\hbar^2}{2m^*\left(\Delta z\right)^2}\begin{pmatrix} 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & \mathrm{e}^{i\kappa_{\mathrm{R}}\Delta z}\,. \end{pmatrix}\,. \tag{4.200}$$

Also Σ_{scat} is a self energy but it is referred to all inelastic scattering events occurring between channels at different energies E. Moreover, $[\mathbf{g}]$ is an *energy broadening function* and is the column vector $[\mathbf{g}_{\text{L}}]$ defined as

$$\begin{cases} [\mathbf{g}]_1 = \mathrm{g}_{\text{L}} = -i\left([\Sigma_{\text{L}}]_{1,1} - [\Sigma_{\text{L}}]^{\dagger}_{1,1}\right) \\ [\mathbf{g}]_i = 0\,, \quad \text{for} \quad i = 2, \ldots, N+1 \end{cases} \tag{4.201}$$

if the source is at the left contact, or $[\mathbf{g}_{\text{R}}]$ as

$$\begin{cases} [\mathbf{g}]_i = 0\,, \quad \text{for} \quad i = 1, \ldots, N \\ [\mathbf{g}]_{N+1} = \mathrm{g}_{\text{R}} = -i\left([\Sigma_{\text{R}}]_{N+1,N+1} - [\Sigma_{\text{R}}]^{\dagger}_{N+1,N+1}\right) \end{cases} \tag{4.202}$$

if the source is at the right contact (daggers refer to Hermitian conjugates). As their names suggest, broadening functions are correlated to the spread of each channel E in the energy domain which is due to particles probability to escape into the leads. A general form of the broadening line-shape $\sigma(E)$ is

$$\sigma(E) \simeq \frac{\mathrm{g}_{\text{L}} + \mathrm{g}_{\text{R}}}{(E-\epsilon)^2 + \left[\left(\mathrm{g}_{\text{L}} + \mathrm{g}_{\text{R}}\right)/2\right]^2}\,, \tag{4.203}$$

where ϵ is the diagonal term of the matrix

$$\mathrm{H}_{\text{L}} + \frac{[\Sigma_{\text{L}}] + [\Sigma_{\text{L}}]^{\dagger}}{2} + \frac{[\Sigma_{\text{R}}] + [\Sigma_{\text{R}}]^{\dagger}}{2}\,. \tag{4.204}$$

As pointed out by S. Datta [203], two important properties differentiate self energies from other Hamiltonians which account for boundary effects: first, Σ strictly depend on energy and, second, they are not Hermitian matrices. For this latter reason their action produce the effect of having an imaginary source term, which affects the broadening function $\mathbf{g}$.

Note that Eq. (4.198) is no more an eigenvalue equation since E does not form a discrete set of values but an energy continuum. If we decide to neglect inelastic scattering then we have to treat energies as independent channels, i.e. we are in the ballistic regime. Now, to solve the problem, we can find the solutions of Eq. (4.198) as we did with the Schrödinger equation or we might notice that its l.h.s. can be expressed by the equation

$$E[\mathbf{I}] - [\mathbf{H}] - [\Sigma_{\text{L}}] - [\Sigma_{\text{R}}] = [\mathbf{G}]^{-1} \tag{4.205}$$

where $\mathbf{G}(E)$ is the $(N+1)\times(N+1)$ non-equilibrium Green's function (NEGF) matrix. From this function we can construct all the relevant quantities of the system as, for instance, its density operator. In order to do that let us firstly define Γ_{L} and Γ_{R} as the matrices composed by all zeros except for their main diagonals $[\mathbf{g}_{\text{L}}]$ and $[\mathbf{g}_{\text{R}}]$, respectively. So, being

$$|\Psi(z)|^2 = \Psi(z)\Psi^*(z) = \Big([\mathbf{G}][\mathbf{g}_{\text{L}}]\Big)\Big([\mathbf{g}_{\text{L}}]^{\dagger}[\mathbf{G}]^{\dagger}\Big) \tag{4.206}$$

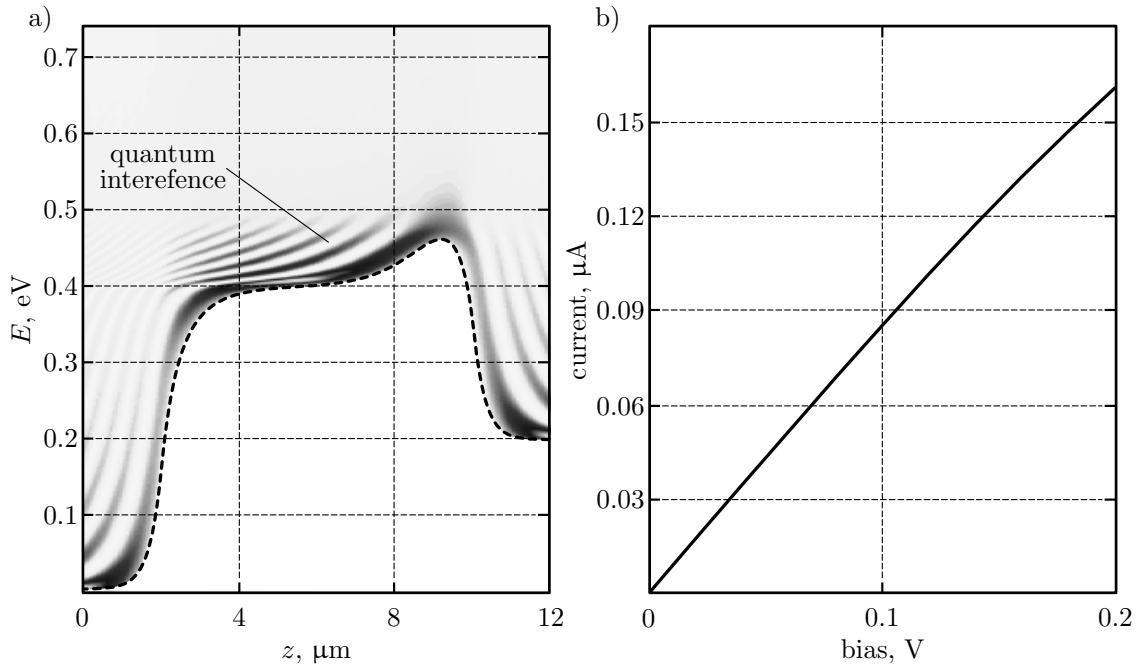

FIGURE 4.17. NEGF calculations performed for a n^+-n-n^+ GaAs structure: (a) Electron local density of states (LDOS) calculated at 0.2 V (dashed line represents the conduction band energy). (b) $I(V)$ characteristics.

then [203]

$$\begin{aligned}\hat{\rho}_{\mathrm{L}}(E) &= \int \mathrm{LDOS}_{\mathrm{L}}(z,E)\, f_{\mathrm{FD}}(E-E_{\mathrm{F_L}})\,\mathrm{d}E \\ &= 2\times\frac{1}{2\pi}\int \frac{\mathrm{d}\kappa}{\mathrm{d}E(\kappa)}\,|\Psi(z)|^2\, f_{\mathrm{FD}}(E-E_{\mathrm{F_L}})\,\mathrm{d}E \\ &= 2\times\frac{1}{2\pi}\int \frac{2m^*\Delta z}{\hbar^2\sin(\kappa\Delta z)}\,|\Psi(z)|^2\, f_{\mathrm{FD}}(E-E_{\mathrm{F_L}})\,\mathrm{d}E \end{aligned} \tag{4.207}$$

and similarly for $\hat{\rho}_{\mathrm{R}}(E)$. Putting together Eqs. (4.206) and (4.207) we have

$$\hat{\rho}_{\mathrm{L}}(E) = 2\times\frac{1}{2\pi}\int [\mathbf{A}_{\mathrm{L}}]\, f_{\mathrm{FD}}(E-E_{\mathrm{F_L}})\,\mathrm{d}E\,, \tag{4.208}$$

where the matrix

$$[\mathbf{A}_{\mathrm{L}}] = [\mathbf{G}][\Gamma_{\mathrm{L}}][\mathbf{G}]^\dagger \tag{4.209}$$

(and also $[\mathbf{A}_{\mathrm{R}}]$) is the $(N+1)\times(N+1)$ *spectral function* matrix.

Now, since

$$n_{\mathrm{L}}(z_i) = \int \mathrm{LDOS}_{\mathrm{L}}(z_i,E)\, f_{\mathrm{FD}}(E-E_{\mathrm{F_L}})\,\mathrm{d}E \tag{4.210}$$

corresponds to the diagonal of $\hat{\rho}_{\mathrm{L}}(E)$ and $n(z_i) = n_{\mathrm{L}}(z_i) + n_{\mathrm{R}}(z_i)$ also local density of states (LDOS) and carrier densities can be computed by means of the NEGF $\mathbf{G}(E)$.

In barrier-like problems, as tunneling ones, one can also compute the transmission coefficient and the related current respectively as [203]

$$\mathcal{T}(E) = \mathrm{Tr}\left\{[\Gamma_\mathrm{L}][\mathbf{G}][\Gamma_\mathrm{R}][\mathbf{G}]^\dagger\right\} \tag{4.211}$$

and

$$I_{\mathrm{L}\to\mathrm{R}} = \frac{2q}{h}\int \mathcal{T}(E)\, f_\mathrm{FD}(E_{\mathrm{F_L}} - E_{\mathrm{F_R}})\,\mathrm{d}E\,. \tag{4.212}$$

Figure 4.17 shows some results coming from NEGF simulations performed on a n^+-n-n^+ GaAs testing structure (the same considered in Ref. [204]) doped with $N_\mathrm{D}^+ = 2\cdot 10^{\ 19}$ cm^{-3} and $N_\mathrm{D} = 2\cdot 10^{18}$ cm^{-3}. In panel (a) we find the calculated LDOS of electrons, which generates typical quantum interference features. Panel (b), instead, shows the $I(V)$ characteristics calculated via Eq. (4.212) by solving the NEGF system for a set of potentials obtained at different bias.

In this last subsection we seen the genuine quantum formalism of non-equilibrium Green's functions for nanosystems and device simulation which includes, by its nature, a number of quantum effects (as tunneling) without introducing specific features or hypotheses. In Chapter 6 we will see how NEGF can be applied to simulate intra-BTBT in LEDs (like in Figure 4.15) however, for the moment, we still want to stress a couple of important considerations about NEGF. On the one hand it is worth noting that, passing through the definition of a density matrix, it is possible to build a formal link between Wigner function and Green's function frameworks. Indeed, by applying a Fourier transform to a particular Green's function $\mathbf{G}^< = i\mathbf{A}f_\mathrm{FD}$ called "G-*lesser*" the WF written in Eq. (4.154) can be recovered. For completeness, consider that $\mathbf{G}^<$ is defined via the standard Green's function $\mathbf{G}$ via

$$\mathbf{G} = \theta(z,z')\,(\mathbf{G}^> - \mathbf{G}^<)\ , \tag{4.213}$$

where $\mathbf{G}^> = -i\mathbf{A}(1-f_\mathrm{FD})$ is the Green's function "G-*greater*" and $\theta(a,b)$ is a function which assumes the value 1 if a is later than b in the system time-contour (called Keldysh contour) or is 0 otherwise. The Green's functions $\mathbf{G}^\lessgtr$ are useful, moreover, to define the lesser/greater boundary self-energies $\Sigma^\lessgtr$, which obey the Keldysh equation

$$[\mathbf{G}^\lessgtr] = [\mathbf{G}][\Sigma^\lessgtr][\mathbf{G}]^\dagger\,. \tag{4.214}$$

The second observation is that, although the Green's function here considered is at least a two-point one-energy function (according to the processes of interest also time and momentum can be considered), which implies a (6+1)-dimension function, its formalism turns out to be very useful since: first one has to deal only with matrices and, second, NEGF provides a robust formalism to solve the single-particle Schrödinger equation with open BC at different possible description levels, from effective mass approximation to tight-binding Hamiltonian, according to the specific problem one wants to study.

Chapter 5

Tunneling in HgCdTe IRPDs

Here we are going to present the author's area of research devoted to studying and modeling direct inter-band-to-band tunneling (inter-BTBT) phenomena in Mid-Wavelength Infra Red (MWIR) MCT photodetectors. In particular, results coming from works presented at the "US Workshop on the Physics and Chemistry of II-VI materials" on October 2014 in Baltimora (Maryland) and at the "42nd Freiburg Infrared Colloquium" on March 2015 (Germany), and also from a paper published in the *Journal of Electronic Materials* on April 24th, 2015 [173], will be proposed and analyzed.

Before presenting our results, let me thank first of all the Electronics and Telecommunications Department (DET) from Politecnico di Torino that allowed the use of TCAD (Technology Computer-Aided Design) tools needed for 2D simulations of IR photodetectors. Then a remarkable tribute also goes to Prof. E. Bellotti and collaborators from Boston University for their precious contribution in setting up the library of physical parameters for MCT material system. Lastly, not by importance, my gratefulness addresses to AIM Infrarot Module GmbH, the company that fabricated and tested all our HgCdTe structures, and in particular to W. Schirmacher, S. Hanna, and H. Figgemeier, whose remarkable contribution made possible this research.

5.1 Background and Motivations

The undiscussed importance of infrared (IR) light detection covers nowadays a wide range of applications (see, for instance, Ref. [205]), from biology to aerospace. Of particular interest are detectors based on the mercury cadmium telluride (MCT) material system [206,207]. As we learned in Chapter 2, this narrow-gap ternary alloy allows to vary the forbidden energy gap through the modification of the Cd molar

fraction, passing stoichiometrically from the semi-metal state (HgTe, for $x = 0$) to the semiconductor one (CdTe, with $x = 1$).

A role of primary importance is assumed by large format MCT photovoltaic IR detectors, namely the focal plane array (FPA) [7, 205, 208, 209]. Depending on their intended use, FPAs are designed for imaging in different regions of the IR spectrum, from SWIR to VLWIR. Within this spectral scale two operating environments can be identified: detection in the endo-atmosphere and in the exo-atmosphere. Endo-atmospheric imaging generally is intended for warm objects with high background irradiance, for example scattered sunlight or irradiance from the earth surface. In this case, SWIR and MWIR are the bands of interest. Exo-atmospheric applications instead deal with cool objects with low background irradiance levels and their operation regimes consist in the LWIR and VLWIR domains [210]. A large development effort has led to the fabrication of high performance MCT large format FPAs able to operate in only one of the possible spectral regions (Ref. [205] and references therein). These FPAs, normally referred to as one-"color" FPAs, are characterized by high operability and have been successfully employed in many systems. The requirements for a new generation of FPAs currently under development call for the possibility of detecting two or more colors at the same time. This is becoming of crucial importance for instance in the situation in which both imaging and object-background discrimination has to be performed, where the nature of the object is uncertain and/or the background itself is changing [210]. This result is actually achieved through both simultaneous [211–214] and sequential [215–218] detection schemes. The successful development of multicolor MCT FPAs requires further improvement of materials quality, adequate processing techniques and a better understanding about the sensor operation, either in terms of pixel performance or interaction between different pixels of the same array.

A second physical issue related to IR detection is represented by the temperature T, either of the scene to be captured and especially of the semiconductor lattice which constitutes each pixel. As a matter of fact, not only photon absorption can produce a signal, but also thermal generation can promote electrons to conduction levels. Such mechanism depends on T, obviously, and is inversely proportional to E_g. In order to get rid of cooling requirements, which noticeably increase production costs, modern FPAs are designed according to the High-Operating-Temperature (HOT) paradigm [60, 219], where devices (often n-on-p junctions) operate in nonequilibrium conditions such that carrier densities stay near their intrinsic concentrations (i.e. well below equilibrium levels).

A third (more fundamental) point to be considered in designing optimized IR photodetectors is the role of quantum processes, due to the trend of progressively shrinking the dimensions of electronic devices in accordance with Moore's law. Probably the most relevant quantum mechanism in this application field is direct tunneling (BTBT). The implications of tunneling in IR detection sensors are related to an electron flux generating an unwanted current flow, even without the presence of photons, i.e. in "dark conditions". Since dark current reduction is a prerequisite for HOT detection ($\geq$ 150 K) [220–223], we are interested in predicting the role of BTBT in IRPDs through physics-based modeling. Furthermore, numeri-

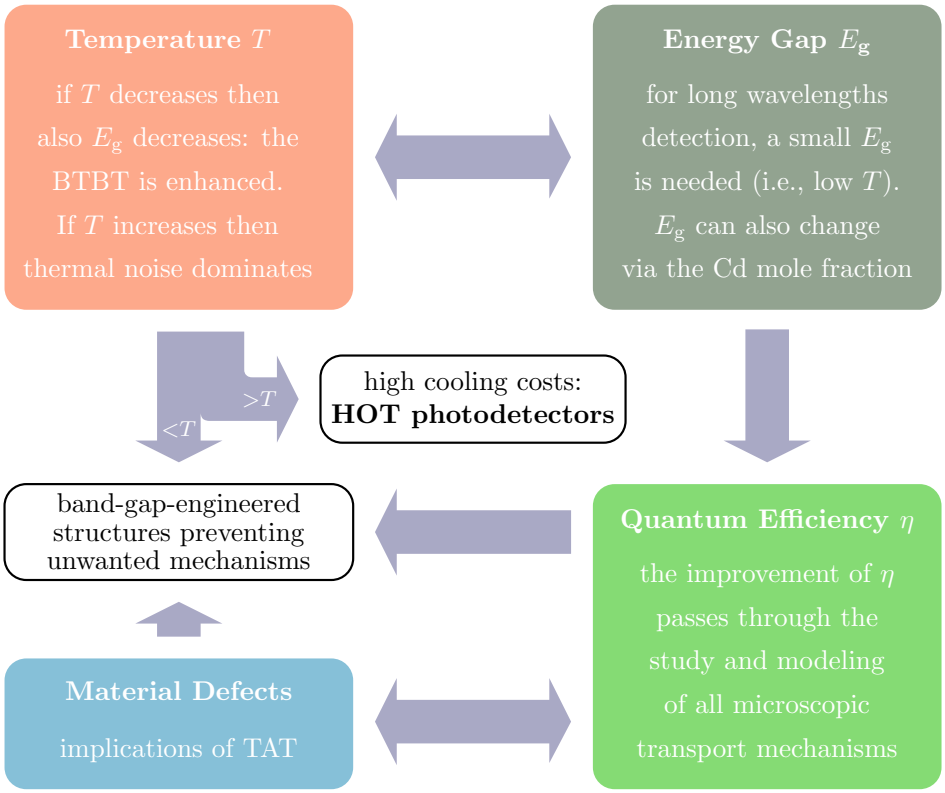

FIGURE 5.1. Simplified chart summarizing the most important physical and technical issues involved in the operating framework of solid-state IR detection with narrow gap materials. Temperature and band gap are mutually interacting and from thermal issues arises the dichotomy between cooled and HOT systems. The consequences of this technological choice are also affected – under the microscopical point of view – by the detector quantum efficiency, a parameter accounting for all the parasitic or intrinsic (quantum) phenomena, like tunneling.

cal simulation can also provide valuable guidelines for performance and reliability improvement of pixel structure and array geometries. The chart in Figure 5.1 summarized the most important physical and technical issues related to IR detection with narrow-gap semiconductors.

Being interested in HOT single-color FPAs composed by MCT pixels, our research is part of the scientific framework just outlined. So, in the next section the fabrication process of two examples of such devices will be briefly summarized, then we will take a look to their electrical dark characteristics before proceeding with numerical simulations.

5.2 Device Fabrication and Characterization

The objects we studied belong to a large detector matrix and are represented by single 15×15 μm pixels consisting in an MCT back-illuminated MWIR detector

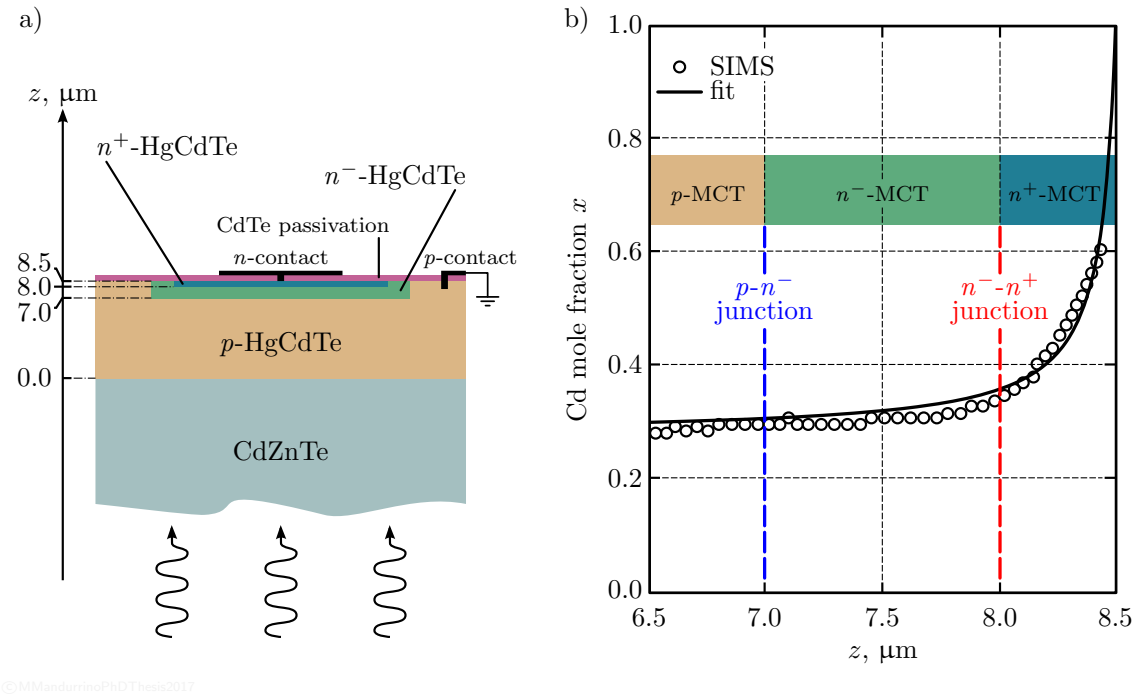

FIGURE 5.2. (a) Cross-sectional view of the MCT pixel constituting our p-n^--n^+ back-illuminated IR photodetector. (b) Profile of cadmium mole fraction as measured (symbol) via secondary-ion mass spectrometry (SIMS) and its least-square fit curve (solid line).

with cut-off wavelength at 80 K equal to $\lambda_{\text{cut-off}} \simeq 5.3$ µm. As one may observe in panel (a) of Figure 5.2, they are grown on CdZnTe substrate (via liquid-phase epitaxy, LPE) and passivated via CdTe superficial growth. In particular, we considered two structures that, apart from the different implantation techniques for the p-type doping, are nominally identical: device hereinafter labeled with "A" is doped via standard Hg-vacancy p-type technology by $N_{\text{A}_\text{A}} = 2 \cdot 10^{16}$ cm^{-3}, whereas device labeled with "B" is doped by implanting Au as acceptor, with nominal density $N_{\text{A}_\text{B}} = 5 \cdot 10^{15}$ cm^{-3} (and low vacancies density [224, 225]). Finally, the junction is obtained through ion implantation into the p-type layer which led to a maximum donor density $N_\text{D} = 10^{18}$ cm^{-3} in both devices.

The structure is composed as follows: starting at depth $z = 0$ from the surface of the CdZnTe substrate (not implemented in our simulations), we have a 7 µm-thick p-type MCT layer, then a 1.5 µm-thick and 10 µm-wide MCT n^--type region doped by $N_\text{D}^- = 4 \cdot 10^{14}$ cm^{-3} whose Cd fraction follows the profile represented in panel (b) of Figure 5.2. This layer incorporates a 0.5 µm of superficial n^+-type MCT in which donor concentration gradually increases according to an isotropic error-function trend that reaches $N_\text{D}^+ = 10^{18}$ cm^{-3} at the surface. A cathode electrode located at the center of the pixel, put in contact with the n^+-type region, and a p-contact deposed on the pixel corner complete the structure. In order to operate with the diode under reverse polarization the n-contact is biased at a positive voltage with respect to the (grounded) reference potential of the p-contact.

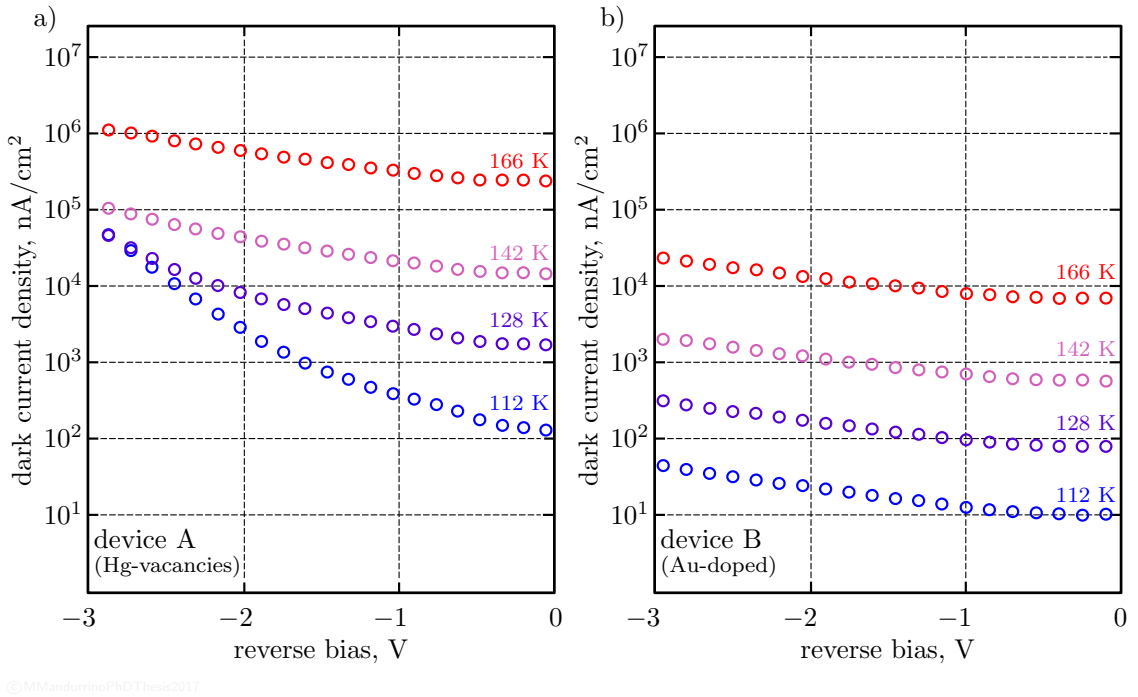

FIGURE 5.3. Dark current densities measured in our AIM Infrarot Module structures: (a) device A and (b) device B. Notice the higher currents in device A with respect to device B, both at high reverse applied bias and also at near-zero voltage (data published in Ref. [173]).

By using a liquid-nitrogen evaporation cryostat allowing to vary the temperature from 80 K to 300 K dark current measurements were performed. As we indicated in Ref. [173], the data was taken thanks to HgCdTe chips hybridized to fan-out circuits, with arrays of photodiodes connected in parallel in order to perform low current measurements. The pixels were surrounded by guard diodes independently biased such that each current contribution coming from outside the detectors are wiped out. Figure 5.3 shows the dark current characteristics for device A and B measured at different temperatures. As one may notice, device A is connoted by higher currents than device B for all bias regimes. This could be due, for instance, to higher defect densities originating from Hg-vacancies. In order to investigate this phenomenon, the eventual presence of tunneling mechanisms and, especially, the different electrical behavior of these two considered devices we set up a numerical investigation, presented in the following section, trying to propose a qualitative and quantitative explanation.

5.3 Simulation Technique

A realistic choice of material parameters is always of crucial importance in numerical simulation of electronic devices. In Chapter 2 we studied the physical/electrical properties of the $Hg_{1-x}Cd_xTe$ material system and, besides this, we also analyzed the most used models describing both optical and transport mechanisms within bulk

MCT. Here we apply all these knowledge and, in particular: E_g has been modeled as a function of temperature and mole fraction according to Eqs. (2.17)–(2.18) and Refs. [32, 33], the electron affinity χ follows the formulation of Ref. [43, p. 1331], for what concerns electron effective masses and mobilities we respectively followed Eq. (2.13), as in Ref. [27], and Eq. (2.20), as in Ref. [7], while for holes $m^*_h = 0.55\, m_0$ (as widely accepted) and $\mu_h(x,T)$ is given by Eq. (2.19), as in Ref. [30]. Regarding recombination processes, we used radiative and Auger lifetime expressions as in Ref. [50] (see, respectively, Eqs. (2.51)–(2.52) for τ_{rad} and Eqs. (2.56)–(2.64) for τ_A), where, as reported in Refs. [226], the Bloch-wavefunction overlap is $|F_{1,2}| = 0.2$ and $\gamma = \tau_{i,A7}/\tau_{i,A1} = 6$. For a detailed description of models and parameters implementation into the numerical solver the reader can also refer to Appendix A.

A particular case is represented by the Shockley-Read-Hall (SRH) generation/recombination. Since we used this mechanism for the preliminary tuning procedure of our simulation setup, the value of SRH parameters have been empirically estimated among those which better allowed to reproduce experimental data with simulations. The results coming from this tuning, which interested the low-bias range ($V_{bias} \leq 0.5$ V) in both device characteristics, have been reported in Figure 5.4 and the related parameters are listed in Table 5.1.

TABLE 5.1. Preliminary tuning of simulation parameters

REGION	σN_{trap} (cm^{-1})	E_{trap} (meV)
device A:		
p-HgCdTe	0.01	30
n-HgCdTe	0.80	70
device B:		
p-HgCdTe	0.0010	30
n-HgCdTe	0.0018	120

List of SRH-related parameters set through a fine tuning of our simulations on the experimental $J(V,T)$ dark current characteristics, whatever the temperature: E_{trap} is the energy of the defect level, measured from the conduction band edge E_C and σN_{trap} is the product of the trap cross-section σ and its density N_{trap} within the device.

The product σN_{trap} is essential for the definition of a proper SRH lifetime since, as already mentioned in this work (see, for instance, Eq. (4.75)),

$$\tau_{n,p} = \frac{1}{\sigma_{n,p} N_{trap}\, v_{th_{n,p}}}\,, \tag{5.1}$$

where

$$v_{th_{n,p}} = \sqrt{\frac{3k_B T}{m^*_{e,h}}}\,, \tag{5.2}$$

as usual for fermions (and as already stated in Eq. (1.34)), and where $\sigma_n = \sigma_p$ is kept constant [227] and only the product σN_{trap} is changed.

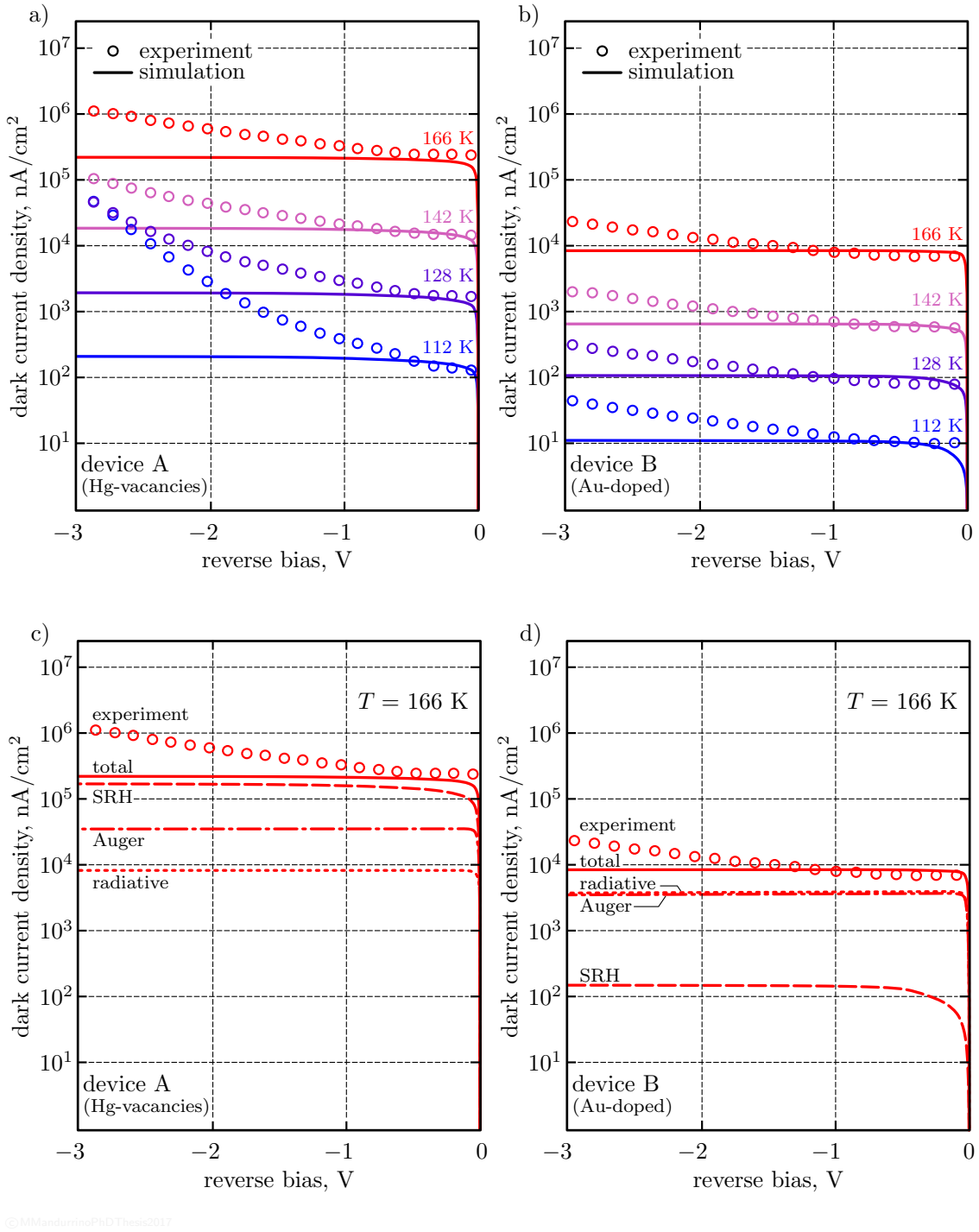

FIGURE 5.4. Experimental (symbols) and calculated (solid lines) dark $J(V,T)$ characteristics for (a) device A and (b) device B at different temperatures. (c) Detail of single GR contributions at $T = 166$ K calculated for device A and, in the same way, (d) for device B. Simulations have been performed including parameters listed in the text and in Table 5.1.

By using a commercial TCAD tool [192] which solves the DD model through an adaptive Delaunay-Voronoï discretization procedure (generating more dense grid where carrier and current density gradients are higher), we exploit the common SRH net rate expression

$$U_{\mathrm{SRH}} = \frac{np - n_i^2}{\tau_p \left(n + n_i \, \mathrm{e}^{\frac{E_{\mathrm{trap}} - E_{\mathrm{F}_i}}{k_{\mathrm{B}} T}} \right) + \tau_n \left(p + n_i \, \mathrm{e}^{\frac{E_{\mathrm{F}_i} - E_{\mathrm{trap}}}{k_{\mathrm{B}} T}} \right)} , \tag{5.3}$$

where E_{trap} here is measured from conduction band E_{C}. The other GR mechanisms have been implemented as already explained at the beginning of this section. Moreover, all the contacts have been assumed as ideally ohmic and the eventual electrical/optical crosstalk effects have been considered negligible, due to the measurement system already described.

As we reported in Ref. [173], and according to Ref. [60], E_{trap} lies very close to the intrinsic Fermi level E_{F_i} both in n-type and p-type HgCdTe doped with As, Cu or Au, while it is expected to be approximately 30 meV below E_{C} in vacancy-doped MCT when $0.2 < x < 0.5$. Considering that the depletion region of both devices mainly extends in the n-layer, the SRH parameter in the p-layer can be reasonably considered irrelevant for dark currents. Taking this into account, we set $E_{\mathrm{trap}} - E_{\mathrm{V}} = 30$ meV for p-type HgCdTe in both devices, as reported in Table 5.1. All the values written there are intended to be valid whatever the device temperature T.

So, for each device only three SRH-related parameters remain free. By setting $\sigma_n N_{\mathrm{trap}}$, $\sigma_p N_{\mathrm{trap}}$ and E_{trap_n} as reported in the same table all the calculated characteristics in Figure 5.4 have been obtained. Panels (a) and (b) show, respectively, the measured and simulated dark characteristics of device A and B at different temperatures, where simulations are performed including in the GR term of the continuity equation SRH, radiative and Auger processes. In panels (c) and (d), instead, only the cases corresponding to the higher temperature ($T = 166$ K) are considered. Here each simulated curve (apart that one labeled as "total") is calculated accounting only for one GR mechanism at time, as if it were the unique contribution to the dark current density. Due to nonlinearity of DD equations, the sum given by considering each single GR term not necessarily coincides with the curve (solid line) obtained when

$$R_{\mathrm{total}} = R_{\mathrm{SRH}} + R_{\mathrm{rad}} + R_{\mathrm{A}} , \tag{5.4}$$

and for this reason we speak in terms of simple contributions, rather than true algebraic components. Anyway, dashed curve refers to the SRH contribution (i.e. $R_{\mathrm{total}} = R_{\mathrm{SRH}}$), dotted curve to radiative ($R_{\mathrm{total}} = R_{\mathrm{rad}}$) and dot-dashed one to the Auger contribution ($R_{\mathrm{total}} = R_{\mathrm{A}}$).

Once we specified such important modellistic considerations, and remembering first of all that these simulations represent only the preliminary step of our investigation (since they have been used just to define a realistic setup for parameters, focusing only on the low-bias regime below 0.5 V), some considerations arise:

1. According to our simulations (see, in particular, panels (c) and (d) of Figure 5.4) device A seems to be SRH-limited since the nonradiative trap-assisted contribution appears as the dominant with respect to radiative and Auger ones. At the same time, device B should not experience this behavior since it is impossible to identify a major contribution. Moreover, SRH contribution in device B is approximately three orders of magnitude lower than in A.

2. The difference just pointed out might be attributable to a less vacancy density of device B with respect to device A resulting, probably, in a lower defect density. As a matter of fact, the tuning procedure gave a smaller product $\sigma_p N_{\text{trap}}$ in the second device: since we assumed that the trap cross-section is constant and always the same, it should be that N_{trap} is lower in device B than in device A. This difference is about three orders of magnitude, as for SRH contributions.

3. Auger reduction in device B by one order of magnitude can be ascribed to a lower acceptor density. This effect is in accordance with theoretical discussions of Ref. [60] and has also been experimentally observed in Ref. [222].

4. Finally, we should not overlook that the three GR mechanisms invoked in these preliminary simulations can not reproduce the experimental dark characteristics at high/medium voltages. Some other processes should be introduced and, since the behavior of measurements seems governed by field-dependent processes with a threshold-like onset, we will try to implement impact ionization (II) and tunneling (TAT, BTBT) models.

5.4 Results: Tunneling Models at Work

We tested the presence of direct tunneling (BTBT) by using the novel expression due to Mandurrino [45] explained in detail in Section 4.2. It is worth remembering that such formalism has been conceived for narrow-gap p-i-n diodes in reverse bias. We are dealing with a p-n^--n^+ structure (with $N_{\text{D}}^- = 4 \cdot 10^{14}$ cm^{-3}) that can be safely considered a suitable case-study for such model since, as demonstrated by the plot in Figure 5.5, the (longitudinal component of the) electric field assumes an almost constant profile across the junction, as should occur in true p-i-n detectors. This property satisfies the only theoretical requirement for the application of the generation rate $G_{\text{Mandurrino}}$.

By setting $D = 1$, Eq. (4.100) becomes

$$G_{\text{Mandurrino}} = A \cdot \mathcal{E} \cdot \exp\left(-\frac{B}{\delta\sqrt{\mathcal{E}}}\right) \tag{5.5}$$

where the unique parameter resulted, from the fitting, to be $\delta = 210$ V$^{1/2}$cm$^{-1/2}$. Since the TCAD simulation suite does not include such *ad hoc* model among its

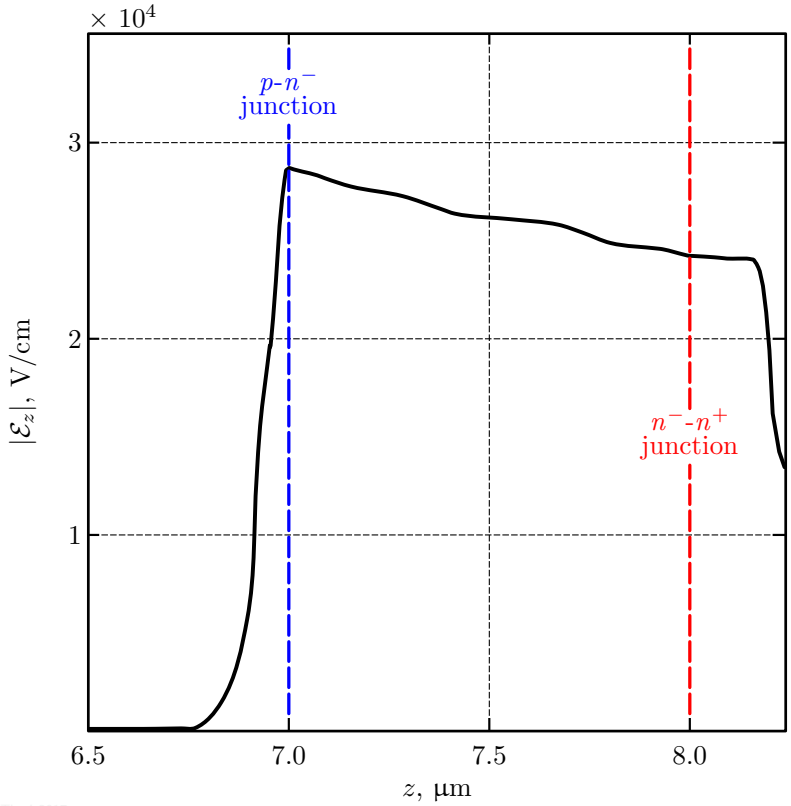

FIGURE 5.5. Vertical component of the electrical field along the two main photodiode junctions simulated in device A at −3 V.

built-in tools, it has been implemented via an external interface allowing to enter customary codes. The C++ routine used to implement he generation rate in Eq. (5.5) into DD equations is reported in Appendix C.

For what concerns TAT, instead, the standard Hurkx formalism has been activated, where the same trap-related properties referred to SRH process and reported in Table 5.1 have been used. Finally, II is included in our simulations through a post-processing procedure which exploits the bias-dependent gain $M(V,T)$ as

$$J_{\text{II}}(V,T) = M(V,T)\, J(V,T)\,, \tag{5.6}$$

(see a very similar trend for APD detectors in Eq. (2.45)) where $J(V,T)$ is the simulated dark current density and the gain $M(V,T)$ makes use of the model [228]

$$M(V,T) = \exp\left[a(T)\, V\, \exp\left(-\frac{V_{\text{th}}}{V}\right)\right], \tag{5.7}$$

being $a(T)$ and $V_{\text{th}} = \mathcal{E}_{\text{th}}\, w_{\text{dep}}$, an opportune threshold potential, the only parameters of this model. We adopted this approximated strategy in place of implementing built-in II models (as the Okuto-Crowell one [229]) since usually they are time-consuming.

Since $w_{\text{dep}} \approx 1$ µm and being $\mathcal{E}_{\text{th}} \approx 30$ kV/cm a typical value in HgCdTe [230], we obtained that $V_{\text{th}} \approx 3$ V, while $a(T)$ remains a free fitting parameter.

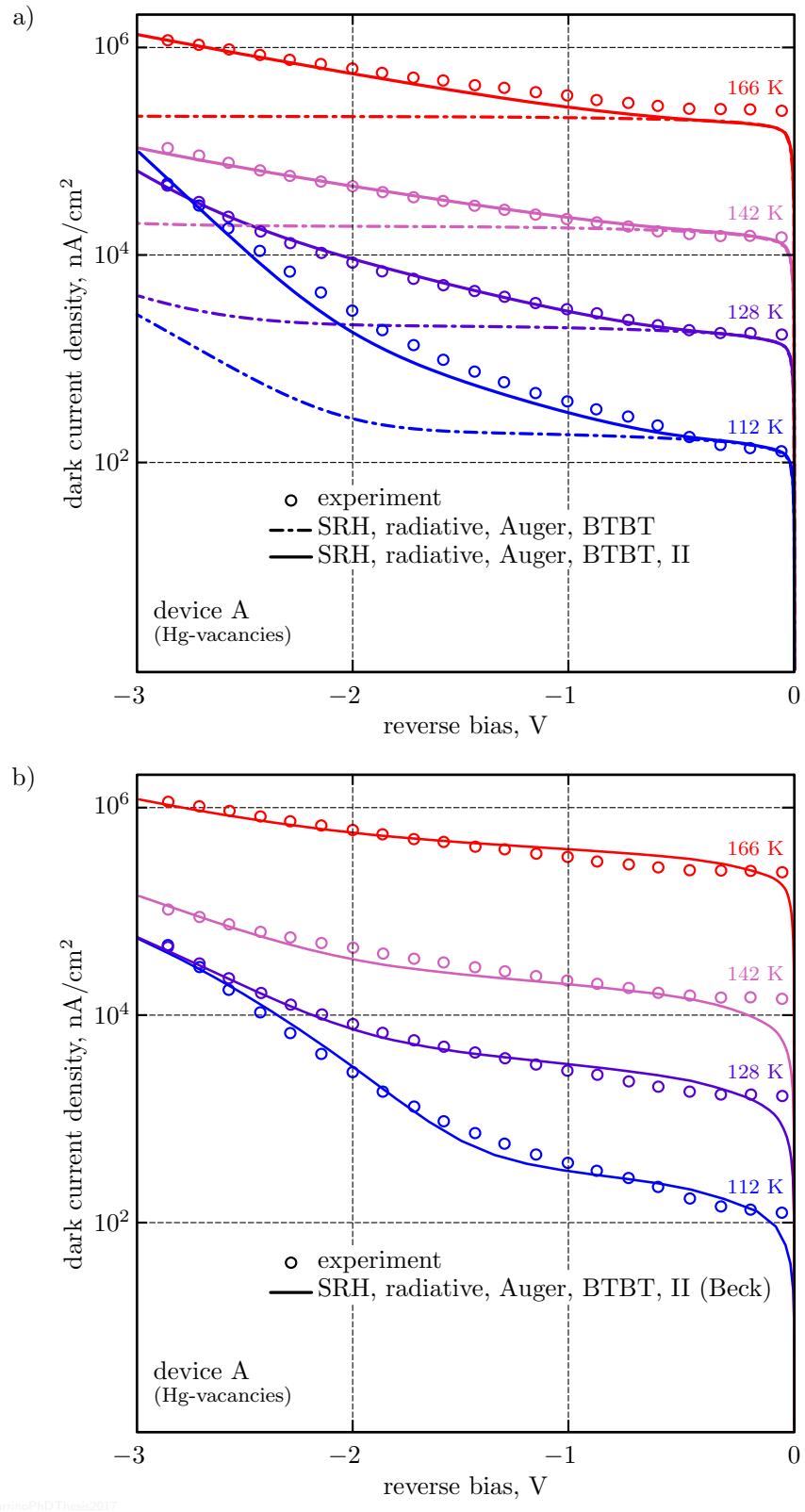

FIGURE 5.6. (a) Measured and calculated dark current densities for device A where simulations consider SRH, radiative, Auger, BTBT (dot-dashed) and also II processes (solid), as published in Ref. [173]. (b) Author's personal reinterpretation of the same data obtained with $\delta = 290$ V$^{1/2}$cm$^{-1/2}$ and introducing the II Beck's model with a parameter $V_{\rm th}(T) = 6.8\,E_{\rm g}(T)/q$ slightly variant by reason of T (according to Refs. [225, 231]) in place of arbitrarily tuning $a(T)$.

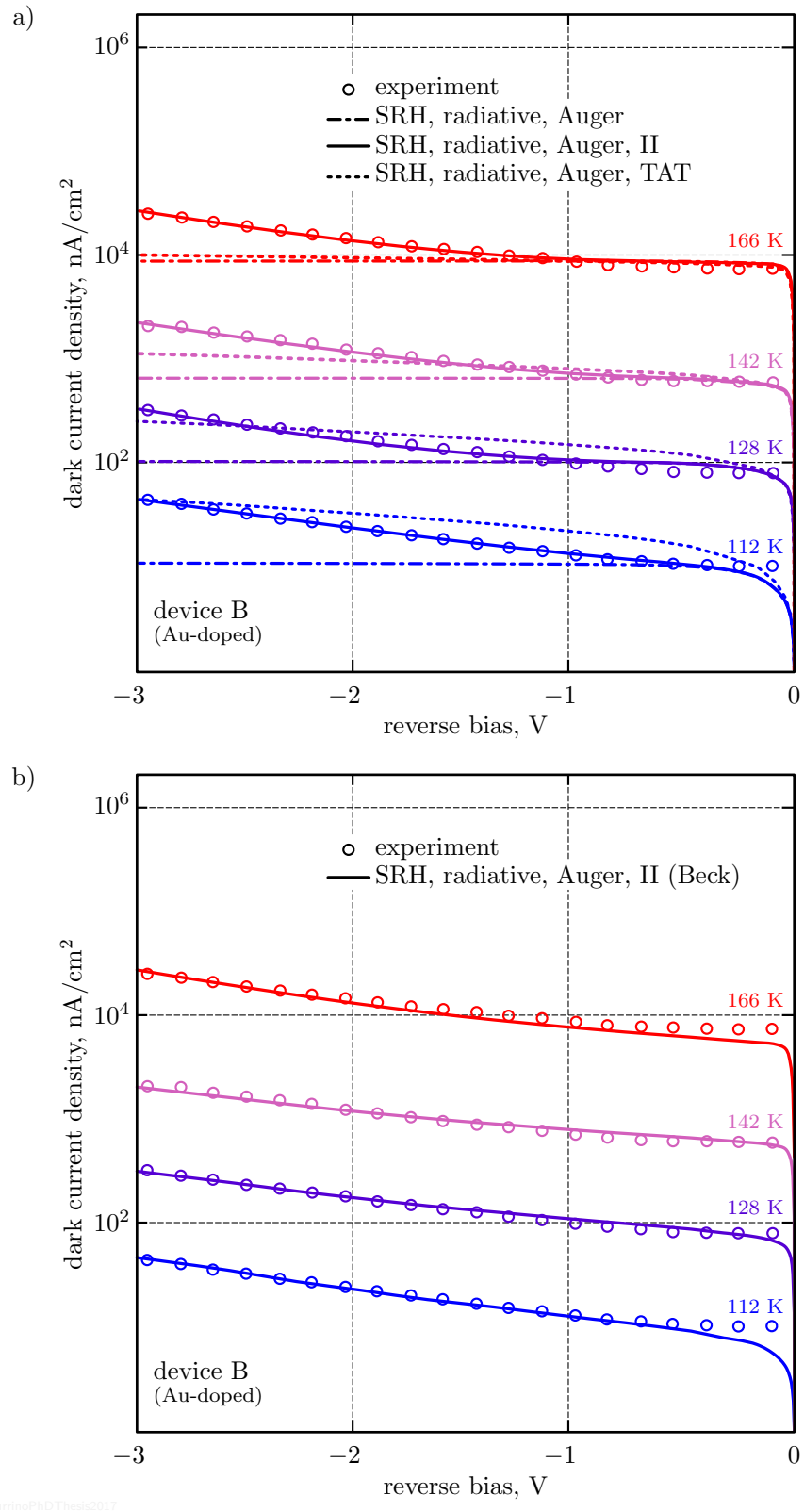

FIGURE 5.7. (a) Measured and calculated dark current densities for device B where simulations consider SRH, radiative, Auger (dot-dashed) and also II processes (solid), as published in Ref. [173]. (b) Alternative approach in simulating dark current densities for device B where SRH, radiative, Auger and II processes (through Beck's model) have been considered.

5.4.1 My novel inter-BTBT formulation

Through the simulation procedure described above we obtained the curves reproduced in Figure 5.6 (device A) and 5.7 (device B). Let us focus now on device A. Simulations represented in panel (a) with dot-dashed lines have been obtained solving the DD equations which include as source terms the SRH, radiative, Auger and BTBT (with the novel formulation presented in this work) mechanisms. The SRH parameters are still the same reported in Table 5.1 while, concerning BTBT, the value $\delta = 210\ \mathrm{V}^{1/2}\mathrm{cm}^{-1/2}$ (constant with T) is the one allowing best agreement with measurements. Solid line curves, instead, are calculated by adding a further term in continuity equations, namely impact ionization (II). In this case Eqs. (5.6) and (5.7) have been used in post-processing, where the values of $a(T)$ coming from the fitting procedure are reported in Table 5.2.

TABLE 5.2. Impact ionization (II) parameters for device A

	panel (a) of Figure 5.6	panel (b) of Figure 5.6
TEMPERATURE, T (K)	$a(T)$ (V^{-1})	$V_{\mathrm{th}}(T)$ (V)
112	0.98	1.6517
128	1.52	1.6722
142	0.91	1.6987
166	0.99	1.7327

List of impact ionization (II) parameters: $a(T)$ refers to the model written in Eq. (5.7) whereas $V_{\mathrm{th}}(T)$ is the main physical quantity entering the Beck's model, calculated through Eq. (5.9).

Although the results reported in panel (a) of Figure 5.6 are quite satisfactory, by the physical standpoint this way of modeling the thermal trend of II is not fully realistic: indeed, $a(T)$ is only a parametric handle one can tune and choose according to the best interpolation curve. A less arbitrary but more rigorous way to proceed, in the author's opinion, consists in introducing a model which captures the correct quantity actually depending on T: this quantity is the threshold voltage V_{th} (that previously was *a priori* fixed as $V_{\mathrm{th}} \approx 3$ V). It strictly depends on the threshold electric field, which determines the II onset of the diode characteristics, and is T-dependent since the field along the depletion region varies with T by reason of the energy gap $E_{\mathrm{g}}(T)$. So, it has been decided to test an alternative approach (see panel (b) of Figure 5.6) by introducing the so-called Beck's model [230, 231]:

$$M_{\mathrm{Beck}}(V,T) = 1 + 2^{\frac{2(V-V_{\mathrm{th}}(T))}{V_{\mathrm{th}}(T)}}\,, \tag{5.8}$$

where a trend considered typical for HgCdTe is [225]

$$V_{\mathrm{th}}(T) = 6.8\,\frac{E_{\mathrm{g}}(T)}{q}\,. \tag{5.9}$$

Using Eq. (5.9), which intrinsically introduces the role of T, we no longer need to emulate the thermal behavior of II by means of external parameters or arbitrary fitting procedures. So, only thanks to parameters related to physical properties of the materials, the simulations of panel (b) have been obtained. It is also worth noting that, as expected from theory and as one may observe in the trend of measured characteristics, V_{th} increases with the increase of T. On the contrary, $a(T)$ does not vary monotonically with temperature.

The second main feature distinguishing simulations of panel (a) and those of panel (b) concerns the implementation of Mandurrino's BTBT formalism. In the first case $\delta = 210\ \text{V}^{1/2}\text{cm}^{-1/2}$ has been chosen, again, by means of tuning procedures. In the latter case, instead, a proper value corresponding to $\delta = 290\ \text{V}^{1/2}\text{cm}^{-1/2}$ has been calculated from physical/geometrical parameters of device A and then used in all the simulations. At any rate, these two modeling strategies plotted in Figure 5.6 return quantitatively comparable and also quite satisfactory results.

5.4.2 Hurkx TAT model

In case of device B the direct tunneling contribution has found to be negligible. For this reason its simulations account only for SRH, radiative and Auger GR processes (dot-dashed curves in panel (a) of Figure 5.7). Besides these mechanisms also II (solid) or Hurkx TAT (dotted) have been tested. As one may notice in the same plot, the best accordance between simulations and experimental data has been found through the inclusion of impact ionization, whose parameters are listed in Table 5.3.

TABLE 5.3. Impact ionization (II) parameters for device B

	panel (a) of Figure 5.7	panel (b) of Figure 5.7
TEMPERATURE, T (K)	$a(T)$ (V^{-1})	$V_{\text{th}}(T)$ (V)
112	0.79	1.6517
128	0.54	1.6722
142	0.62	1.6987
166	0.52	1.7327

As in previous table, list of impact ionization (II) parameters: $a(T)$ refers again to the model written in Eq. (5.7) whereas $V_{\text{th}}(T)$ enters into the Beck's model and has been calculated through Eq. (5.9). Note that parameters in the last column coincide with the ones of device A.

Since II simulations follow the same procedure discussed above, also in case of device B an alternative modeling approach involving again Beck's model has been attempted. This author's personal reinterpretation can be found in panel (b) of Figure 5.7 and its related parameters are reported in the last column of Table 5.3 (but they coincide with the same parameters of device A written in Table 5.2). As one may see, the two approaches give almost the same results, both quantitative and qualitative. Anyway, a slight difference can be observed at the highest temperature,

where Beck's model underestimates the experimental current density in the low-bias regime.

5.5 Final Remarks

We analyzed two back-illuminated HgCdTe-based MWIR photodetectors constituted by nominally identical single-pixel structures differing only for the implanted p-type doping dose. Moreover, device A is characterized by standard Hg-vacancy doping (with $N_{\mathrm{A_A}} = 2 \cdot 10^{16}$ cm^{-3}) while device B is doped via Au implantation ($N_{\mathrm{A_B}} = 5 \cdot 10^{15}$ cm^{-3}).

Through the solution of drift-diffusion (DD) equations, obtained via a commercial TCAD tool, simulations of both device A and B have been performed. As previously discussed concerning preliminary simulations, device A has been found to be SRH-limited while no dominant processes can be observed in device B. This behavior probably stems from the higher defect density of the first device with respect to the second one. This fact seems to support the hypothesis according to which reducing the vacancies density is a prerequisite for HOT detection. Probably also the lower acceptor density in the absorber region contributes to this result.

Implementing the author's formalism (see Section 4.2 and Ref. [45]) to describe band-to-band tunneling (BTBT) and also including impact ionization (II) as post-processing correction via Eqs. (5.6) and (5.7), we obtained very satisfactory reproduction of measured reverse dark characteristics $J(V,T)$ for both devices in the range between 112 K and 166 K. This is the first time in the literature that tunneling rate $G_{\mathrm{Mandurrino}}$, theoretically derived in Ref. [45], is (proficiently) applied to model BTBT in real devices.

Through simulations we can infer that II is important in device A as well in device B, while TAT seems to be irrelevant in both structures. Instead BTBT has shown to be probably correlated with the doping concentration of the p-type region since it is present only in device A where N_{A} is four times higher than in device B. This could be explained through a different band bending which, in turn, affects the tunneling path. An alternative explanation could be that, in the bias range we considered, the onset of direct transitions has not yet occurred: a lower defect density could actually decrease the electric field in the junction such that BTBT arises at higher (reverse) voltages.

Finally, the reader can also find some unpublished results provided by the author which include a more realistic parametrization of II and BTBT through the rigorous description of thermal trends – in the first model – and of the δ parameter in Mandurrino's formalism – for what concerns the second model.

Chapter 6

Tunneling in InGaN/GaN LEDs

This chapter is intended to present the author's research in the field of theoretical investigation and physics-based numerical modeling of interband trap-assisted tunneling (TAT) in III-nitride blue light-emitting diodes (LEDs). The core of this production is constituted by essentially two journal papers and three conference papers. The first object has been submitted to the Special Issue of *Physica Status Solidi* on "Nitride Semiconductors" on September 30th, 2015 and published on January 28th, 2015 [87]. A second paper on these topics, as already mentioned in the present dissertation, has been published under invitation of his Editor in Chief J. Piprek in the *Journal of Computational Electronics* on January 21st, 2015 in the Special Issue entitled "Simulation of GaN-based Light-Emitting Diodes" [145]. Finally, the last three important papers consist in the work presented on September 1st, 2014 at NUSOD 2014 in Palma de Mallorca (Spain) [232] and the other ones presented on May 8th, 2015 at the Fotonica AEIT Conference on Photonics Technologies, in Turin (Italy) [233] and on September 8th, 2015 at the XIV SPIE International Conference on "Solid State Lighting and LED-based Illumination Systems" that was held in San Diego, California [234]. Besides such most representative papers, other works have been presented in minor or national conferences: see the complete list of publications reported at the beginning of the present thesis.

A second area of interest has been represented by some investigations on intra-band-to-band tunneling (intra-BTBT) involvinq quasi-confined states in the LED active region. Some preliminary results about this topic have been already presented at international conferences, as occurred for the invited talk given on February 15th, 2016 at the XXIV SPIE International Conference on "Physics and Simulation of Optoelectronic Devices" in San Francisco, California [235].

All the scientific production was possible thanks to the collaboration with the group of Padua University leaded by M. Meneghini, G. Meneghesso and E. Zanoni.

They provided experimental characterizations, as well as precious feedback on technology and material-related issues. Also the contribution of OSRAM Opto Semiconductors GmbH was fundamental since no research would be viable without their testing single quantum-well (SQW) structures. A special mention also goes to Prof. G. Verzellesi from Modena and Reggio Emilia University for motivating and fostering TAT investigations on LEDs and for his expert and valuable oversight on some theoretical and modeling aspects. Last but not least, the author aknowledges and thanks the Electronics and Telecommunications Department (DET) of Politecnico di Torino for all the logistic support, especially for what concerned TCAD licensing, and also the Italian University Consortium CINECA, for its computational facilities.

6.1 Background and Motivations

The barrier encompassing the active region of GaN-based light-emitting diodes (LEDs) usually provides a very small probability to allow direct (interband) tunneling transitions (i.e., BTBT) due to typically long tunneling paths. However, electrons and holes may experience a shorter distance if they encounter structural defects along their path that generate inter-gap states. Such defects are represented by point defects or by (extended) dislocations. So, besides behaving as SRH-centers of recombination, crystal defects may also play an essential role in the carrier transport framework as tunneling enhancers, a feature not always included in modeling studies.

The importance of defects over the electrical behavior of GaN-based LEDs has been observed and pointed out in the literature by many researchers [18,86,236–243]. However, quite rarely these works account at the same time for both data analysis and device simulations. The purpose of the author's Ph.D. involvement in this research area was to fill this gap by integrating a rigorous physical interpretation of electrical/optical measurements with self-consistent physics-based simulations of trap-assisted tunneling (TAT), where model parameters are inspired by such considerations on experimental data.

With this aim we intended to verify the hypothesis according to which GaN-based blue LEDs in forward bias conditions are affected by trap-assisted tunneling of electrons coming from conduction band (CB) in the n-type region of the diode and travelling "under" the quantum well (QW) towards defects located on the p-side and, at the same time, affected by TAT of holes originating from valence band (VB) in the p-type region and travelling "under" the QW towards defects in the diode n-side (see main picture of Figure 6.1). Strictly speaking in more physical terms, the microscopic interaction we suppose to occur is the same as depicted, for electrons, in panel (a): after a pure tunneling transition, the carrier scatters with an opportune occupation statistics of the oscillating trap by means of multiphonon emission (MPE) process, which preserves the energy conservation principle. After that, the electron may (non-radiatively) recombine, at the trap site, with an hole in VB. A symmetrical process may occur for holes, obviously.

Besides this relevant effect, an additional lower-order physical mechanism may

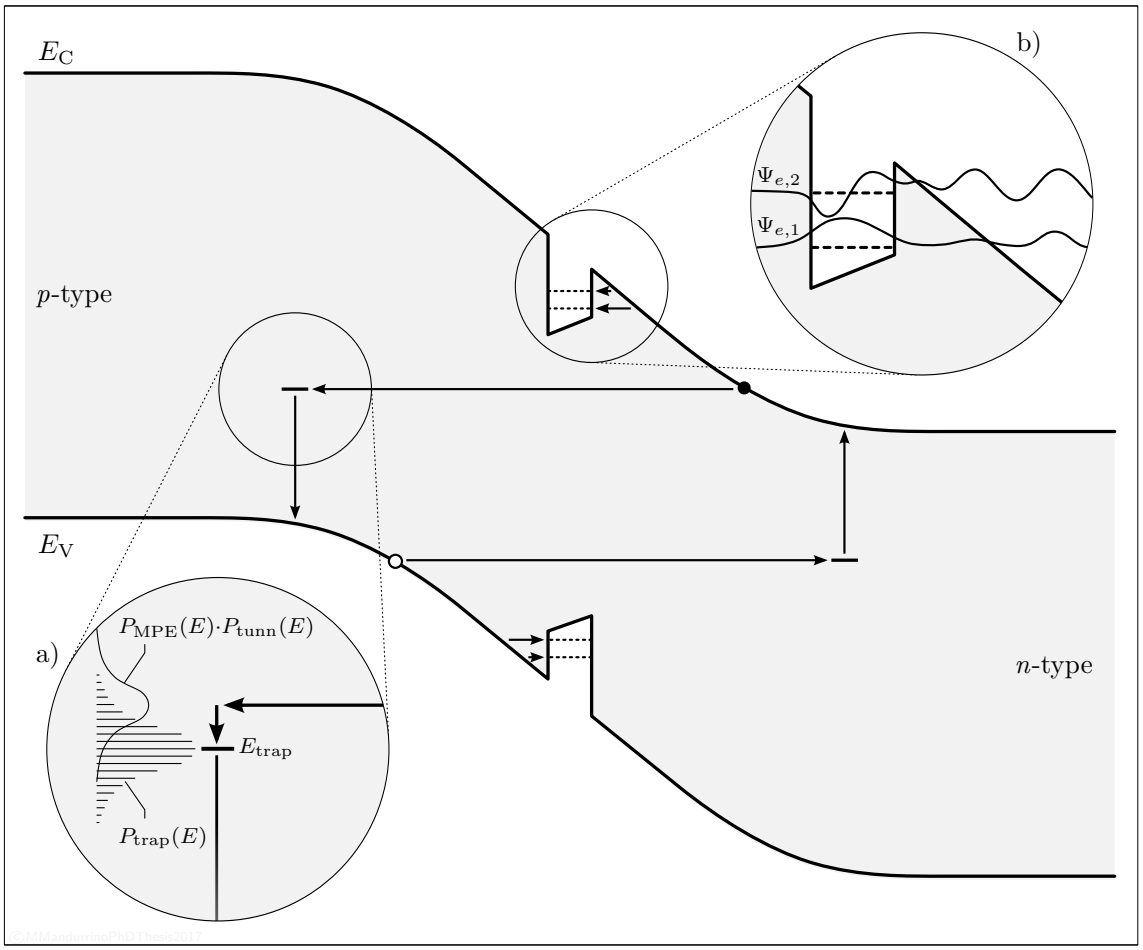

FIGURE 6.1. Schematic picture of different tunneling phenomena occurring in GaN-based single quantum well (SQW) blue light-emitting diodes (LEDs): (a) electron trap-assisted tunneling (TAT) and (b) electron intra-band-to-band tunneling (intra-BTBT) from conduction band towards the quantum well (energies not to scale).

take place: the quantization due to carrier confinement in the active region creates quasi-bound states (QBS) within the well. According to the classical picture, under zero or low injection regime typically only the ground state is occupied while levels characterized by higher energies are empty. However, even at equilibrium, the probability of having intra-band wavefunction leakages between QW and the surrounding barriers (on the n-side in case of electrons and on the p-side for holes) is not negligible. Thus, direct BTBT from band edges filling these unoccupied QBS may occur (see panel (b) of Figure 6.1), with important consequences on their occupation dynamics, on the radiative/non-radiative balance of currents and, probably, also on the LED efficiency droop. In order to investigate on this quantum effect – as far as we know, totally neglected in the today literature – part of my third-year Ph.D. efforts have been spent in studying and modeling intra-BTBT towards the active region of single-QW (SQW) GaN-based blue LEDs.

Figure 6.1 summarizes at a glance all the processes we shall see in the present chapter. In panel (a): TAT of electrons from conduction band in the n-side towards

a defect in the p-side at an energy given by the joint probability $P_{\mathrm{MPE}}(E) \cdot P_{\mathrm{tunn}}(E)$, then capture into the trap with energy E_{trap} via multiphonon emission according to the occupation probability $P_{\mathrm{trap}}(E)$ of the vibrational quasi-levels. A symmetrical process is supposed to occur for holes. In panel (b): electron intra-band-to-band-tunneling from CB in the n-side towards quasi-bound states in the quantum well. Again, a similar processes may occur in valence band for holes.

6.2 Devices Characterization and Data Analysis

As in the case of IR photodetectors, we analyzed two different blue light-emitters, both manufactured by OSRAM Opto Semiconductors. They are almost identical except for the indium content, 15% in device A and 18% in device B, the QW thickness, respectively of 3 nm and 1.5 nm, and finally the substrate, which is represented by highly conductive SiC in device A and Si in device B. In the following all the other (common) features: with area 250×250 $\mu\mathrm{m}^2$, devices are grown starting from a n^+-GaN buffer layer (with $N_{\mathrm{D}}^+ = 5 \cdot 10^{18}$ cm^{-3}), then they include a 10 nm-thick n^--GaN barrier doped by $N_{\mathrm{D}}^- = 10^{16}$ cm^{-3}, an n^--InGaN quantum well followed by a 2 nm-thick n^--GaN second barrier, both doped with the same doping concentration N_{D}^-. Finally, a 40 nm-thick p^+-$\mathrm{Al}_{0.15}\mathrm{Ga}_{0.85}\mathrm{N}$ electron-blocking layer (EBL) and a p^+-GaN contact region, both with $N_{\mathrm{A}}^+ = 2 \cdot 10^{19}$ cm^{-3}, complete the structure.

Both devices were electrically and optically characterized by our colleagues from Padua University. Some experimental data can be found in Figure 6.2. As widely accepted, it is possible to find some TAT signatures from the analysis of the experimental $I(V)$ characteristics [86, 88, 236, 237, 239–243]. Not only, also important parameters concerning traps can be deduced: for instance, X. A. Cao and coworkers [244, 245] found a correlation between growth quality and TAT by showing the implications of different growing techniques for fabricating InGaN/GaN structures on GaN and sapphire substrates. In these two works also the importance of trap density N_{trap} as the main feature determining the excess current under low-forward-bias conditions was pointed out. In effect, as one may notice from our experimental $I(V)$ characteristics, this operating regime is connoted by high ideality factors η (curves grow less rapidly). This behavior was already observed in 1996 by H. C. Casey *et al.* [246], who proposed the TAT via defects located in the active region as the mechanism which mostly contributes to generate such effect. Moreover, as they highlighted, the $I(V)$ slope at low-bias and in semilogarithmic scale is almost the same whatever T when TAT dominates. This implies a slight decreasing trend of η as a function of T. In fact, referring to the Shockley equation (see Eq. (3.59)), if the slope

$$\frac{\mathrm{d}[\ln(I)]}{\mathrm{d}V} = \frac{q}{\eta k_{\mathrm{B}} T} \tag{6.1}$$

is constant then η must decrease by reason of the increase of T.

In a paper appeared in the following issue of *Applied Physics Letters*, P. Perlin *et al.* [247] described essentially the same effect. In addition, they also highlighted

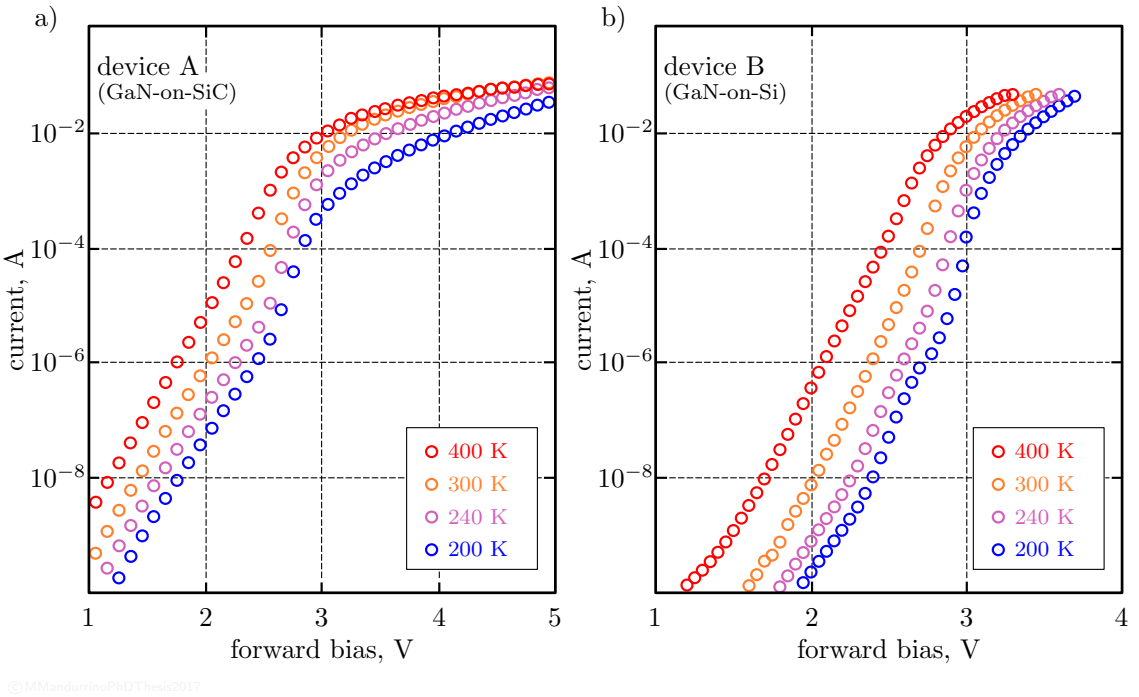

FIGURE 6.2. Subset of measured $I(V,T)$ forward characteristics of our single quantum well (SQW) InGaN/GaN blue light-emitting diodes (LEDs) manufactured by OSRAM Opto Semiconductors: (a) device A and (b) device B (see text for more details about the structures).

the presence of different bias-dependent transport regimes within each of the characteristics they measured, associated to as many ideality factors.

An important step towards a better comprehension of TAT implication on LEDs electrical behavior was put by L. Hirsh and A. S. Barrière [248] who noticed that, due to the marginal role of temperature on defect-related transport, the plot of $\log(I/V^2)$ versus V^{-1} is a straight line with T-independent slope. Few years later other researchers [245,249,250] came to the same conclusion, demonstrating that the rigorous explanation of this phenomenon involves the energy parameter E_0 defined as

$$E_0 = \eta k_B T\,, \tag{6.2}$$

a quantity strictly related to the tunneling barrier encountered by carriers during the first part of the process. Since the temperature has a negligible effect on the energy gap of GaN-based materials (see, e.g., Figure 3.7) and, in turn, on the shape of this barrier, then E_0 should be T-insensitive: when T increases, η decreases in such a way that E_0 remains almost constant.

A possible explanations about the the presence of different $I(V)$ slopes has been provided by resorting to the often used formula [237, 247, 249–251]

$$E_0 = \frac{4qh}{\pi}\sqrt{\frac{N_{SCR}}{\epsilon\, m^*_{tunn}}}\,, \tag{6.3}$$

where ϵ is the dielectric constant, N_{SCR} the reduced doping concentration at the

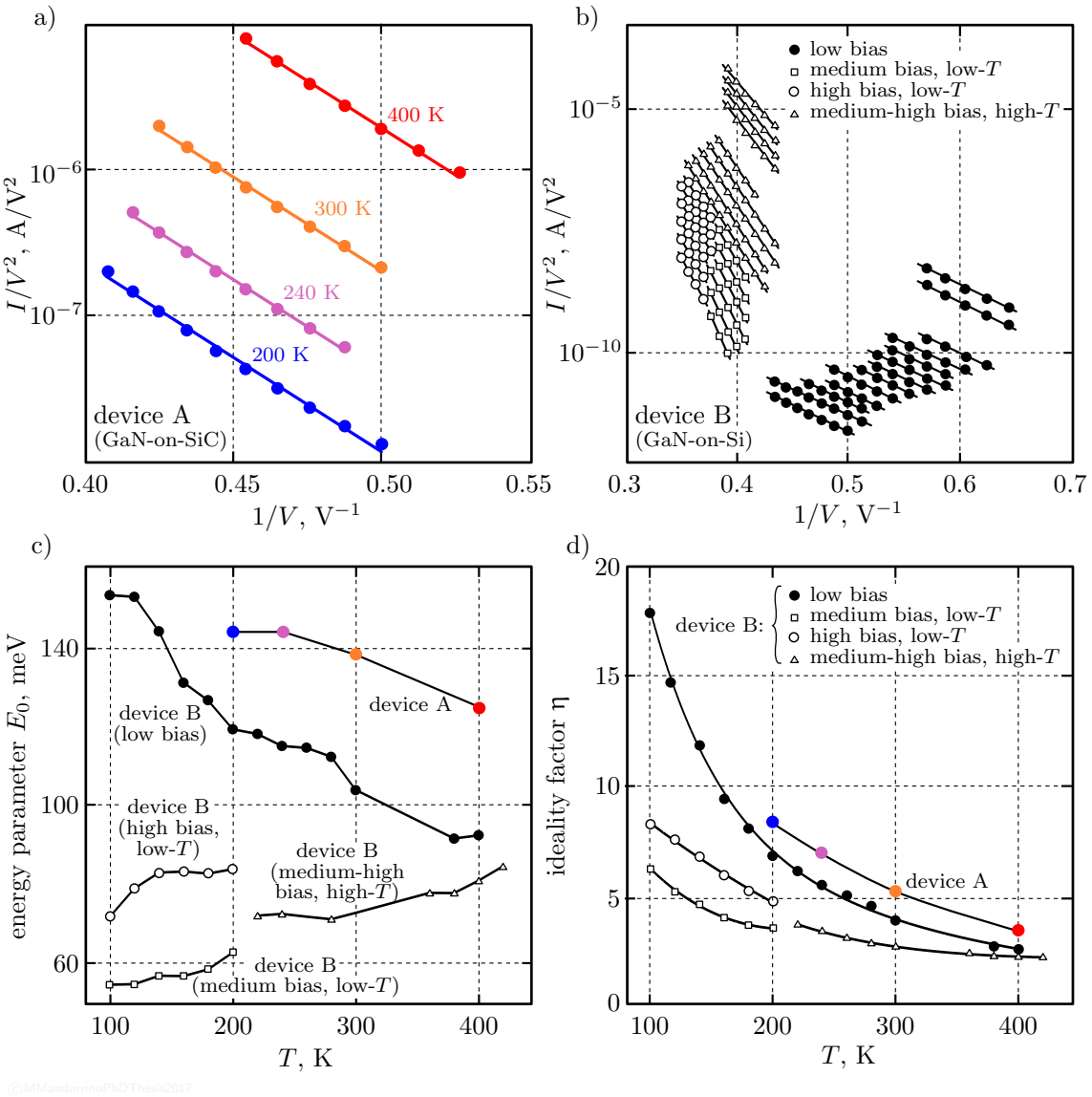

FIGURE 6.3. Trend of I/V^2 as a function of $1/V$ in device A (a) and device B (b): symbols represent calculated values while straight lines are the interpolating functions. Temperature-dependent energy parameters E_0 (c) and ideality factors (d) extracted from experimental forward $I(V,T)$ characteristics plotted in Figure 6.2. Bias-dependent data have been differentiated for device B.

edges of the space-charge region (SCR) and m^*_{tunn} the tunneling effective mass. Eq. (6.3) allows to justify the occurrence of different ideality factors and, in turn, of different E_0 by thinking each slope domain as a conduction regime generated by the tunneling of a specific particle with a proper effective mass m^*_{tunn}. As we shall see, this fact will be taken into consideration when we will introduce heavy hole tunneling.

More recently, in 2010, R. Nana and coworkers proposed an alternative mechanism at the basis of the observed difference in the slope of forward LED characteristics, which involves the energy of traps E_{trap} rather their concentration: by

stimulating the optically active defect levels of blue and UV emitters via electroluminescence (EL), the authors of Ref. [252] ascribed each bias-dependent slope domain of $I(V)$ to defect-assisted transitions occurring via traps with only specific energy. By changing the bias, the energy of allowed transitions changes accordingly, by selecting particular trap levels to assist the process. This selection phenomenon determines, at the end, different current channels with different transport properties and, finally, different $I(V)$ slopes.

We have seen that the decreasing trend of η with respect to T is a clear signature of TAT. Not only the behavior of the ideality factor as a function of temperature is crucial but also its value can give some information about the process from which it is generated. Ideality factors in the range from 1 to 2 are widely considered as coming from recombination transitions in the depletion region or diffusion-type transport in neutral regions [86,246,249–254], whereas tunneling usually determines $\eta \gg 2$ [250,255].

Figure 6.2 shows the forward electrical characteristics for both LEDs fabricated by OSRAM Opto Semiconductors (curves refer to a selected subset of all measurements, which have been performed in the range between 100 K and 420 K). Notice that, below the turn-on, device B is connoted by lower currents with respect to device A. By plotting the experimental I/V^2 at fixed T as a function of $1/V$ we obtained the data points reported in panels (a) and (b) of Figure 6.3. Straight lines are the fitting curves obtained through the simplified Shockley equation

$$I(V) \simeq I_0 \exp\left(\frac{qV}{E_0}\right) = I_0 \exp\left(\frac{qV}{\eta k_B T}\right), \tag{6.4}$$

where I_0 is a prefactor related to the diode saturation current. As one may observe, fit curves (which become straight lines in semilogarithmic scale) are parallel for both LEDs. Due to its more complex electrical behavior, we differentiate the trend of I/V^2 versus $1/V$ for device B into four sub-domains, denoted by different symbols: filled dots refer to the low bias regime (1.5 V - 2.3 V) whatever the temperature, squares to the medium bias range (2.45 V - 2.65 V) at low temperature (< 220 K), empty circles to high bias (2.65 V - 2.85 V) again at low-T and triangles to the medium-high bias regime (2.3 V - 2.8 V) at high-T (≥ 220 K).

By using Eq. (6.4) it is possible to extract from each fitting line the variations of E_0 and η with respect to the temperature. The values obtained from this procedure are plotted, respectively, in panels (c) and (d) of Figure 6.3. The almost constant trend of E_0 (except for the case at low-bias, that we shall discuss), the decreasing behavior of η versus T and its values, always greater or highly greater than 2, suggest the presence of trap-assisted tunneling. It is worth noting that, as pointed out in Ref. [249], $\eta(T)$ has a decreasing exponential behavior. In fact, solid line fitting curves in panel (d) of Figure 6.3 have been obtained through functions of the form

$$\eta(T) = a \cdot \exp(-b\,T) + c \cdot \exp(-d\,T). \tag{6.5}$$

We might observe that $T = 220$ K is a crucial temperature for device B since it separates the medium- and high-bias regime into two main regions of interest, both

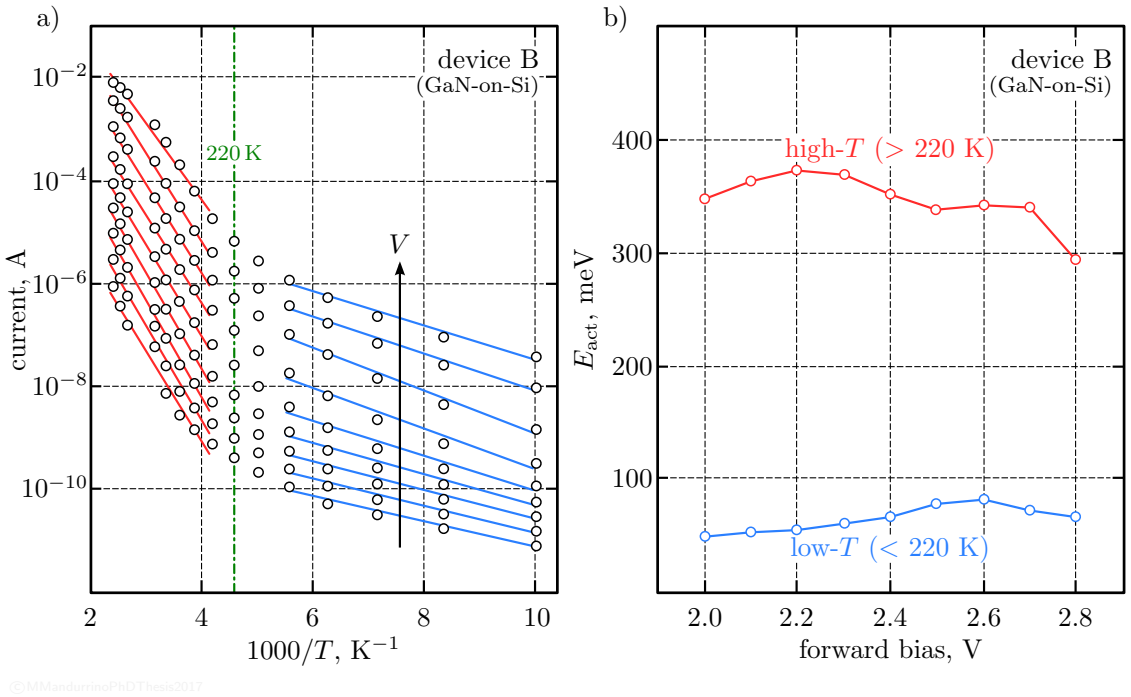

FIGURE 6.4. (a) Arrhenius plot and (b) extracted activation energies E_{act} as a function of the inverse temperature in device B. Each set of $I(1/T)$ data has been obtained from experimental characteristics for a set of fixed bias in the range 2.0 V to 2.8 V in eight steps (0.1 V each one).

for what concerns the energy parameter and the ideality factor. For T above this value the thermal trend of these quantities seems univocal, while below it at least two behaviors have been observed. The other interesting feature about device B is that the values assumed by the energy E_0 are willing in very different ways depending on the bias range. At high applied forward voltages the energy parameter is nearly constant and always lower than 90 meV. For bias $V = 1.5$ V to 2.3 V, instead, E_0 starts from very high values ($\simeq$ 150 meV) at 100 K then it decreases rather rapidly with temperature and reaches values comparable with those of the higher bias regime. Since 90 meV has proven to be a distinctive threshold between electron and hole trap-assisted tunneling [237, 247, 249–251] this effect, as well as the role played by the temperature $T = 220$ K, deserves more attention. So, in the next subsection we will analtze in more detail the experimental data searching for further evidences of multiple particle tunneling mechanisms.

6.2.1 Supplementary analysis on LED-B about hole TAT

We present here some unpublished results from a deep analysis on device B looking at the possibility to have some specific fingerprints of hole tunneling. Following a well known procedure (see, e.g., Ref. [248]), a study of trap-assisted activation energies as a function of the temperature has been performed. The conceptual tool

we exploited is the well known Arrhenius law. Thus, from the relation

$$E_{\text{act}} \propto -k_{\text{B}}T \log(I) \tag{6.6}$$

the function

$$I \simeq \exp\left(-\frac{E_{\text{act}}}{k_{\text{B}}T}\right) \tag{6.7}$$

can be obtained, where E_{act} represents the TAT activation energy and also the slope of Eq. (6.7) plotted in semilogarithmic scale. Figure 6.4 presents, in panel (a), the result of performing a least-square fit with Eq. (6.6) on experimental $I(1/T)$ at fixed bias. As already observed, the critical temperature $T = 220$ K plays a distinctive role, dividing the data into two sub-domains: below it high activation energies are expected while above lower values should manifest. Extracting E_{act} from the slope of each line a plot like the one reported in panel (b) can be obtained. From this graph one can see that activation energies in the low-T regime (< 220 K) are of the order of 350 meV, while for $T \geq 220$ K values between 50 and 100 meV are observed. The former trend is typical of electron TAT while the latter of hole TAT [248].

Considering the Arrhenius plot, along with the fact that – as already discussed for panel (c) of Figure 6.3 – energy parameters lower than 90 meV under medium- and high-bias regimes are typical of hole trap-assisted tunneling, we can infer that:

1. at low temperature and low bias electron trap-assisted tunneling governs the transport;
2. at low temperature and medium bias regime hole TAT dominates;
3. at high temperature and medium- to high-bias regime hole TAT, again, dominates;

where the high- and low-T regime is determined by $T = 220$ K while the critical value for high bias is approximately $V_{\text{crit}} = 2.65$ V. The low-bias trend denoted by filled dots in Figure 6.3 is, in the author's opinion, a thermal transition occurring from pure electron TAT at around 100 K to a pure hole TAT for $T = 420$ K, when E_0 assumes values more typical of holes. A further confirmation about these conclusions comes from a simple test performed on particle effective masses. Since E_0 is a function of m^* (see Eq. (6.3)), the ratio $E_{0,\text{e}}/E_{0,\text{h}}$ corresponds to

$$M = \sqrt{\frac{m^*_{\text{h}}}{m^*_{\text{e}}}}\,. \tag{6.8}$$

Introducing typical values of electrons and heavy holes effective masses in III-nitrides, then Eq. (6.8) gives $M_{\text{GaN}} = 2.74$, $M_{\text{AlGaN}} = 2.34$ and $M_{\text{InGaN}} = 2.98$, where both Al and In mole fractions are equal to 0.15. If we compute at $T = 100$ K the ratio between E_0 extracted from the characteristics in the low- and medium-bias regime (see first filled dot and square data points in panel (c) of Figure 6.3) one obtains $M = 2.798$, which is perfectly in accordance to what theoretically just calculated above for GaN, $Al_{0.15}Ga_{0.85}N$ and $In_{0.15}Ga_{0.85}N$.

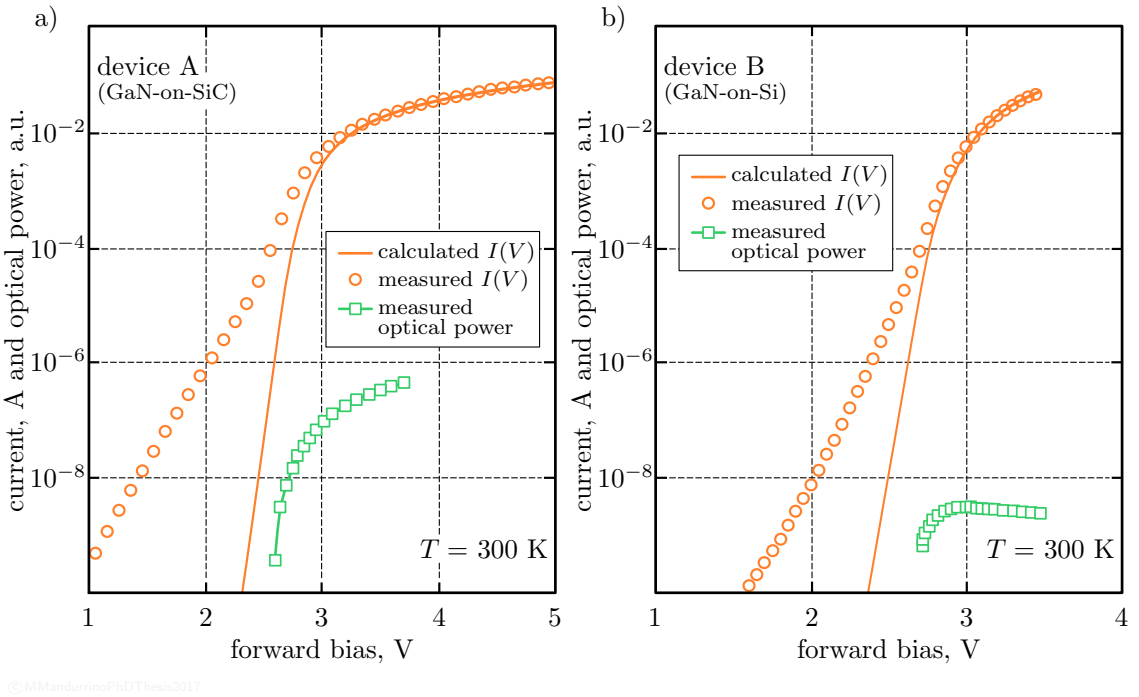

FIGURE 6.5. Experimental optical characteristics (squares), experimental (circles) and simulated (solid line) electrical characteristics for device A (a) and device B (b) at room-temperature. Simulations account for SRH, radiative and Auger generation/recombination terms only.

6.3 Semiclassical Approach for interband TAT

Now we proceed in describing the simulation procedure adopted in order to investigate the presence of electron/hole TAT in devices A and B. As for IR photodetectors, the semiclassical approach passes through the use of a commercial TCAD simulator [192] which implements a build-in drift-diffusion (DD) solver. The multiphonon emission (MPE) and elastic capture rates written in Eqs. (4.128) and (4.129) are used within the SRH-like TAT recombination rate to calculate electron and hole currents due to the tunneling source term. Again, thanks to a Delaunay-Voronoï discretization scheme we divided our LEDs into spatial sub-domains where the quantities related to the carrier transport are computed at run-time. Material parameters, moreover, are those described in Chapter 3 (for a complete description about models and material libraries implementation see also Appendix B).

By using a 2D representation of our LEDs and implementing in the DD system only SRH, radiative and Auger GR source terms we obtained the curves reproduced in plots of Figure 6.5, where measured and calculated electrical characteristics and experimental optical characteristics are represented for both devices at room-temperature. It is well evident that standard generation/recombination models can not explain the excess currents below the optical turn-on, i.e. in the low optical power regime, observed in LED grown either on SiC and on Si. According to previous discussions made about the presence of different tunneling particles we inferred

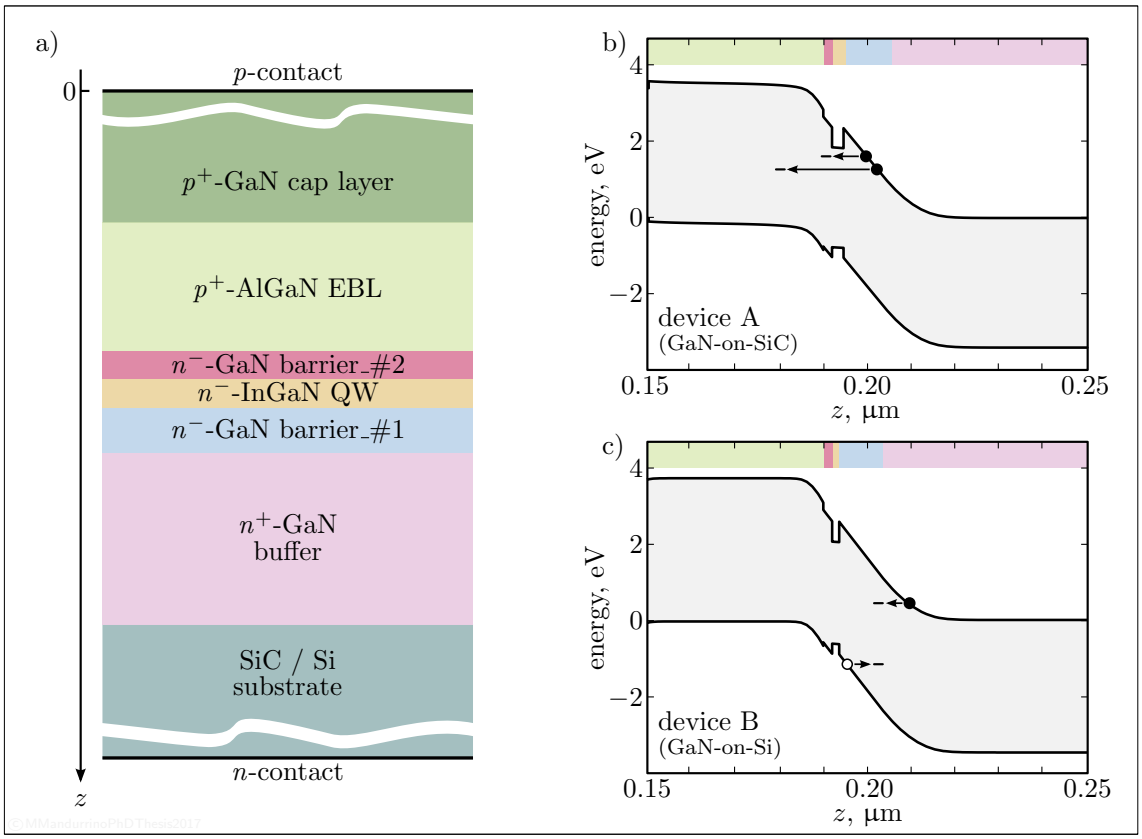

FIGURE 6.6. (a) Simplified structure (not to scale) of our single quantum well (SQW) light-emitting diodes (LEDs) manufactured by OSRAM Opto Semiconductors and their room-temperature band diagrams calculated at equilibrium representing the trap-assisted tunneling (TAT) mechanisms we studied in device A (b) and device B (c).

that the observed excess current, namely the discrepancy between model and experiment, can be in part explained with tunneling. Thus we supposed the occurrence of the following TAT mechanisms in the LED active region (see Figure 6.6): in device A, electron tunneling from the n^--GaN barrier on the n-side towards defects in the second n^--GaN barrier and in the following p^+-AlGaN electron-blocking layer (EBL); in device B, electron tunneling from the n^+-GaN buffer layer into the contiguous n^--GaN barrier and, moreover, hole tunneling from the active region into the same barrier.

These are the TAT mechanisms we will implement in the TCAD simulator in order to verify the initial hypothesis about their occurrence. Due to the high number of physical parameters at play, we shall start the description of numerical results with a multi-parametric preparatory study, which allows to put in evidence the MPE model sensitivity on the variation of its relevant parameters.

6.3.1 Parameters calibration

The first interesting parameter we want to discuss does not involve tunneling at all, so for this reason it is preferable to analyze it before introducing more deeply

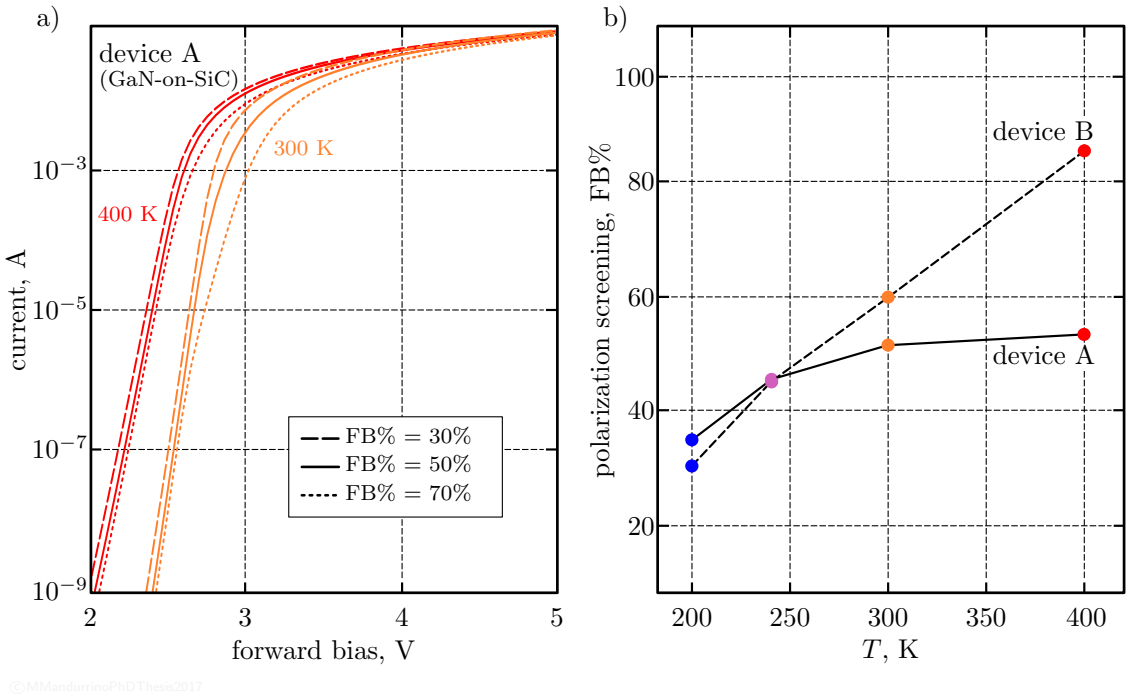

FIGURE 6.7. (a) $I(V)$ characteristics for device A at 300 K and 400 K calculated by tuning the percentage of charge density obtained through Fiorentini-Bernardini model (polarization screening, FB%). (b) Empirical variation of FB% coming from the fitting procedure of calculated characteristics on measured data in the range 200 K to 400 K.

the results of our major interest. The parameter – and, thus, the effects we are talking about – concerns the spontaneous polarization (see the related theory in Subsection 3.2.3) occurring at heterointerfaces of our devices, i.e. between the EBL and the surrounding cap layer, by one side, and the (second) GaN barrier and, even more interesting for us and for lighting applications, at the heterostructure present in correspondence of the InGaN quantum well and the surrounding n^--GaN barriers. As we shall see in a moment, the effects of sheet polarization charges at these interfaces heavily reflect on the electrical behavior of both diodes. Not only, a thermal trend of the polarization charge density seem to occur from the comparison between measured and calculated characteristics (without tunneling).

In order to obtain a satisfactory reproduction of measured $I(V)$ characteristics of device A above the optical turn-on, it has been noticed that the percentage of polarization charge density (*polarization screening*) obtained via Fiorentini-Bernardini model should be tuned in the range ∼35% to ∼55% when the device temperature grows from 200 K to 400 K. This behavior can be probably ascribed to a model underestimation, in our devices, of the lattice constants a_0 and c_0 variation with T. As reported by Figure 6.7 in panel (a), as the parameter FB% increases all the simulated characteristics shift towards higher bias and the threshold voltage increases as well. Panel (b), instead, synthesizes at a glance the empirical increase of FB% versus T as resulting from the tuning performed in order to fit experimental $I(V)$ characteristics without TAT models.

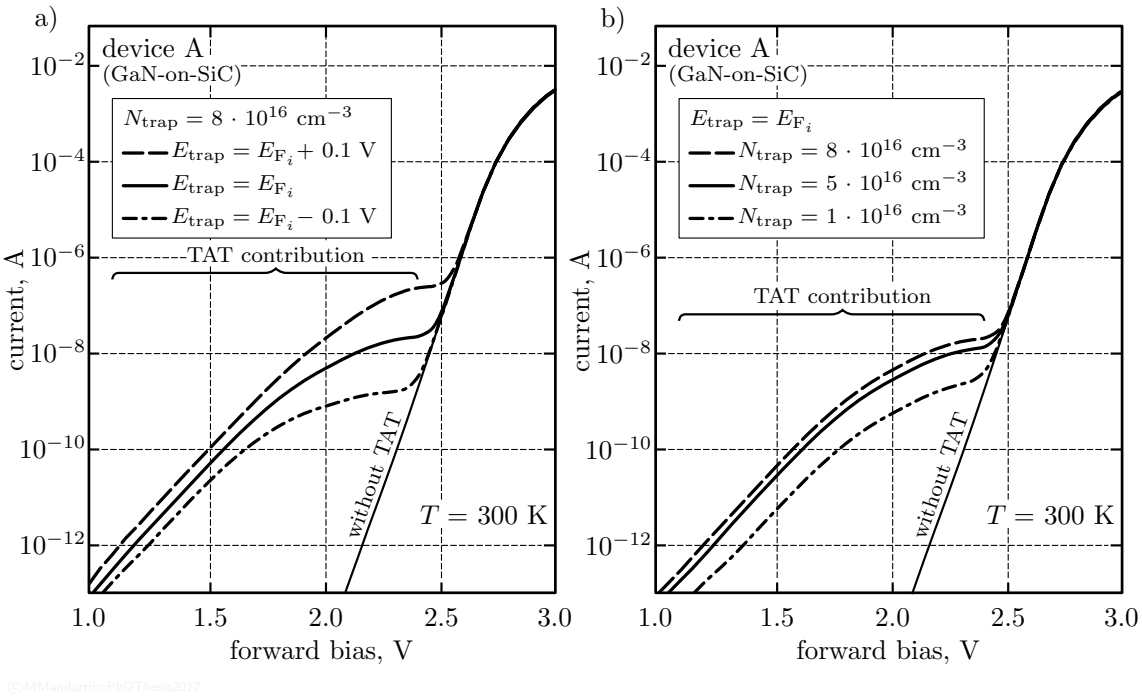

FIGURE 6.8. Simulated electrical characteristics for device A at room-temperature with and without TAT mechanisms. (a) Fixing the trap density at $N_{trap} = 8 \cdot 10^{16}$ cm^{-3}, E_{trap} has been tuned around E_{F_i}. (b) Symmetrically, keeping constant the trap energy E_{trap} at the intrinsic Fermi level, N_{trap} is changed from 1 to $8 \cdot 10^{16}$ cm^{-3}.

In the following we will perform some tests, with electron TAT models activated, only on structure A since the differences between the two LEDs are such that the same qualitative trend of parameters can be applied to both devices. The first set of tests concerns trap-related parameters. Where not otherwise specified, the other TAT parameters are $\Omega_{trap} = 2.24 \cdot 10^{-8}$ μm^2, $m^*_{tunn_n} = 0.2\ m_0$ [105], $m^*_{tunn_p} = 1.5\ m_0$ [105] (for hole TAT in device B), $S = 10$ and $\hbar\omega_0 = 91.2$ meV [95]. Moreover, TAT is implemented by means of a further spatial grid which assists traveling carriers. Such *non-local mesh* was built up and customized by the author dividing the tunneling window – in the case of device A, from the SCR edge in the n^+-GaN to the entire EBL while, for device B, from the SCR edge to the barrier/QW interface – into suitable sub-domains which computationally allow the integration of multiple tunneling paths.

In Figure 6.8 we can appreciate the role of trap energy E_{trap} and density N_{trap} at room-temperature. Panel (a) shows the implications of changing E_{trap} on the calculated electrical characteristics when single-trap levels are implemented with the same arbitrary density $N_{trap} = 8 \cdot 10^{16}$ cm^{-3} both in the barrier and in the EBL. Symmetrically, panel (b) refers to the variations of N_{trap} in both layers affected by the presence of traps when their energies are arbitrarily fixed at the intrinsic Fermi level $E_{trap} = E_{F_i}$.

It is important to note that, as expected, if the trap density increases the tunneling current increases as well. This was pretty expectable. What is less obvious is

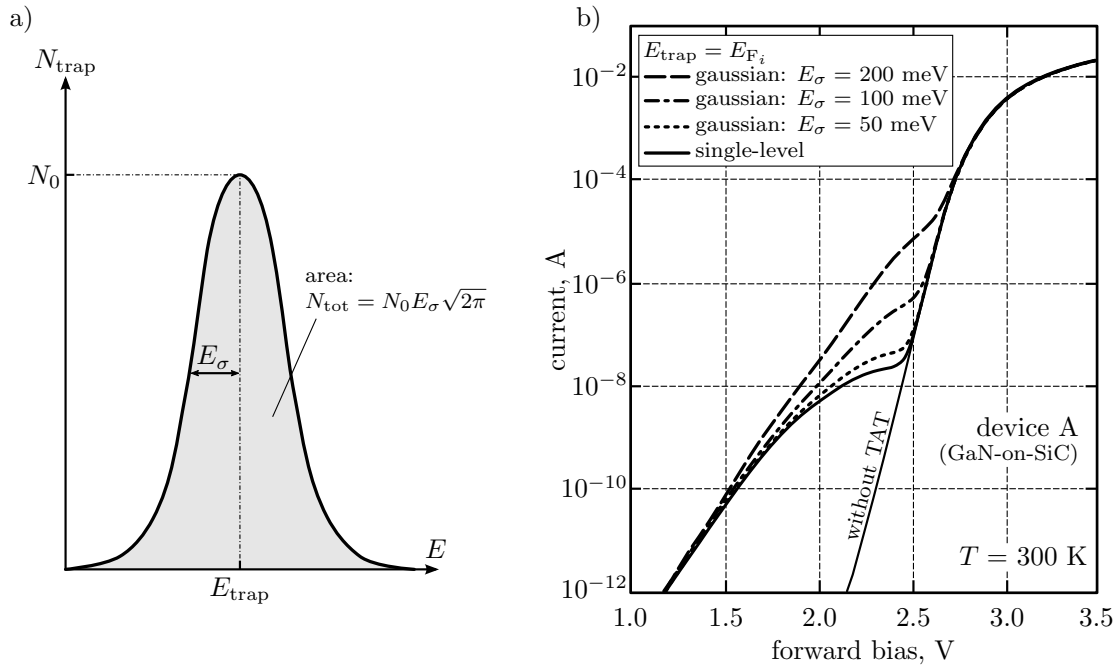

FIGURE 6.9. (a) Representation of trap density of states (DOS) and its related parameters. (b) Room-temperature simulations of device A including the single-level traps case ($N_{\text{trap}} = 8 \cdot 10^{16}$ cm^{-3}) and with gaussian trap DOS where three different configurations of the broadening energy E_σ (with $N_{\text{tot}} = 8 \cdot 10^{16}$ cm^{-3}) have been used. All simulations are performed by imposing $E_{\text{trap}} = E_{F_i}$.

that, unlike the SRH recombination which is enhanced for traps lying on the midgap, here the electron multiphonon-elastic TAT current is maximized when traps are close to the conduction band (CB) edge, i.e. when the tunneling path (for electrons) is minimized. The opposite occurs for holes (here not included). We can conclude that this kind of TAT applied to such LED structures is limited by pure tunneling and not by the ensuing recombination.

The test presented so far are referred to single-level traps, whose energy distribution corresponds to a Dirac-delta function centered at the value E_{trap}. It is worth noting what happens if we implement a certain distribution function for trap DOS as in the case of a Gaussian distribution

$$N_{\text{trap}}(E) = N_0 \exp\left[-\frac{(E - E_{\text{trap}})^2}{2E_\sigma^2}\right], \tag{6.9}$$

where N_0 is the peak concentration and E_σ is the standard deviation of the distribution $N_{\text{trap}}(E)$ (see panel (a) of Figure 6.9). Introducing a trap DOS like in Eq. (6.9) and keeping the total trap spatial density $N_{\text{tot}} = N_0 E_\sigma \sqrt{2\pi}$ fixed at the arbitrary value $N_{\text{tot}} = 8 \cdot 10^{16}$ cm^{-3}, different simulations have been performed by tuning only the parameter E_σ. The results are reported in panel (b) of Figure 6.9. The case with single-level traps, where $N_{\text{trap}} = N_{\text{tot}}$ in order to have homogeneous results, is compared with simulations in which the gaussian broadening assumes the

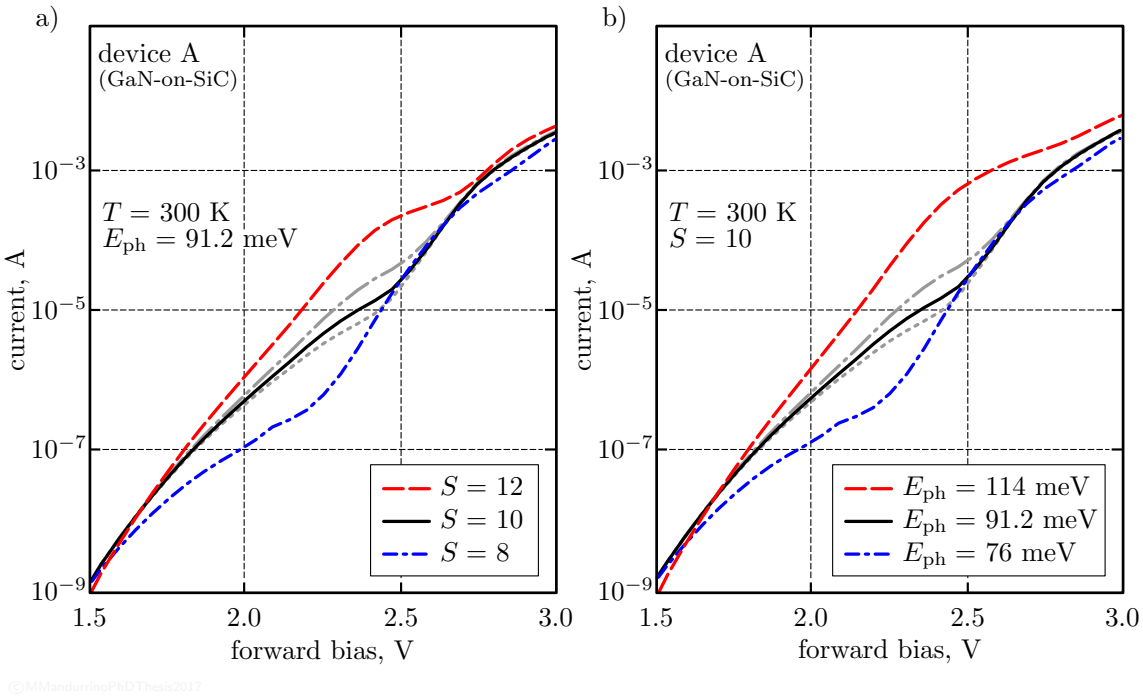

FIGURE 6.10. Room-temperature simulations of TAT conducted on structure A by (a) fixing the phonon energy and varying the Huang-Rhys factor and, viceversa, (b) by fixing S and tuning the energy E_{ph}. Light-grey lines are identical in both panels and represent the case of constant relaxation energy $\epsilon_r = 912$ meV obtained with imposing $S = 8$ and $E_{ph} = 114$ meV (dot-dashed) or $S = 12$ and $E_{ph} = 76$ meV (dotted).

values 50 meV, 100 meV and 200 meV. As expected, if E_σ decreases the simulation tends to the case of single-level traps, since for $E_\sigma \to 0$ the gaussian distribution becomes a Dirac-delta function with $N(E_{trap}) = N_{trap}$.

The last batch of testing simulations was conducted on the most important MPE parameter: the lattice relaxation energy $\epsilon_r = S\hbar\omega_0$. We treated this energy both as collective parameter and also as composed by the product of other two parameters, i.e. the Huang-Rhys factor S and the phonon energy $E_{ph} = \hbar\omega_0$. The simulations resulting from the tuning of these quantities are reported in Figure 6.10 for device A at room-temperature. In panel (a) we kept constant the phonon energy at its typical value for III-nitride materials ($E_{ph} = 91.2$ meV) and we simulated the effects of using different Huang-Rhys factors. In panel (b) we did the inverse so, by keeping S constant ($S = 10$) we explored the implications of varying the energy E_{ph}. Both subfigures include a third (identical) test in which either S and E_{ph} change in such a way that their product remains always equal to $\epsilon_r = 912$ meV: dot-dashed light-gray line indicates the case with $S = 8$ and $E_{ph} = 114$ meV while dotted curve represents the case with $S = 12$ and $E_{ph} = 76$ meV.

The first observation which arises from both plots is that the phonon energy E_{ph} variation has a stronger impact on the TAT contribution to the forward $I(V)$ characteristics of our devices than the Huang-Rhys factor. Since the trend seems to follow the rule according to which, by fixing one parameter – S or E_{ph}, it doesn't

matter – and increasing the other one the TAT contribution increases too, then one would be say that TAT currents increase with the increase of ϵ_r. However, a second interesting feature is that, as represented by light-gray curves, the TAT current can increase even if ϵ_r is constant, provided that E_{ph} sufficiently increases at the expense of S.

All these considerations make the multiphonon-elastic model a rather complicate workbench for TAT simulations, since the high number of parameters and their related degree of freedom. Anyway, both the calibration tests and the analysis conducted on experimental results, as we discussed, can clarify the path towards data interpretation allowing the understanding of electrical measurements which can be as reasonable as possible.

6.3.2 TAT simulation results

Finally we implemented the tunneling mechanisms reported in panels (b) and (c) of Figure 6.6 by using the material libraries discussed in Chapter 3 and Appendix B and with the TAT parameters specified in Tables 6.1 and 6.2. Polarization screening, moreover, varies as reported in panel (b) of Figure 6.7 while, as in part already declared, $\Omega_{\text{trap}_n} = 2.24 \cdot 10^{-8}$ µm^2, $\Omega_{\text{trap}_p} = 10^{-9}$ µm^2, $m^*_{\text{tunn}_n} = 0.2\ m_0$, $m^*_{\text{tunn}_p} = 1.5\ m_0$, $S = 10$ and $\hbar\omega_0 = 91.2$ meV.

TABLE 6.1. Electron TAT parameters in device A

	BARRIER		**EBL**	
TEMPERATURE, T (K)	E_{trap} (eV)	N_{trap} (cm^{-3})	E_{trap} (eV)	N_{trap} (cm^{-3})
200	1.3098	$2 \cdot 10^{17}$	0.8639	$9.9 \cdot 10^{15}$
240	1.3178	$2 \cdot 10^{17}$	0.9027	$9.9 \cdot 10^{15}$
300	1.3182	$2 \cdot 10^{17}$	0.9844	$9.9 \cdot 10^{15}$
400	1.4423	$2 \cdot 10^{17}$	1.1005	$9.9 \cdot 10^{15}$

TABLE 6.2. Electron/hole TAT parameters in device B

	ELECTRON TAT		**HOLE TAT**		
TEMPERATURE, T (K)	E_{trap} (eV)	N_{tot} (cm^{-3})	E_{trap} (eV)	N_{tot} (cm^{-3})	E_σ (eV)
200	1.7015	$3 \cdot 10^{17}$	1.472	$5 \cdot 10^{16}$	0.05
240	1.7081	$3 \cdot 10^{17}$	1.498	$5 \cdot 10^{16}$	0.30
300	1.7173	$3 \cdot 10^{17}$	1.506	$5 \cdot 10^{16}$	0.50
400	1.7253	$3 \cdot 10^{17}$	1.521	$5 \cdot 10^{16}$	0.60

List of single-level/gaussian trap-assisted tunneling (TAT) parameters, where the trap energy is calculated from band edges: below CB for electrons and above VB for holes.

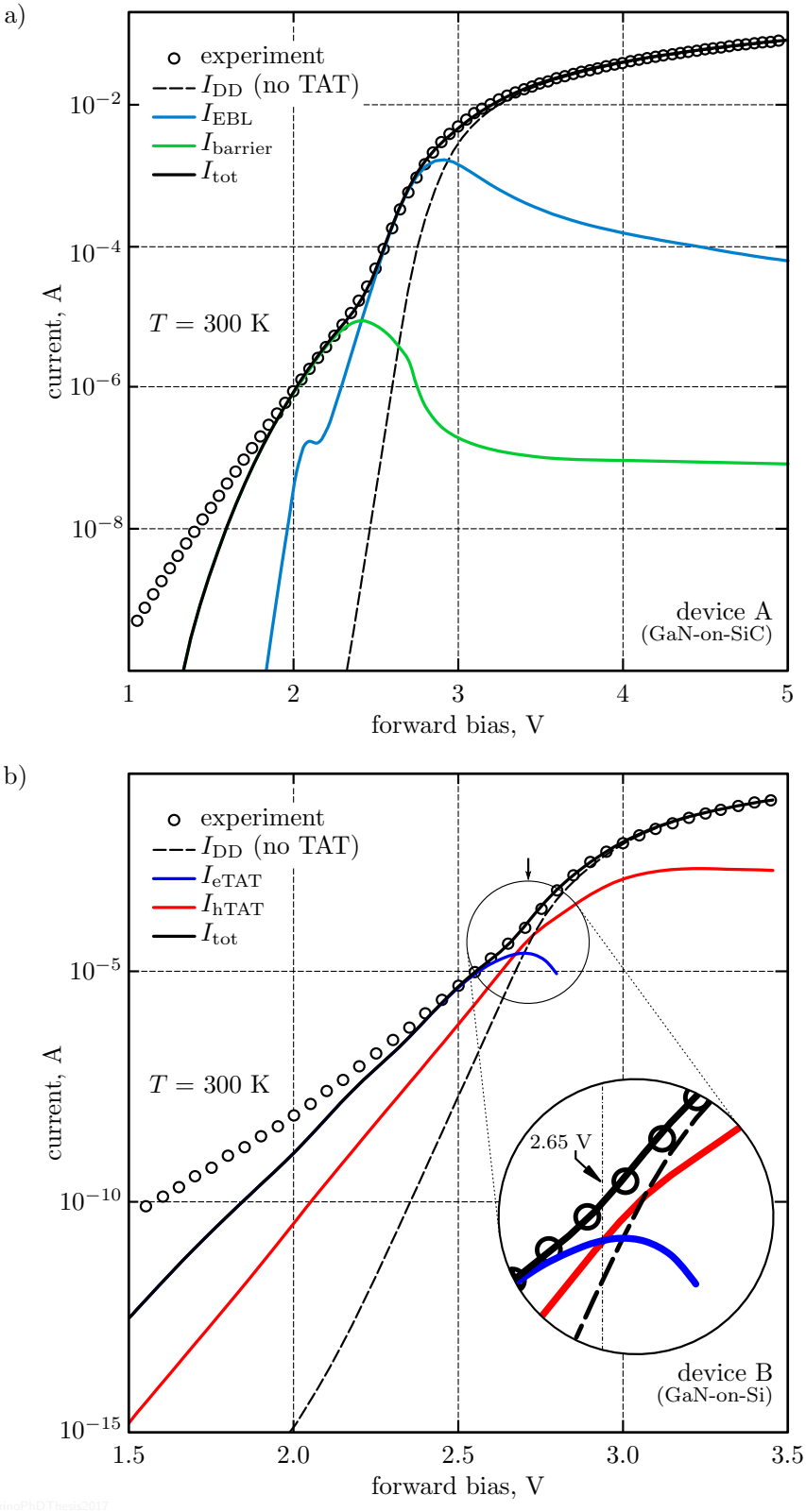

FIGURE 6.11. Room-temperature comparison between calculated and measured $I(V)$ characteristics of (a) device A and (b) device B, including the separated current contributions coming from each TAT process, implemented with the single-level/gaussian parameters of Tables 6.1 and 6.2.

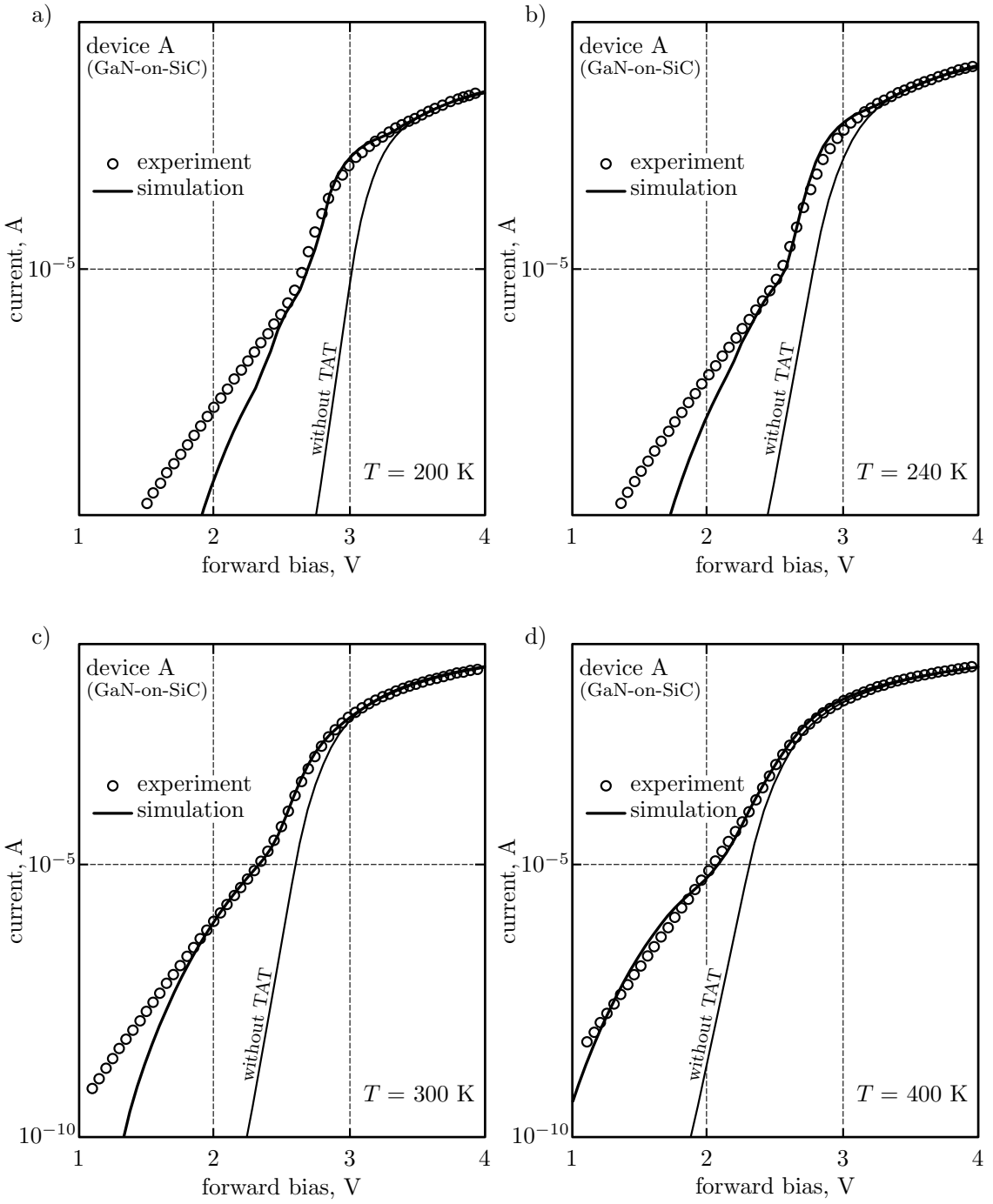

FIGURE 6.12. Measured and calculated $I(V)$ characteristics of device A at different temperatures. Simulations include all the TAT mechanisms discussed in the text, where single-level traps with parameters reported in Table 6.1 have been implemented.

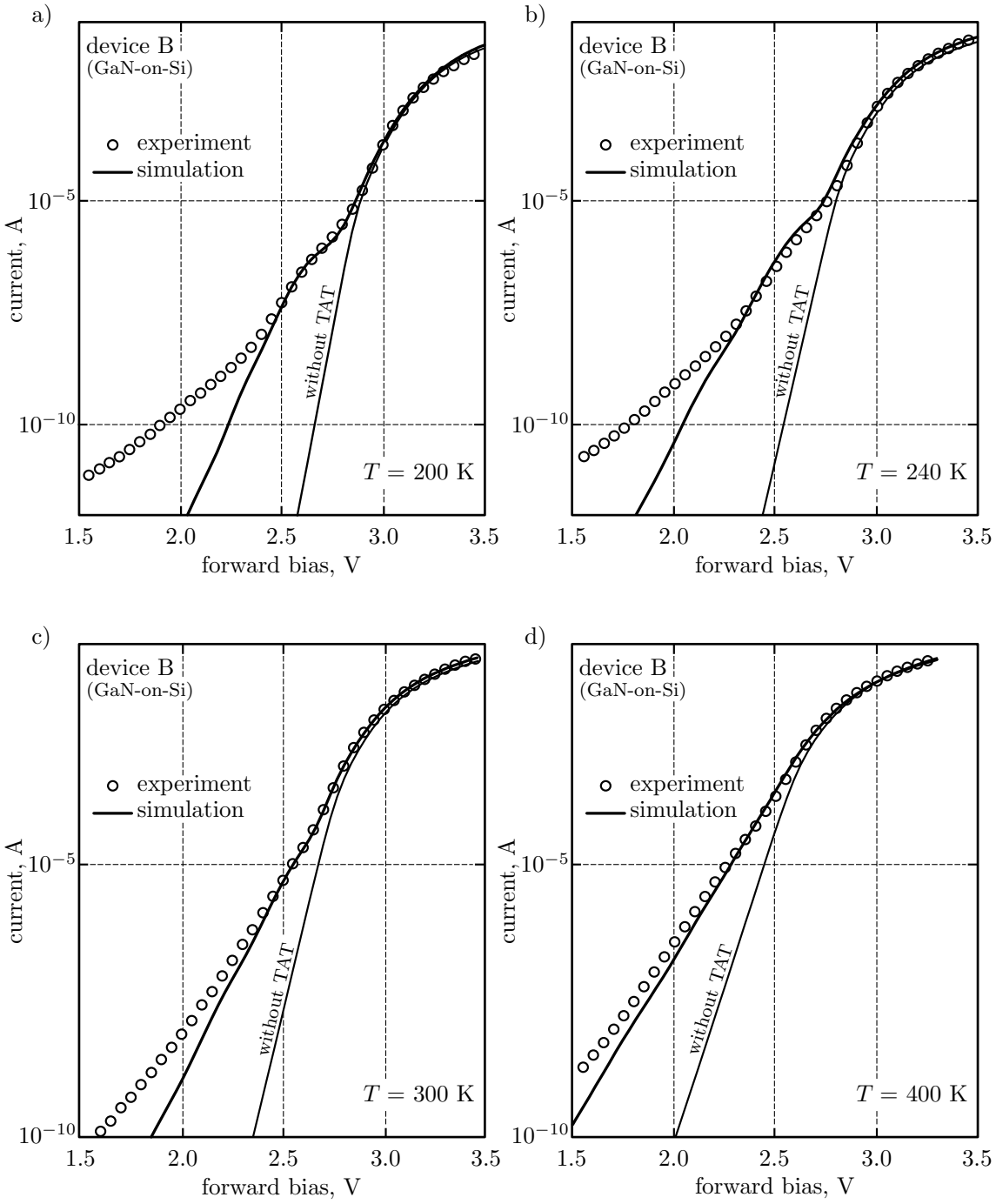

FIGURE 6.13. Measured and calculated $I(V)$ characteristics of device B at different temperatures. Simulations include all the TAT mechanisms discussed in the text, where gaussian traps with parameters reported in Table 6.2 have been implemented.

In Figure 6.11 one can appreciate the comparison between measured and calculated electrical characteristics for both devices. In panel (a) the single contributions due to electron TAT towards (deep) defects in the EBL (light-blue solid line) and in the GaN barrier (green solid) of device A are also included. It is interesting to note that the trap capture/occupation dynamics is pretty evident: at low bias the current coming from each TAT channel increases then, at a certain electric field, it experiences a maximum (here the occupation probability is the highest) and, finally, a partial decrease occurs. In panel (b) we find almost the same behavior. Here, moreover, we included some unpublished results: first of all we differentiated the contribution coming from electron TAT towards the GaN barrier (blue solid) and the hole TAT towards acceptor defects in the same layer (red solid) of device B (as in the previous structure, we hypothesized the presence of deep traps). Then in the zoomed inset the intersection between these two contributions is magnified in order to appreciate the fact that they cross each other exactly at the bias 2.65 V which is, as deduced from data analysis (see Subsection 6.2.1), the threshold determining the onset of hole tunneling dominance over electron TAT, in accordance with Refs. [249, 250]. All the simulations described above include in the drift-diffusion system the SRH, radiative and Auger GR processes. In particular, those concerning single tunneling contributions, have been performed – as in the case of IR photodetectors – considering each time the process under study as the only source term within the continuity equation. For this reason we speak again in terms of contributions rather than components, since the sum of each single contribution not necessarily coincides with the simulation performed activating all the processes.

With the same procedure, and applying the TAT parameters for deep traps listed in Tables 6.1 and 6.2, the characteristics in the range between 100 K and 400 K have been obtained for both devices, as reported in Figures 6.12 and 6.13. Apart for the operative region corresponding to the very-low-bias regime, either device A and B experimental data are well reproduced by simulations accounting for the TAT processes already discussed. A reason that confirms the importance of simulations is that from the parameters implemented it is possible to infer some properties about the simulated structures. In our case, Tables 6.1 and 6.2 and the other set of libraries can allow to extract useful information concerning the traps which are at the origin of TAT mechanisms. First of all, a small decreasing trend of trap energies with the increase of temperature can be observed. This behavior, in agreement with what reported on a theoretical basis by the authors of Ref [183], can be attributed in part to a rearrangement of energy levels due to the gap narrowing with the increase of T (see Subsection 3.2.3). Second, the presence of (deep) defects below and above the active region, up to the p^+-AlGaN EBL, in the LED grown on SiC and only below the QW in the grown-on-Si device could indicate a deeper propagation of defects originating from SiC with respect to Si substrates. At the same time trap densities, as they result from modeling parameters, are higher in device B than in device A.

At any rate this work demonstrated, for the first time in the literature, the efficacy of the multiphonon-elastic TAT picture in modeling GaN-based light-emitters, correctly describing the experimental characteristics of two different SQW LED test

structures below the optical turn-on, and above 2 V, in a current range wide at least five orders of magnitude.

6.4 Quantum Approach for intra-BTBT

The last process we have to analyze is the intra-band-to-band tunneling (intra-BTBT) occurring between the n-barrier close to the quantum well and the QW itself (panel (b) of Figure 6.1). Due to the quantization of energy levels this kind of tunneling requires to be described by a modeling method which self-consistently solves both the Schrödinger problem and also the (quantum) transport equations within the whole device. Even if several approaches have been proposed, as in the case of Density Gradient (DG) formalism, a density-matrix-derived method often implemented in commercial TCAD suites [192] as quantum corrections to the DD (or hydrodynamic) model (see, e.g., Refs. [256–267]), semiclassical modeling can be inadequate to this purpose. Anyway, the full-quantum NEGF approach appears as the most indicate tool to describe such kind of microscopic phenomena.

In order to investigate such process in a realistic case, a new LED structure has been conceived for our simulations. Since NEGF formalism needs to be defined along all the device, by construction, we created a short diode whose structure has been inspired by the one simulated in Ref. [147]. So, starting from the n-side, we have: a 6.5 nm-thick n^+-GaN layer (with $N_D^+ = 10^{19}$ cm^{-3}) then a 3 nm-thick n^--GaN barrier, the active region made of a 2 nm-thick n^--$In_{0.2}Ga_{0.8}N$ quantum well and, again, a second 3 nm-thick n^--GaN barrier, all with doping concentration $N_D^- = 10^{16}$ cm^{-3}. A 4 nm-thick p^+-$Al_{0.13}Ga_{0.87}N$ EBL and a 6.5 nm-thick p^+-GaN cap layer, both doped with $N_A^+ = 10^{19}$ cm^{-3}, conclude the structure (see Figure 6.14).

Input data for the NEGF module has been obtained through the same TCAD tool used in previous simulations. So, first the device is implemented into the DD solver, then conduction and valence band profiles, as well as quasi-Fermi levels, are extracted by using our material libraries (see Appendix B) and by imposing SRH, radiative and Auger GR processes and 100% of polarization screening at heterointerfaces computed via Fiorentini-Bernardini model. Energy profiles and all material parameters are managed by means of HDF5 libraries, working as input files for the NEGF routine. The Green's functions method, finally, is implemented under the scattering regime (see Subsection 4.4.2) i.e. considering besides the boundary self-energies, which define the ballistic picture, also proper carrier-phonon scattering self-energies $\Sigma^{\lessgtr}_{scat}$ such that

$$[\mathbf{G}^{\lessgtr}] = [\mathbf{G}]\big([\Sigma^{\lessgtr}] + [\Sigma^{\lessgtr}_{scat}]\big)[\mathbf{G}]^{\dagger} \tag{6.10}$$

and the the broadening function matrices become

$$[\Gamma_s] = i\big([\Sigma^{>}_{scat}] - [\Sigma^{<}_{scat}]\big)\,, \tag{6.11}$$

where $\Sigma^{\lessgtr}$ are the standard lesser/greater boundary self-energies.

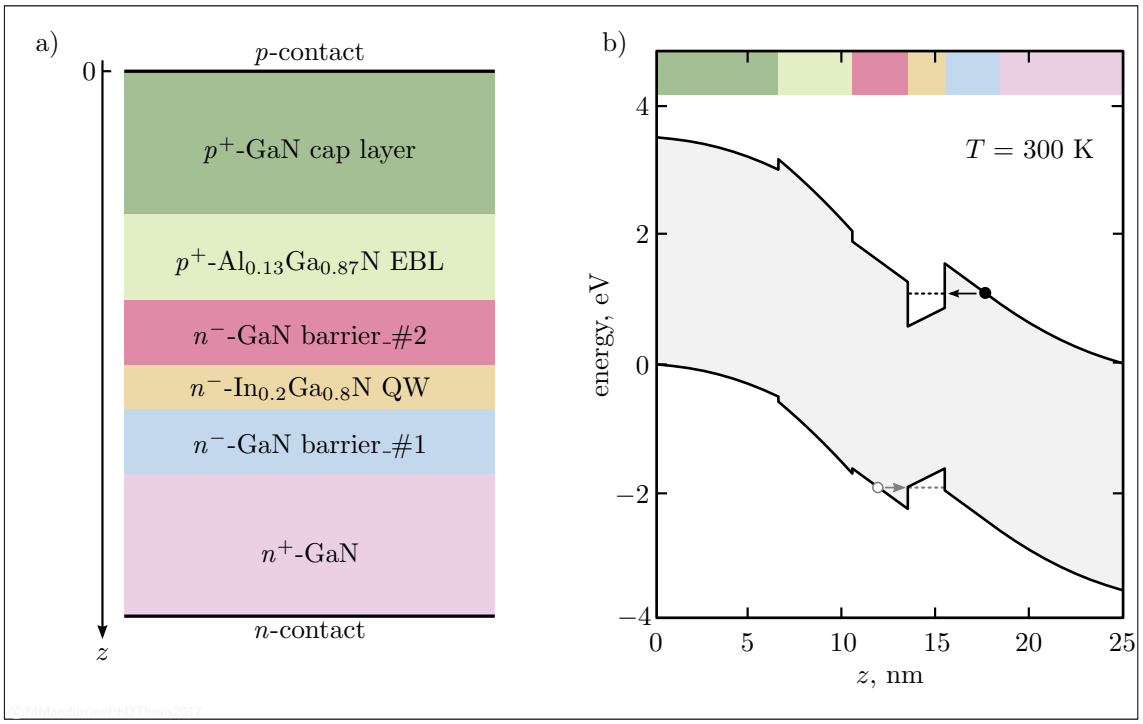

FIGURE 6.14. (a) Simplified structure (all thicknesses are to scale) of our case-study single quantum well (SQW) light-emitting diode (LED) and its room-temperature band diagram calculated at equilibrium including the electron intra-band-to-band tunneling (intra-BTBT) mechanism we studied with NEGF formalism (in grey the symmetrical process involving holes – not discussed here – is also depicted). The whole structure represents the core of NEGF implementation scheme, except for contacts at coordinates $z = 0$ nm and $z = 25$ nm which are, in our formalisms, the leads/boundary regions.

Polar-optical phonon self energies have been introduced as

$$\Sigma^{\lessgtr}_{\text{scat}}(\mathbf{k}, E) = \frac{q^2\hbar\omega}{4\pi^2}\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right)\int_0^{\mathbf{q}_{\text{t}\max}} \mathbf{q}_{\text{t}}\, F(\mathbf{q}_{\text{t}}, \Delta z, \mathbf{k}, \ell_{\text{D}})\, \mathrm{d}\mathbf{q}_{\text{t}} \times \left(f_{\text{BE}}\, \mathbf{G}^{\lessgtr}(\mathbf{q}_{\text{t}}, E \pm \hbar\omega) + (f_{\text{BE}} + 1)\, \mathbf{G}^{\lessgtr}(\mathbf{q}_{\text{t}}, E \mp \hbar\omega)\right) \quad (6.12)$$

with the hypothesis of having only local scattering events and under the first-order Born approximation, and where $\mathbf{q}_{\text{t}}$ and $\mathbf{q}_z$ are the phonon transversal/longitudinal wavevectors, f_{BE} the Bose-Einstein occupation function, $\Delta z = z_i - z_j$ (with z_i and z_j two lattice sites), ℓ_{D} the inverse Debye screening length,

$$F(\mathbf{q}_{\text{t}}, \Delta z, \mathbf{q}_{\text{t}}, \ell_{\text{D}}) = \int_0^{\pi/a} \cos(\mathbf{q}_z\, \Delta z) \left(\frac{1}{\sqrt{\left(\mathbf{q}_z^2 + \mathbf{q}_{\text{t}}^2 + \ell_{\text{D}}^2 + \mathbf{k}^2\right)^2 - 4\mathbf{k}^2\mathbf{q}_{\text{t}}^2}} - \frac{\mathbf{q}_z^2 + \mathbf{q}_{\text{t}}^2 + \ell_{\text{D}}^2 + \mathbf{k}^2}{\left(\left(\mathbf{q}_z^2 + \mathbf{q}_{\text{t}}^2 + \ell_{\text{D}}^2 + \mathbf{k}^2\right)^2 - 4\mathbf{k}^2\mathbf{q}_{\text{t}}^2\right)^{3/2}} \right) \quad (6.13)$$

and a the lattice constant in transport direction.

Analogously to the Boltzmann transport picture, Eq. (6.12) can be represented as a comoposition of a scatter-in and a scatter-out component, respectively

$$\begin{aligned}\Sigma^{<}_{\text{scat}}(\mathbf{k},E) &= \Sigma^{<}_{\text{em}}(\mathbf{k},E) + \Sigma^{<}_{\text{ab}}(\mathbf{k},E) \\ &= D_{\text{el-ph}}\ (f_{\text{BE}}+1)\ \mathbf{G}^{<}(E+\hbar\omega) + D_{\text{el-ph}}\, f_{\text{BE}}\, \mathbf{G}^{<}(E-\hbar\omega) \quad (6.14)\end{aligned}$$

and

$$\begin{aligned}\Sigma^{>}_{\text{scat}}(\mathbf{k},E) &= \Sigma^{>}_{\text{em}}(\mathbf{k},E) + \Sigma^{>}_{\text{ab}}(\mathbf{k},E) \\ &= D_{\text{el-ph}}\ (f_{\text{BE}}+1)\ \mathbf{G}^{>}(E-\hbar\omega) + D_{\text{el-ph}}\, f_{\text{BE}}\, \mathbf{G}^{>}(E+\hbar\omega)\,, \quad (6.15)\end{aligned}$$

which both include an absorption and emission term Σ_{ab} and Σ_{em}, and where $D_{\text{el-ph}}$ represents the electron-phonon interaction strength.

The whole structure we described constitutes the core of NEGF formalism while the ideal contacts located at its extremities represent the left and right "leads", which are folded in the points at $z = 0$ nm and $z = 25$ nm, respectively.

6.4.1 NEGF for direct tunneling towards QBS

Figure 6.14 shows, in panel (a), the schematic structure of our case-study SQW test LED while, in panel (b), the room-temperature band diagram at equilibrium, as it was extracted from DD calculations, is also reported. It includes a schematic representation of the electron (and hole, not implemented) intra-BTBT we want to simulate.

Extracting the electron local density of states (LDOS) for the considered case-study LED structure one obtains the plot represented in panel (a) of Figure 6.15. The first observation concerns the presence of, essentially, two distinct regions within the device where concentration of electronic states are higher (dark patterns): the second GaN barrier well on the left of the well and the well itself. We know that these features typically correspond to opportune quasi-bound states (QBS) levels occurring in the structure. Notice also that, by virtue of NEGF formalism, they are distributed along the energy axis with a certain spreading width, somehow related to the broadening function $\mathbf{g}$. Strictly speaking, we say that LDOS – and also all the other output quantities – are "energy-resolved". The pattern according to which the LDOS distributes in the space/energy range here considered is typical of quantum interference due to the presence of heterointerfaces at the well boundaries, which reflects the coherent superposition of (scattering) states. A second important consideration arising from the same graph deals with the features appearing below the fundamental level in the quantum well: they are due to phonon sub-bands and strictly come from the NEGF scattering picture. The same property can be noticed also in panel (b) where we represented the calculated (energy-resolved) electron density.

Figure 6.16, finally, shows a comparison between QBS energies calculated either with the author's FEM-based 1D Schrödinger solver and via the NEGF code developed by our group. The number and position of states is slightly different but

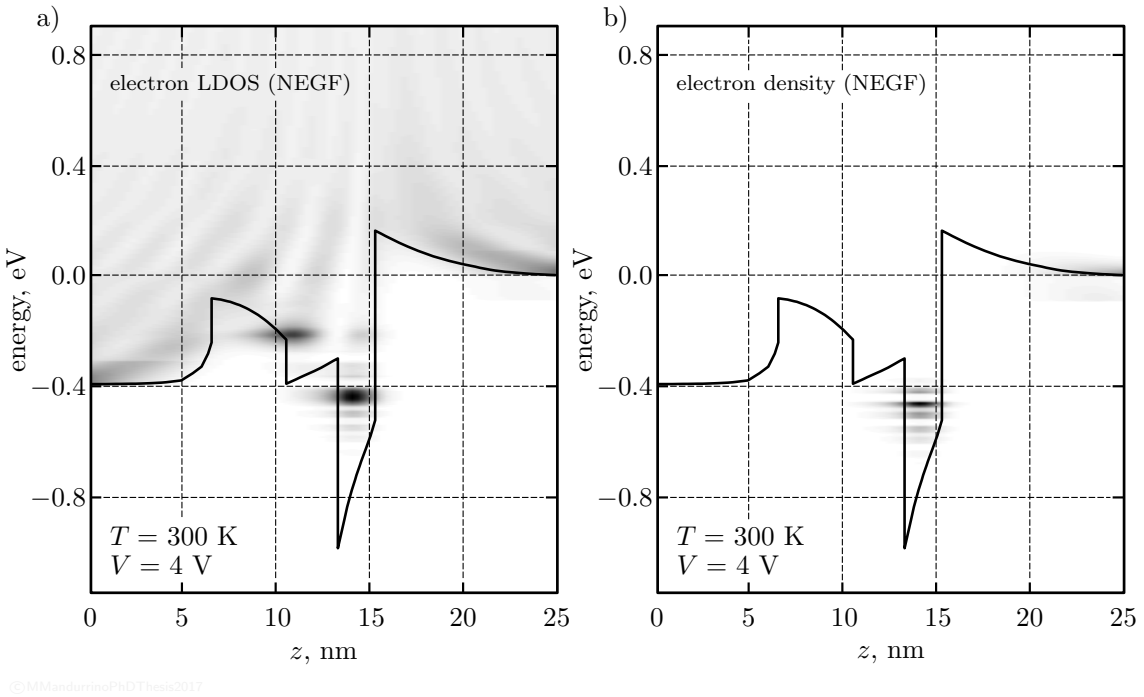

FIGURE 6.15. NEGF (a) electron local density of states (LDOS) and (b) electron density calculated within the scattering picture in our SQW test structure. Both subplots show the presence of quasi-bound states (QBS) in the well, as one may deduce from dark patterns within the active region. The other features in panel (a) can be ascribed to quantum interference at heterointerfaces.

the most important fact, by the tunneling standpoint, is that the highest level is the same in both cases. Moreover, panel (a) also presents the calculated wavefunctions of each QBS while panel (b) includes the NEGF electron current density. It is worth noting that the last eigenvalue detected by both approaches generates a remarkable current channel occurring between the first GaN barrier and the well (the dark area in panel (b)). This fact, along with the non-vanishing trend of wavefunction at the same eigenenergy – as highlighted in panel (a) – is a clear fingerprint of intra-band-to-band tunneling towards QBS we were interested in.

6.5 Final Remarks

In this chapter we have seen the application of different tunneling processes to InGaN/GaN single quantum well blue LEDs simulation. The theory about III-nitrides, and the related material properties exposed in Chapter 3 through TCAD-suitable libraries developed in my first-year of Ph.D., here have found their worth usage. First we chosen two real structures, fabricated by OSRAM Opto Semiconductor (and characterized at Padua University) with slightly different features. Then we inferred from electrical/optical characteristics about the presence of trap-assisted tunneling (TAT) due to electrons and holes travelling from their respective majority-carrier regions into the active layer via material defects and, then, recombining there. Because

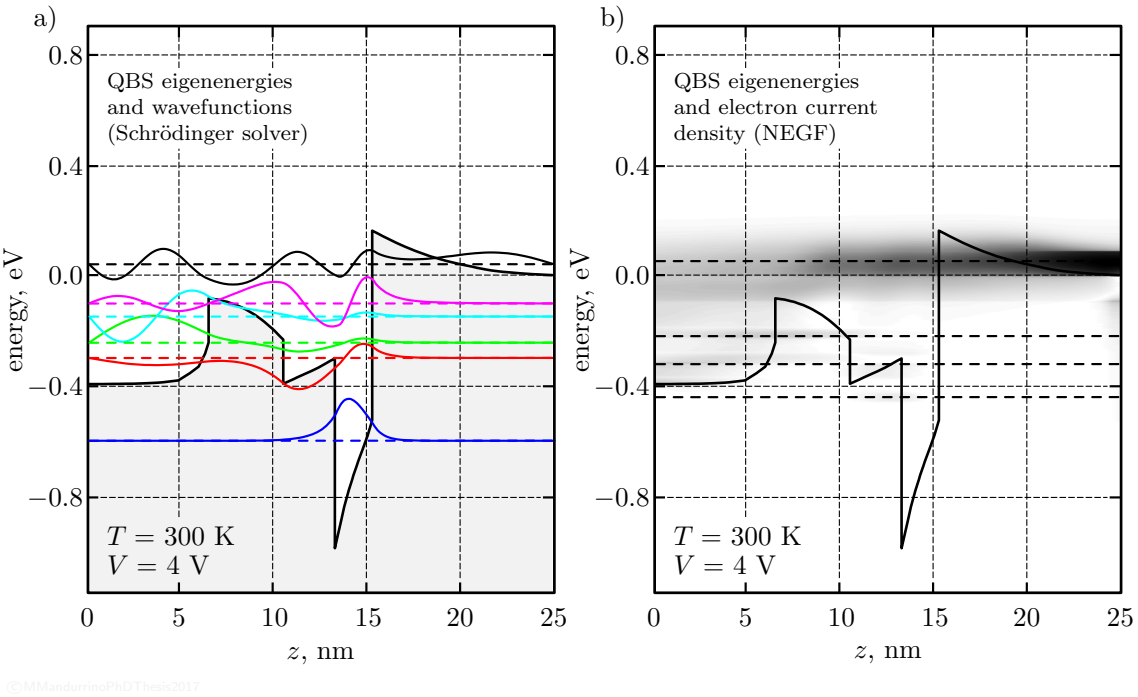

FIGURE 6.16. (a) Eigenenergies and wavefunctions of quasi-bound states (QBS) calculated with the FEM-based 1D Schrödinger solver presented in this work. (b) NEGF QBS eigenenergies and electron current density. Notice the non-vanishing wavefunction at the highest eigenenergy connecting the right-barrier and the well, in panel (a), and the corresponding current channel at the same energy occurring between these layers, in panel (b).

of the band diagram shape, also the occurrence of intra-band-to-band (intra-BTBT) established between the barrier region and quasi-bound states (QBS) in the quantum well has been supposed.

By using the semiclassical framework describing TAT via multiphonon emission (MPE) and elastic carrier/trap scattering event – whose theoretical foundations have been presented in detail in Chapter 4 – we simulated the above mentioned TAT processes. The numerical results demonstrated that the inclusion of such models contributes to accurately reproduce the experimental $I(V,T)$ characteristics in a wide range of currents (at least five orders of magnitude) and temperatures (from 200 K to 400 K) with very good agreement that, otherwise, simple classical DD framework without any tunneling model would not recover. Furthermore, Subsections 6.2.1 and 6.3.2 presented some unpublished results, for instance in what concerns important data analysis intended to discriminate between electron and hole TAT as well as about the implementation of gaussian trap DOS, a feature almost neglected in the literature.

The last section of the present chapter dealt with the second type of tunneling here claimed to occur: intra-BTBT. Since this mechanism involves genuine quantum properties, its physics can be hardly captured by semiclassical formulations. Probably this is why intra-BTBT is rarely investigated in device-level LED modeling studies. As a matter of fact, we made use of the full-quantum NEGF formalism

discussed in Section 6.4 to study these effects in a SQW testing structure. The role of QBS, however, is also well reproduced by the simple author's FEM-based 1D Schrödinger solver. Indeed, comparing results from this approach and from our NEGF code one can come to the same conclusion: a wavefunction/current leakage between the GaN barrier and QBS in the QW occurs. This effect could be important because it may have relevant implications on transport and, thus, also affect the electrical/optical LED behavior.

Chapter 7

Summary and Conclusions

This dissertation, and all the work behind it, pursued the scope of studying different tunneling mechanisms in different optoelectronic devices, by means of different physical formalisms. We started in Part I with a theoretical introduction about basic solid-state physics (Chapter 1) then we moved towards the description of narrow-gap HgCdTe alloy system and HgCdTe-based infrared photodetectors (Chapter 2) and finally we analyzed the family of wide-gap GaN-based materials and their utilize for solid-state lighting, describing also LED properties under the transport and quantum confinement perspective (Chapter 3). This propaedeutical section included some important parts of the author's Ph.D. work, as setting up a self-consistent library of nitride properties suitable for LED numerical simulations, as an example. Moreover, such introduction was actually necessary in order to gradually define the framework and the particular point of view through which tunneling has been considered in the present dissertation.

Part II collected several scientific results about what represented the main core of my research production in the last three years. This chapter is intended to go over the last part again to bring out and highlight the most relevant aspects introduced by the present dissertation. Since I've been involved in a program with a very strong theoretical component, the entire Chapter 4 was fully dedicated to introduce the basic formalism to analyze tunneling under the semiclassical perspective, then the main tunneling mechanisms occurring in semiconductor devices have been explained (Section 4.1) and then, by gradually increasing the degree of complexity, some other theoretical features have been added to the dissertation in order to introduce two main methodological results/improvements:

1. a novel formalism which I developed and that has been applied for the first time to describe direct tunneling in real HgCdTe-based IR photodetectors (Section 4.2);

2. define the physical background concerning the trapping processes in order to model trap-assisted tunneling in GaN-based LEDs (Section 4.3).

The first aspect has been presented in my Master's thesis and here, after being fully developed, it found its first (satisfactory) application in describing the dark current of standard technology (Hg-vacancy p-type doping) HgCdTe IR photodetectors. The second one, instead, can be found as theoretical introduction to the 2015 special issue paper I published in the *Journal of Computational Electronics* by invitation of its then Editor in Chief J. Piprek. The chapter also presented some advanced methods for full-quantum tunneling modeling represented by the density-matrix-based Wigner function method and the Schrödinger-equation-based Non-Equilibrium Green's Function (NEGF) method (Section 4.4), giving also some noticeable examples coming from the literature or showing preliminary results performed by simulating with a NEGF code developed by our Computational Electronics group in Politecnico di Torino.

In Chapter 5 two case-studies of HgCdTe-based IR photodetectors manufactured by AIM Infrarot Module have been presented. They were differing only by the p-type doping dose and technology since device A was doped with the standard Hg-vacancy procedure whereas device B included Au-acceptors implantation. Since the previous literature, and due to narrow-gap materials in in these devices (which in turns produces a short transition path between valence and conduction band), the author inferred the presence of direct band-to-band tunneling (BTBT) affecting the dark electrical behavior under reverse bias of such structures. In order to verify the occurrence of BTBT the above mentioned novel formalism, theoretically presented in my Master's thesis, here has been revised and adapted to our devices. The model was initially conceived for p-i-n diodes being its unique prerequisite a constant profile of the longitudinal component of the electric field along the junction.

Our structures were made by growing a p-type HgCdTe layer followed by a n^--type region and then by a thin n^+-type HgCdTe layer. Being not too much different from the ideal case, we supposed a constant field also in our devices. This fact was confirmed by SIMS profiles (see panel (b) Figure 5.2) thus we felt justified to apply the new model. Actually this re-formulation of the BTBT generation rate, which is a modification of the standard Kane's theory including an empirical parameter D proposed in Refs. [174–176] for Si-based devices and a geometrical/physical device-related parameter δ (in $\mathrm{V}^{1/2}\mathrm{cm}^{-1/2}$) introduced in my Master's thesis [45], demonstrated to contribute to a very accurate matching between measurements and simulations. In particular, a set of published results concerning simulated dark characteristics which included SRH, radiative, Auger recombination, this new BTBT model and also impact ionization described by Eqs. (5.6) and (5.7) provided very satisfactory comparisons between experimental and calculated dark $J(V,T)$ in device A in the range between 112 K and 166 K (see panel (a) of Figure 5.6), where we fixed $D = 1$ and δ was determined via best-fitting procedure.

Besides these results also an unpublished personal author's reinterpretation of such simulations have been presented: by using the Beck's model to describe the

impact ionization (which introduces a more realistic T-dependence) and by using a δ fully adherent to our devices parameters this thesis demonstrates the possibility to obtain an alternative and equally satisfactory version of simulated dark characteristics well reproducing the experimental data which are more rigorous by the physical modeling standpoint (see panel (b) of Figure 5.6). A similar procedure has been followed for device B. Unlikely the previous device here we noticed, probably due to the lower vacancy concentration, that introducing BTBT is useless to recover measured dark currents in the temperature range here considered – again, from 112 K to 166 K – (see panel (a) of Figure 5.7). This is likely the effect of a different field at the junction, by reason of less vacancies in the device, generating a different band bending and, thus, making negligible the presence of direct band-to-band transitions.

Also for device B this thesis proposes a never published alternative – and personal – reinterpretation of numerical simulations which passes through the use of the T-dependent Beck's model. Again, this procedure provided satisfactory agreement between calculated and measured dark currents (see panel (b) of Figure 5.7). In conclusion, it can be claimed that the novel BTBT formulation I developed allowed us to obtain an excellent reproduction of the experimental electrical behavior of our structures in dark conditions. We hope that in the future such model can find a place in commercial simulators among the other built-in functions since we believe it could be of crucial importance in narrow-gap-based IR detectors semiclassical simulations.

Chapter 6 dealt with the analysis of two InGaN/GaN single quantum well (SQW) blue light-emitting diodes (LEDs) manufactured by OSRAM Opto Semiconductors. The first LED, device A hereinafter, differs from the second structure, device B, by the QW composition (respectively with In content equal to 15% and 18%) and thickness (3 nm versus 1.5 nm, respectively) and also by their substrate (SiC and Si for device A and B). We first analyzed the experimental data and, in particular, a subset of electrical and optical characteristics within the range 200 K to 400 K. As reported in literature, by fitting the forward $I(V)$ characteristics of the diodes with a simplified Shockley equation (see Eq. (6.4)) and extracting its parameters, if the values of the ideality factor η are higher than 2 and decreasing with T and, at the same time, the energy parameter E_0 is quite T-independent, then we can infer the occurrence of TAT. This was exactly our case. So, through a complex data analysis entirely reported in Section 6.2 we were able to suppose the presence of electrons tunneling "under the well" of device A from the n-side towards defects in the left barrier and in the p-AlGaN electron-blocking layer (EBL), where they recombine with holes in valence band at the trap site (see panel (b) of Figure 6.6). Moreover this thesis supposed that such process can be modeled via the MPE formalism theoretically derived in Chapter 4.

The results coming from the same analysis performed on device B, however, demonstrated a more complex behavior: in particular the fingerprint of a bi-polar TAT was found. Indeed, two main E_0 and η trends have been observed depending on the bias and temperature regime, since at low-bias and low-T electron TAT seems to dominate while at higher applied voltages hole TAT should govern the electrical

transport. A more deep and never published analysis on this second structure shown that for V_{bias} greater than $V_{\text{crit}} = 2.65$ V (especially at $T > 220$ K) the effective mass of carriers most responsible for defect-assisted transitions corresponds to the heavy hole mass in III-nitride materials. Thus we supposed the existence of electron TAT from the n-side towards defects in the barrier and, viceversa, hole TAT from the active region towards the same layer (see panel (c) of Figure 6.6).

In Section 6.3 we proceeded by simulating the electrical behavior of our devices including, for the first time in the literature for such kind of devices, the MPE model into the tunneling source term of drift-diffusion (DD) equations. Simulations shown that below the optical turn-on of both devices the transport is dominated by defect-assisted phenomena. Moreover device A confirmed to be governed only by electron TAT, as inferred, and trap-related parameters revealed the presence of deep-level (acceptor-like) defects extending from the substrate up to the entire EBL (see panel (a) of Figure 6.11). Device B, instead, confirmed to be affected by both electron and hole tunneling but, unlikely the previous structure, here deep-levels seem to reach only the barrier lying below the active region. Anyway, simulations demonstrated that at $V_{\text{bias}} = 2.65$ V and room-temperature (so, $T > 220$ K) hole contribution starts to dominate over the electron one, in perfect accordance to what theoretically predicted (see panel (b) of Figure 6.11). In short, the most important results can be summarized as follows:

1. the first evidence about the effectiveness of the MPE theory applied to TAT simulations in GaN-based blue LEDs was carried out, both for electron and bi-polar tunneling;

2. the TAT model was demonstrated to be fully reliable for what concerns electrical simulations performed below the optical turn-on of the structures under study in the range between 200 K and 400 K;

3. through our modeling procedure we were able to infer useful information about the energy, the concentration and also the location of acceptor/donor traps generating the observed TAT mechanisms;

4. thanks to our simulations, it was possible to observe and discriminate the contribution coming from each TAT channel for both devices (see Figure 6.11);

5. last but not least, the present study constitutes a quite rare example of combined theoretical, analytical and numerical investigation on GaN-based light-emitters.

We finally hope that such study could boost the research over the transport implications of defects in III-nitride materials in order to improve the industrial processes and reduce the impact of quantum effects in solid-state lighting applications.

Not only TAT mechanisms were supposed to occur in InGaN/GaN LEDs. The typical band structure of such devices suggests also the presence of an interesting

quantum process involving quasi-bound states (QBS) within the QW. Since the triangular barrier surrounding the well on the n-side can assist the intra-band direct tunneling of electrons into the QW (similarly, holes can tunnel from the second barrier on the p-side towards QBS in the well), in Section 6.4 we supposed also the occurrence of such intra-BTBT mechanism (see Figure 6.14). Since in the author's opinion only a genuine quantum formalism could capture such kind of phenomenon it has been decided to employ the NEGF routine to verify whether intra-BTBT has an impact on SQW LED structures. Being the NEGF method highly expensive by the computational standpoint, a new (realistic) structure has been implemented, whose total length is of only 25 nm (see Figure 6.14). The results of NEGF simulations revealed a strong carrier leakage in correspondence of the above mentioned barriers (see Figure 6.15), confirming the initial hypothesis about the relevance of intra-BTBT mechanisms in GaN-based SQW light-emitters. To the same conclusion one should come after performing simulations through the 1D Schrödinger solver developed by the author, which reveals a clear QBS wavefunction leakage in the same region of the device.

To conclude, we presented the most advanced methods and results concerning tunneling modeling and simulations in modern optoelectronic devices, both for light emission and detection. Several physical/mathematical methodologies have been adopted and the processes described range from direct transitions (both in emitters and in photodetectors) to trap-assisted ones, always providing opportune theoretical foundations of data analysis, models derivation and numerical approaches choice. Moreover the results of our calculations shown to be well supported by experimental data and literature information, providing in some cases the first evidence of both models efficacy and phenomena occurrence. Lastly, we tried to give at least a pale feeling of how crucial is the art of physics-based numerical modeling within today electronics and its broad application areas.

Appendices

Appendix A

HgCdTe Parameters Implementation

The MCT material library here summarized accounts for the physical parameters discussed in Chapter 2. Since most of them are x-dependent quantities, the library for HgCdTe material system is built by means of multiple-declaration statements divided into a certain number of subdomains spanning the range $x = 0$ to $x = 1$ according to the scheme

```
*
* parameter(x) = ...
* equation of the x-dependent parameter to be computed
*
*-----------------------------------
* 0 <= x <= x1
*-----------------------------------
  Xmax(0)      = 0
  parameter(0) = < >
  Xmax(1)      = x1
  parameter(1) = < >
*-----------------------------------
* x1 < x <= x2
*-----------------------------------
  Xmax(2)      = x2
  parameter(2) = < >
*-----------------------------------
* x2 < x <= x3
*-----------------------------------
  Xmax(3)      = x3
  parameter(3) = < >
*-----------------------------------
*
* and so on up to x = 1 ...
*
```

remembering that, in our devices, the cadmium mole fraction change at mesh nodes in accordance with the fitting law plotted in panel (b) of Figure 5.2.

In order to manage the T-dependence, instead, a single .par library file is used for each simulation temperature. Such files are automatically generated by a MATLAB source code (thanks to the work done by D. D'Orsogna and M. Moresco from Boston University).

In the following we shall analyze the models related to some of the most important quantities needed to HgCdTe MWIR photodetectors simulation with the commercial suite Sentaurus Device by Synopsys® [192] by showing the code implemented for the fixed temperature $T = 85$ K.

A.1 Energy Gap and Electron Affinity

The Hansen-Seiler [32, 33] formula (see Eqs. (2.17) and (2.18)) for the computation of MCT $E_g(x, T)$, as well as the electron affinity model [43]

$$\chi(x, T) = 4.23 - 0.813\,(E_g(x, T) - 0.083)\ , \tag{A.1}$$

are implemented through the third-order fitting formula

$$y(x) = F + A\,x + B\,x^2 + C\,x^3\,. \tag{A.2}$$

So, the .par file accounts for such bandgap-related code

```
*
* Eg(x)  = F + A x + B x^2 + C x^3
* Chi(x) = F + A x + B x^2 + C x^3
* 0 <= x <= 1
*
*---------------------------------
Bandgap {
* Eg = Eg0 + alpha * Tpar^2 / (beta + Tpar) - alpha * T^2 / (beta + T)
  Xmax(0)     = 0
  Chi0(0)     = 4.5074
  Eg0(0)      = -0.25821
  alpha(0)    = 0
  beta(0)     = 0
  Xmax(1)     = 1
  Chi0(1)     = 2.9916
  Eg0(1)      = 1.6062
  alpha(1)    = 0
  beta(1)     = 0
  B(Chi0(1)) = 0.65853
  C(Chi0(1)) = -0.67642
  B(Eg0(1))  = -0.81
  C(Eg0(1))  = 0.832
}
```

where only one unique mole fraction interval has been considered and in which all `alpha` and `beta` statement are set equal to zero in order to neglect the T-dependence of the Sentaurus built-in formula written at the line 8 of the previous code.

A.2 Permittivity

The (high-frequency) T-independent MCT permittivity $\epsilon(x)$ (see Eq. (2.33)) is implemented, according to Refs. [7, 43], through the model of Eq. (A.2) via the code

```
*
* eps(x) = F + A x + B x^2 + C x^3
* 0 <= x <= 1
*
*-----------------------------------
Epsilon {
  Xmax(0)       =  0
  epsilon(0)    = 20.5
  Xmax(1)       =  1
  epsilon(1)    = 10.7
  B(epsilon(1)) =  5.7
  C(epsilon(1)) =  1.0162e-13
}
```

where, again, the interval is uniquely one.

A.3 Effective Mass

In the case of electron effective masses (see Eq. (2.13)) the x domain has been divided into 10 segments while, again, the third-order fitting formula is used. Since Sentaurus uses equation

$$m_{\mathrm{e}}^{*}(T) = m_m + \sqrt[3]{m_l \left(6a \frac{E_{\mathrm{g}}(0)}{E_{\mathrm{g}}(T)}\right)^2}, \tag{A.3}$$

in order to exploit Eq. (A.2) we have to set as zero all the `a` and `ml` declarations:

```
*
* Formula1: me/m0 = [ (6 * a[Eg(0)/Eg(T)])^2 *  ml ]^(1/3) + mm
* me(x) = F + A x + B x^2 + C x^3
*
*-----------------------------------
eDOSMass {
  Formula = 1
  a  = 0
  ml = 0
*-----------------------------------
* 0 <= x <= 0.1
*-----------------------------------
  Xmax(0)  =  0
  mm(0)    = -0.024312
  Xmax(1)  =  0.1
  mm(1)    = -0.0066767
  B(mm(1)) = -0.54671
```

```
  C(mm(1)) =  1.3154
*-----------------------------------
* 0.1 < x <= 0.2
*-----------------------------------
  Xmax(2)  =  0.2
  mm(2)    =  0.0064493
  B(mm(2)) = -0.30057
  C(mm(2)) =  0.37231
*-----------------------------------
*
* ...
*
}
```

The hole effective masses, instead, are set as $m_{\mathrm{h}}^{*} = 0.55\ m_0$, whatever the temperature and the mole fraction:

```
hDOSMass {
 * Formula1: mh = m0*{[(a+bT+cT^2+dT^3+eT^4)/(1+fT+gT^2+hT^3+iT^4)]^(2/3) + mm}
   Formula = 1
      a = 0
      b = 0
      c = 0
      d = 0
      e = 0
      f = 0
      g = 0
      h = 0
      i = 0
      mm = 0.55  # [m0]
}
```

where all the built-in parameters are imposed to be zero in order to suppress all the unwanted dependencies of the pre-implemented model

$$m_{\mathrm{h}}^{*}(T) = m_m + \left(\frac{a + bT + cT^2 + dT^3 + eT^4}{1 + fT + gT^2 + hT^3 + iT^4}\right)^{2/3} . \tag{A.4}$$

In this way only m_m (i.e. `mm`) survives.

A.4 Mobility

Here the mole fraction range is divided into 51 subdomains while, again, Eq. (A.2) has been used. In order to recover this fitting equation with the models in Eqs. (2.19) and (2.20), the built-in function

$$\mu_{\mathrm{e,h}}(T) = \mu_{\mathrm{Max_{e,h}}} \left(\frac{T}{300\ \mathrm{K}}\right)^{-\alpha_{\mathrm{e,h}}} , \tag{A.5}$$

must be treated by imposing the exponent α as equal to zero. Thus:

```
*
* mu_const = mumax (T/T0)^(-Exponent)
* mu(x) = F + A x + B x^2 + C x^3
*
*----------------------------------------
ConstantMobility {
*----------------------------------------
* 0 <= x <= 0.02
*----------------------------------------
  Xmax(0)     = 0, 0
  mumax(0)    = 92940.4968, 929.405
  Exponent(0) = 0, 0
  Xmax(1)     = 0.02, 0.02
  mumax(1)    = 116158.7834, 1161.5878
  Exponent(1) = 0, 0
  B(mumax(1)) = 5589587.0996, 55895.871
  C(mumax(1)) = 5495919.7093, 54959.1971
*----------------------------------------
*
* ...
*
}
```

where each parameter is declared as a couple of values, respectively, for electrons and holes.

A.5 Doping

The T-independent behavior of dopants (ionization energies and incomplete ionization) follows the carrier statistics

$$\begin{cases} n_{\mathrm{D}} = N_{\mathrm{D}} \dfrac{1}{1 + g_{\mathrm{D}}\, \mathrm{e}^{\left(\frac{n}{n_1}+1\right)}} = N_{\mathrm{D}} \dfrac{1}{1 + g_{\mathrm{D}}\, \mathrm{e}^{\left(\frac{E_{\mathrm{F}}-E_{\mathrm{D}}}{k_{\mathrm{B}}T}\right)}} \\ p_{\mathrm{A}} = N_{\mathrm{A}} \dfrac{1}{1 + g_{\mathrm{A}}\, \mathrm{e}^{\left(\frac{p}{p_1}+1\right)}} = N_{\mathrm{A}} \dfrac{1}{1 + g_{\mathrm{A}}\, \mathrm{e}^{\left(\frac{E_{\mathrm{A}}-E_{\mathrm{F}}}{k_{\mathrm{B}}T}\right)}} \end{cases}, \tag{A.6}$$

where $g_{\mathrm{D}} = 2$ and $g_{\mathrm{A}} = 4$ are the so-called degeneracy factors. Sentaurus uses these formulas only when n_{D} and p_{A} are greater than `NdCrit` and `NaCrit`, respectively. So we considered them as null in order to have Eq. (A.6) activated whatever the doping concentrations. Also the doping-dependence of n_{D} and p_{A} has been neglected by means of `alpha = 0` statement, where `alpha` different from zero would be part of the model via

$$\begin{cases} n_1 = N_{\mathrm{C}}\, \mathrm{e}^{-\frac{E_{\mathrm{C}}-E_{\mathrm{D}}-\alpha_{\mathrm{D}}\sqrt[3]{N_{\mathrm{A}}+N_{\mathrm{D}}}}{k_{\mathrm{B}}T}} \\ p_1 = N_{\mathrm{V}}\, \mathrm{e}^{-\frac{E_{\mathrm{A}}-\alpha_{\mathrm{A}}\sqrt[3]{N_{\mathrm{A}}+N_{\mathrm{D}}}-E_{\mathrm{V}}}{k_{\mathrm{B}}T}} \end{cases}. \tag{A.7}$$

So we set

```
Ionization {
  * Parameters for incomplete ionization (replace '*' with doping species)
  * E_*_0    ionization energy of donor/acceptor
  * g_*      degeneracy factor
  * alpha_*  doping dependent shift parameter
  *-- Donor --
        E_D_0   = -1.89e-3  # [eV]
        g_D     =  2
        alpha_D =  0        # [eV cm]
        NdCrit  =  0        # [cm-3]
  *-- Acceptor --
        E_A_0   =  2.01e-2  # [eV]
        g_A     =  4
        alpha_A =  0        # [eV cm]
        NaCrit  =  0        # [cm-3]
}
```

A.6 GR Mechanisms

In the case of SRH processes the definition of lifetimes passes through the so-called Scharfetter model

$$\tau_{\mathrm{SRH}} = \tau_{\mathrm{min}} + \frac{\tau_{\mathrm{Max}} - \tau_{\mathrm{min}}}{1 + \left(\frac{N_{\mathrm{A}}+N_{\mathrm{D}}}{N_{\mathrm{ref}}}\right)^{\gamma}}\,, \tag{A.8}$$

which assumes an univocal value if we state

$$\tau_{\mathrm{SRH}} = \tau_{\mathrm{min}} = \tau_{\mathrm{Max}}\,, \tag{A.9}$$

where τ_{SRH} is the desired value for electron/hole lifetimes. Our x- and T-independent parameters, thus, become

```
Scharfetter {
  * relation and trap level for SRH recombination:
  * tau = taumin + ( taumax - taumin ) / (1 + ( N/Nref )^gamma)
        taumin = 1.5e-6, 1.5e-6  # [s]
        taumax = 1.5e-6, 1.5e-6  # [s]
        Etrap  = 0               # [eV]
}
```

where the values γ and N_{ref} are negligible since we adopted the hypothesis of Eq. (A.9).

For what concerns, instead, radiative and Auger mechanisms we have again x- and T-dependent parameters (see Section 2.3.2) where, for a fixed temperature, again the third-order interpolation holds. Supposing $T = 85$ K, we have, respectively

```
*
* R_Radiative = C (n p - ni_eff^2)
* Brad(x) = F + A x + B x^2 + C x^3
* 0 <= x <= 1
*
*-----------------------------------
RadiativeRecombination {
        Xmax(0) =  0
        C(0)    =  3.2280e-11
        Xmax(1) =  0.18
        C(1)    =  3.2280e-11
        Xmax(2) =  1
        C(2)    =  6.7069e-10
        B(C(2)) = -5.4721e-10
        C(C(2)) =  4.2358e-10
}
```

with one x-interval, and

```
*
* R_Auger = ( C_n n + C_p p ) (n p - ni_eff^2)
* with C_n,p = (A + B (T/T0) + C (T/T0)^2)
* Caug(x) = F + A x + B x^2 + C x^3
*
*-------------------------------------------------
Auger {
*-------------------------------------------------
* 0 <= x <= 0.18
*-------------------------------------------------
        Xmax(0) = 0, 0
        A(0)    = 1.0375e-24, 2.119e-25  # [cm^6/s]
        B(0)    = 0, 0
        C(0)    = 0, 0
        Xmax(1) = 0.18 #, 0.18
        A(1)    = 1.0375e-24, 2.119e-25  # [cm^6/s]
        B(1)    = 0, 0
        C(1)    = 0, 0
*-------------------------------------------------
*
* ...
*
}
```

which is divided into 52 x-intervals.

Appendix B

GaN/InGaN/AlGaN Parameters Implementation

Here we describe the procedure used to implement some of the most important physical parameters of III-nitride materials listed in Chapter 3 into the commercial TCAD simulator, Sentaurus Device by Synopsys® [192]. Since some built-in functions does not coincide with the equations used in our material library, in the following we present input parameters, models and implementation codes that have been used to introduce our formalisms into the simulator, and in particular within the "parameter file" .par.

B.1 Energy Gap

The room temperature GaN bandgap is computed by placing $T = 300$ K in the following Varshni equation:

$$E_{\mathrm{g,GaN}}(T) = E_{\mathrm{g,GaN}}(0\ \mathrm{K}) - \frac{9.09 \cdot 10^{-4} \cdot T^2}{830 + T}\,, \tag{B.1}$$

while, for $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{N}$ holds

$$E_{\mathrm{g,AlGaN}}(x,T) = (1-x)E_{\mathrm{g,GaN}}(T) + xE_{\mathrm{g,AlN}}(T) - x(1-x)\,, \tag{B.2}$$

where

$$E_{\mathrm{g,AlN}}(T) = E_{\mathrm{g,AlN}}(0\ \mathrm{K}) - \frac{17.99 \cdot 10^{-4} \cdot T^2}{1462 + T}\,, \tag{B.3}$$

in which, again, $T = 300$ K. Finally, for $\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}$:

$$E_{\mathrm{g,InGaN}}(x,T) = (1-x)E_{\mathrm{g,GaN}}(T) + xE_{\mathrm{g,InN}}(T) - 3\,x(1-x)\,, \tag{B.4}$$

where

$$E_{\mathrm{g,InN}}(T) = E_{\mathrm{g,InN}}(0\ \mathrm{K}) - \frac{2.45 \cdot 10^{-4} \cdot T^2}{624 + T}\,. \tag{B.5}$$

For the values of mentioned parameters see Table B.1. It is worth stressing that we do not include any kind of band gap tailoring, so the previous expressions refer to values E_g which are fixed with respect to variables different from mole fraction and temperature. In order to ensure this fact, it is recommended to use the option `EffectiveIntrinsicDensity(NoBandGapNarrowing)` in the "Physics" section of the SDevice command file.

TABLE B.1. III-nitrides bandgap input parameters

PARAMETER	VALUE
GaN energy gap at 0 K, $E_{g,GaN}(0\ K)$	3.507 eV
GaN energy gap at 300 K (calculated), $E_{g,GaN}(300\ K)$	3.435 eV
AlN energy gap at 0 K, $E_{g,AlN}(0\ K)$	6.230 eV
InN energy gap at 0 K, $E_{g,InN}(0\ K)$	0.735 eV
GaN electron affinity, χ_{GaN}	4.070 eV
InGaN and AlGaN conduction band offset, Δ_{off}	0.670 eV

Bandgap-related parameters of GaN, InN and AlN.

As an example, supposing to have mole fractions $x = 0.15$ for both Al and In, the room-temperature energy gaps for ternary alloys are $E_{g,AlGaN}(300\ K) = 3.88$ eV and $E_{g,InGaN}(300\ K) = 4.6$ eV.

B.2 Electron Affinity

GaN affinity is assumed to be temperature-independent since its parameters are defined as for $In_xGa_{1-x}N$ with $x = 0$, which actually suppresses the T-dependence. In fact in

$$\chi_{InGaN}(x,T) = \chi_{GaN} - \left(E_{g,InGaN}(x,T) - E_{g,GaN}(x,T)\right)\ \Delta_{off} \tag{B.6}$$

the term within the parentheses is null as long as $x = 0$, whatever the temperature (see also Eqs. (B.1) and (B.4)). This means that, in the Sentaurus parameter file, the GaN Bandgap-Affinity section must be defined as follows:

```
Bandgap {
* Eg = Eg0 + alpha * Tpar^2 / (beta + Tpar) - alpha * T^2 / (beta + T)
   !(
  set Eg0GaN [expr 3.507 - 9.09e-4 * $T * $T / (830 + $T)]
   )!
  Chi0  = 4.07
  Eg0   = !(puts -nonewline "$Eg0GaN")!
  alpha = 0.0
  beta  = 0.0
```

```
  Tpar  = 0.0
}
```

Firstly we suppress the standard formula (written in the second line of the code above) by plugging `alpha` and `beta` equal to zero; in this way `Chi0` no longer depends on temperature `T`. Then, in the fourth line Eq. (B.1) is implemented. Finally, since all the bandgap-related values we inserted refer to $T = 0$ K, we have also to set `Tpar` as null to be sure to neglect the second addend in the computation of `Eg`. Concerning the alloys AlGaN and InGaN the implementation of Eq. (B.6) and

$$\begin{aligned} \chi_{\mathrm{AlGaN}}(x,T) \;=\; & \chi_{\mathrm{GaN}} - (E_{\mathrm{g,AlGaN}}(x,T) - E_{\mathrm{g,AlGaN}}(300\ \mathrm{K})) \\ & - (E_{\mathrm{g,InGaN}}(x,300\ \mathrm{K}) - E_{\mathrm{g,GaN}}(300\ \mathrm{K}))\ \Delta_{\mathrm{off}}\,, \end{aligned} \tag{B.7}$$

passes, respectively, through the codes

```
Bandgap {
* Eg = Eg0 + alpha * Tpar^2 / (beta + Tpar) - alpha * T^2 / (beta + T)
   !(
  set Eg0GaN [expr 3.507 - 9.09e-4 * $T * $T / (830 + $T)]
  set Eg0InN [expr 0.735 - 2.45e-4 * $T * $T / (624 + $T)]
  set Egx    [expr (1 - $x) * $Eg0GaN + $Eg0InN * $x - 3 * $x * (1 - $x)]
  set ChiGaN [expr 4.07 - ($Egx - $Eg0GaN) * 0.67]
   )!
  Chi0  = !(puts -nonewline "$ChiGaN")!
  Eg0   = !(puts -nonewline "$Egx")!
  alpha = 0.0
  beta  = 0.0
  Tpar  = 0.0
}
```

and

```
Bandgap {
* Eg = Eg0 + alpha * Tpar^2 / (beta + Tpar) - alpha * T^2 / (beta + T)
   !(
  set Eg0GaN    [expr 3.507 -  9.09e-4 *  $T *  $T / ( 830 +  $T)]
  set Eg0GaN300 [expr 3.507 -  9.09e-4 * 300 * 300 / ( 830 + 300)]
  set Eg0AlN    [expr 6.23  - 17.99e-4 *  $T *  $T / (1462 +  $T)]
  set Eg0AlN300 [expr 6.23  - 17.99e-4 * 300 * 300 / (1462 + 300)]
  set Egx       [expr (1 - $x) * $Eg0GaN    + $Eg0AlN    * $x - $x * (1 - $x)]
  set Egx300    [expr (1 - $x) * $Eg0GaN300 + $Eg0AlN300 * $x - $x * (1 - $x)]
  set ChiGaN    [expr 4.07 - ($Egx300 - $Eg0GaN300) * 0.67 - ($Egx - $Egx300)]
   )!
  Chi0  = !(puts -nonewline "$ChiGaN")!
  Eg0   = !(puts -nonewline "$Egx")!
  alpha = 0.0
  beta  = 0.0
  Tpar  = 0.0
}
```

About the meaning of `Chi0` we have to spend few words. According to the Sentaurus Device User Guide [192] the T-dependent electron affinity is computed through

$$\chi(T) = \chi_0 + \frac{(\alpha_1 + \alpha_2)\, T^2}{2\,(\beta_1 + \beta_2 + T)} + k\, E_{\text{bgn}} \tag{B.8}$$

where, by default, $\alpha_2 = \beta_2 = 0$, the bandgap-narrowing energy shift E_{bgn} depends on the correction model activated (if any) and its adjustable coefficient k (the so-called `Bgn2Chi`) assumes the standard value 0.5. Clearly, without any bandgap-narrowing model ($E_{\text{bgn}} = 0$) and by imposing also that $\alpha_1 = \beta_1 = 0$ (as in our case), Eq. (B.8) for GaN reduces to $\chi(T) \equiv \chi_0$. For InGaN and AlGaN, moreover, Eqs. (B.6) and (B.7) are still valid.

B.3 Permittivity

The T-independent InGaN and AlGaN permittivities follow the interpolation rules

$$\epsilon_{\text{InGaN}} = \epsilon_{\text{GaN}} + x\ (\epsilon_{\text{InN}} - \epsilon_{\text{GaN}}) \tag{B.9}$$

and

$$\epsilon_{\text{AlGaN}} = \epsilon_{\text{GaN}} + x\ (\epsilon_{\text{AlN}} - \epsilon_{\text{GaN}})\ , \tag{B.10}$$

where the needed parameters are reported in Table B.2.

TABLE B.2. III-nitrides permittivity input parameters

PARAMETER	VALUE
GaN permittivity, ϵ_{GaN}	9.5 ε_0
InN permittivity, ϵ_{InN}	15.0 ε_0
AlN permittivity, ϵ_{AlN}	8.5 ε_0

Permittivities in GaN, InN and AlN.

The codes used in the parameter file are

```
Epsilon {
  epsilon = 9.5
}
```

for GaN,

```
Epsilon {
   !(
  set perm_GaN    9.5
  set perm_InN   15
  set diff_val   [expr $perm_InN - $perm_GaN]
  set perm_InGaN [expr $diff_val * $x + $perm_GaN]
   )!
```

```
  epsilon = !(puts -nonewline "$perm_InGaN")!
}
```

for InGaN and

```
Epsilon {
   !(
  set perm_GaN   9.5
  set perm_AlN   8.5
  set diff_val   [expr $perm_AlN - $perm_GaN]
  set perm_AlGaN [expr $diff_val * $x + $perm_GaN]
   )!
  epsilon = !(puts -nonewline "$perm_AlGaN")!
}
```

for AlGaN. As an example, assuming $x = 0.15$ for both ternary alloys, we have $\epsilon_{\mathrm{InGaN}} = 10.325\ \epsilon_0$ and $\epsilon_{\mathrm{AlGaN}} = 9.35\ \epsilon_0$.

B.4 Effective Mass

Since the simulator implements the following models for electron and hole effective masses

$$m_{\mathrm{e}}^*(T) = m_m + \sqrt[3]{m_l \left(6a \frac{E_{\mathrm{g}}(0)}{E_{\mathrm{g}}(T)}\right)^2} \tag{B.11}$$

and

$$m_{\mathrm{h}}^*(T) = m_m + \left(\frac{a + bT + cT^2 + dT^3 + eT^4}{1 + fT + gT^2 + hT^3 + iT^4}\right)^{2/3}, \tag{B.12}$$

our fixed values (reported in Table B.3) can be implemented through the codes

```
eDOSMass {
 * Formula1: me/m0 = [ (6 * mt)^2 *  ml ]^(1/3) + mm
 * mt = a[Eg(0)/Eg(T)]
   Formula = 1
      a  = 0
      ml = 0
      mm = < >  # [m0]
}
```

and

```
hDOSMass {
 * Formula1: mh = m0*{[(a+bT+cT^2+dT^3+eT^4)/(1+fT+gT^2+hT^3+iT^4)]^(2/3) + mm}
   Formula = 1
      a = 0
      b = 0
      c = 0
      d = 0
      e = 0
      f = 0
```

```
      g = 0
      h = 0
      i = 0
      mm = < >  # [m0]
}
```

which disable the original formulas and introduce, via the value of the parameter `mm` in brackets, our effective masses written in Table B.3.

TABLE B.3. III-nitrides effective mass input parameters

PARAMETER	VALUE
GaN electron effective mass, $m^*_{\mathrm{e,GaN}}$	0.200 m_0
GaN hole effective mass, $m^*_{\mathrm{h,GaN}}$	1.500 m_0
InGaN electron effective mass, $m^*_{\mathrm{e,InGaN}}$	0.188 m_0
InGaN hole effective mass, $m^*_{\mathrm{h,InGaN}}$	1.670 m_0
AlGaN electron effective mass, $m^*_{\mathrm{e,AlGaN}}$	0.200 m_0
AlGaN hole effective mass, $m^*_{\mathrm{h,AlGaN}}$	1.090 m_0

Effective electron and hole mass in GaN, InGaN and AlGaN.

Hole masses are not differentiated into heavy and light holes because we used the same value for both carriers.

B.5 Mobility

In this case, our model

$$\mu_{\mathrm{e,h}}(T) = \mu_{\mathrm{Max_{e,h}}} \left(\frac{T}{300\ \mathrm{K}} \right)^{-\alpha_{\mathrm{e,h}}}, \tag{B.13}$$

where $\mu_{\mathrm{Max}} = \mu(300\ \mathrm{K})$ is the (constant) mobility at room-temperature, perfectly coincides with the built-in function in Sentaurus. Thus, according to Table B.4 the following code has been implemented in the .par file:

```
ConstantMobility {
 * mu_const = mumax (T/T0)^(-Exponent)
      mumax    = 300, 10   # respectively, electron and hole value [cm2/(Vs)]
      T0       = 300       # [K]
      Exponent = 1.5, 2.0
}
```

TABLE B.4. III-nitrides mobility input parameters

PARAMETER	VALUE
electron room-T mobility, $\mu_{\mathrm{Max_e}}$	300.0 $\mathrm{cm^2V^{-1}s^{-1}}$
hole room-T mobility, $\mu_{\mathrm{Max_h}}$	10.0 $\mathrm{cm^2V^{-1}s^{-1}}$
electron exponential factor, α_{e}	1.5
hole exponential factor, α_{h}	2.0

Mobility-related parameters in GaN, InGaN and AlGaN.

B.6 Doping I: Ionization Energy

As already mentioned in Chapter 3, dopant ionization energies in GaN-based material system are chosen as $E_{\mathrm{D,ion}} = 20$ meV either for donors in GaN, InGaN and AlGaN while acceptors follow the values reported in Table B.5, which are due to the well-known influence of the Al and In mole fractions.

TABLE B.5. III-nitrides incomplete ionization input parameters

PARAMETER	VALUE
donor ionization energy, $E_{\mathrm{D,ion}}$	20 meV
GaN acceptor ionization energy, $E_{\mathrm{A,ion_{GaN}}}$	199 meV
InGaN acceptor ionization energy, $E_{\mathrm{A,ion_{InGaN}}}$	170 meV
AlGaN acceptor ionization energy, $E_{\mathrm{A,ion_{AlGaN}}}$	210 meV

Doping-related parameters in GaN, InGaN and AlGaN.

B.7 Doping II: Incomplete Ionization

We wanted to introduce the dopants statistics

$$\begin{cases} n_{\mathrm{D}} = N_{\mathrm{D}} \dfrac{1}{1 + g_{\mathrm{D}}\, \mathrm{e}^{\left(\frac{E_{\mathrm{F}} - E_{\mathrm{D}}}{k_{\mathrm{B}} T}\right)}} \\ p_{\mathrm{A}} = N_{\mathrm{A}} \dfrac{1}{1 + g_{\mathrm{A}}\, \mathrm{e}^{\left(\frac{E_{\mathrm{A}} - E_{\mathrm{F}}}{k_{\mathrm{B}} T}\right)}} \end{cases}, \tag{B.14}$$

where g_{D} and g_{A} are opportune degeneracy factors. Since the simulator implements the same formulas with the exception that for concentrations higher than the critical values `NdCrit` and `NaCrit`, the ionization becomes complete. To suppress this arbitrary behavior it is sufficient to impose $N_{\mathrm{crit}} = 0$ always. The code introducing both ionization energies parameters and such ionization model is the following

```
Ionization {
  * Parameters for incomplete ionization (replace '*' with doping species)
  * E_*_0    ionization energy of donor/acceptor
  * g_*      degeneracy factor
  * alpha_*  doping dependent shift parameter
  *-- Donor --
        E_D_0   = < >  # [eV]
        g_D     = 2
        alpha_D = 0    # [eV cm]
        NdCrit  = 0    # [cm-3]
  *-- Acceptor --
        E_A_0   = < >  # [eV]
        g_A     = 4
        alpha_A = 0    # [eV cm]
        NaCrit  = 0    # [cm-3]
}
```

where, as usual, brackets must include our material values and where the parameter `alpha` is set to be zero since, otherwise, it would introduce an unwanted doping dependence into the incomplete ionization process.

B.8 GR Mechanisms

In LEDs the SRH recombination has been modeled via its usual trap-related lifetimes

$$\tau_{\mathrm{SRH}} = \frac{1}{\sigma\, v_{\mathrm{th}}\, N_{\mathrm{trap}}}\,. \tag{B.15}$$

In order to call this formula we have to activate, in the SDevice command file, the flag `Recombination(SRH(DopingDependence))`, enabling the doping-dependent Scharfetter model

$$\tau_{\mathrm{SRH}} = \tau_{\mathrm{min}} + \frac{\tau_{\mathrm{Max}} - \tau_{\mathrm{min}}}{1 + \left(\frac{N_{\mathrm{A}}+N_{\mathrm{D}}}{N_{\mathrm{ref}}}\right)^{\gamma}} \tag{B.16}$$

where N_{ref} is an arbitrary reference concentration and γ a dimensionless parameter. In order to make Eqs. (B.15) and (B.16) comparable we have to set $\tau_{\mathrm{Max}} = \tau_{\mathrm{min}}$ and equal to the desired value of SRH lifetime. The following code

```
Scharfetter {
  * relation and trap level for SRH recombination:
  * tau = taumin + ( taumax - taumin ) / (1 + ( N/Nref )^gamma)
        taumin = < >, < >  # [s]
        taumax = < >, < >  # [s]
        Etrap  = 0         # [eV]
}
```

has been used, where brackets should include the electron/hole lifetimes (see Table B.6) and `Etrap = 0` means that we are considering traps at the intrinsic Fermi level.

TABLE B.6. III-nitrides SRH input parameters

PARAMETER	VALUE
GaN SRH lifetime, $\tau_{\mathrm{SRH_{GaN}}}$	$5 \cdot 10^{-8}$ s
InGaN SRH lifetime, $\tau_{\mathrm{SRH_{InGaN}}}$	$5 \cdot 10^{-8}$ s
AlGaN SRH lifetime, $\tau_{\mathrm{SRH_{AlGaN}}}$	$1 \cdot 10^{-7}$ s

Doping-dependent (Scharfetter) SRH-related electron/hole lifetimes in GaN, InGaN and AlGaN.

For what concerns radiative and Auger recombination, we acted by setting directly their coefficients B_{rad} and $C_{n,p}$, respectively. So the .par codes read

```
RadiativeRecombination {
  * R_Radiative = C (n p - ni_eff^2)
        C = < >  # [cm^3/s]
}
```

for radiative, and

```
Auger {
  * R_Auger = ( C_n n + C_p p ) (n p - ni_eff^2)
  * with C_n,p = (A + B (T/T0) + C (T/T0)^2)
        A  = < >, < >  # [cm^6/s]
        B  = 0, 0      # [cm^6/s]
        C  = 0, 0      # [cm^6/s]
}
```

for Auger, where bracketed values follow Table B.7 and where coefficients `B` and `C` have been set equal to zero in order to suppress the second and the third terms in the model written at the line 3 of the previous code.

TABLE B.7. III-nitrides radiative/Auger recombination input parameters

PARAMETER	VALUE
radiative coefficient, B_{rad}	$2 \cdot 10^{-11}$ cm^{3}s^{-1}
Auger coefficient, $C_{n,p}$	$2 \cdot 10^{-30}$ cm^{6}s^{-1}

Radiative and Auger-related recombination coefficients in GaN, InGaN and AlGaN.

Appendix C

Novel BTBT Formulation: C++ Routine

```
#include <iostream>
#include <cmath>
#include <stdlib.h>
#include "PMIModels.h"

class Band2Band_pin_MM : public PMI_Recombination {

protected:
    double pi, q, m0, hbar;

public:
    Band2Band_pin_MM (const PMI_Environment& env);
   ~Band2Band_pin_MM();

void Compute_r
    (const double t, const double n, const double p,
     const double nie, const double f, double& r);

void Compute_drdt
    (const double t, const double n, const double p,
     const double nie, const double f, double& drdt);

void Compute_drdn
    (const double t, const double n, const double p,
     const double nie, const double f, double& drdn);

void Compute_drdp
    (const double t, const double n, const double p,
     const double nie, const double f, double& drdp);

void Compute_drdnie
```

```
    (const double t, const double n, const double p,
     const double nie, const double f, double& drdnie);

void Compute_drdf
    (const double t, const double n, const double p,
     const double nie, const double f, double& drdf);

};

    Band2Band_pin_MM::
    Band2Band_pin_MM (const PMI_Environment& env) :

    PMI_Recombination (env)

    {;}

    Band2Band_pin_MM::
   ~Band2Band_pin_MM ()
 {

 }

void Band2Band_pin_MM::
        Compute_r (const double t, const double n, const double p,
                   const double nie, const double f, double& r)

        { double mt, eg, eg300, A, B, D;
          double pi, q, m0, hbar;

   const PMIBaseParam* gamma = ReadParameter ("gamma");
   double g = *gamma;   // device-related parameter of Mandurrino BTBT rate

   const PMIBaseParam* Temp = ReadParameter ("Temp");
   double T = *Temp;

   double x = ReadxMoleFraction();

      pi = 3.14159265358979;
      q = 1.602e-19;
      m0 = 9.109e-31;
      hbar = 1.054e-34;
      eg300 = 0.283668 * q;
      D = 1;   // dimensionless parameter of Mandurrino BTBT rate

      // x- and T-dependent energy gap (eg) and electron tunneling mass (mt)
      eg1 = -0.302 + 1.93 * x - (0.81 * x * x) + (0.832 * x * x * x);
      eg2 = 5.35e-4 * ((T * T * T -1822) / (T * T - 255.2)) * (1 - 2 * x);
      eg_eV = (eg1+eg2);
      eg = eg_eV * q;
      mt = m0 / (5.73 * (2/eg + 1/(eg+1)));

      A = -(1e-2) * sqrt(2*mt) * pow(q,2) / (4*pow(pi,3)*pow(hbar,2)*sqrt(eg));
      B = (1e-2) * pi * sqrt(mt*pow(eg300,3)) / (2*sqrt(2)*q*hbar);
```

```
      if (f == 0) {
          r = 0.0;
      }
      else {
        r = A * pow(g,(D-1)) * pow(f,((D+1)*0.5))
            * exp( -B / ( g * sqrt(f) ) );
          }
      }

void Band2Band_pin_MM::
    Compute_drdt (const double t, const double n, const double p,
                  const double nie, const double f, double& drdt)
        {;}

void Band2Band_pin_MM::
    Compute_drdn (const double t, const double n, const double p,
                  const double nie, const double f, double& drdn)
        {;}

void Band2Band_pin_MM::
    Compute_drdp (const double t, const double n, const double p,
                  const double nie, const double f, double& drdp)
        {;}

void Band2Band_pin_MM::
    Compute_drdnie (const double t, const double n, const double p,
                    const double nie, const double f, double& drdnie)
        {;}

void Band2Band_pin_MM::
        Compute_drdf (const double t, const double n, const double p,
                      const double nie, const double f, double& drdf)
        { double r, mt, eg, eg1, eg2, eg_eV, eg300, A, B, D;
          double pi, q, m0, hbar;

   const PMIBaseParam* gamma = ReadParameter ("gamma");
   double g = *gamma;   // device-related parameter of Mandurrino BTBT rate

   const PMIBaseParam* Temp = ReadParameter ("Temp");
   double T = *Temp;

      pi = 3.14159265358979;
      q = 1.602e-19;
      m0 = 9.109e-31;
      hbar = 1.054e-34;
      eg300 = 0.283668 * q;
      D = 1;   // dimensionless parameter of Mandurrino BTBT rate

      // x- and T-dependent energy gap (eg) and electron tunneling mass (mt)
      eg1 = -0.302 + 1.93 * x - (0.81 * x * x) + (0.832 * x * x * x);
      eg2 = 5.35e-4 * ((T * T * T -1822) / (T * T - 255.2)) * (1 - 2 * x);
      eg_eV = (eg1+eg2);
      eg = eg_eV * q;
```

```
      mt = m0 / (5.73 * (2/eg + 1/(eg+1)));

      A = -(1e-2) * sqrt(2*mt) * pow(q,2) / (4*pow(pi,3)*pow(hbar,2)*sqrt(eg));
      B = (1e-2) * pi * sqrt(mt*pow(eg300,3)) / (2*sqrt(2)*q*hbar);

    if (f == 0) {
        r = 0.0;
         }
    else {
        r = A * pow((1.24172e2),(d-1)) * pow(f,((d+1)*0.5))
            * exp( -B / ( 1.24172e2 * sqrt(f) ) );
         }

    if (r == 0) {
        drdf = 0.0;
         }
    else {
        drdf = A * pow(g,D) * exp( -B / (g * sqrt(f)) ) * pow(f,((D*0.5)-1))
               * ( (B / (2 * g * g)) + (sqrt(f) * (g * (1 + D))) );
         }
}

PMI_Recombination* new_PMI_Recombination (const PMI_Environment& env)
    { return new Band2Band_pin_MM (env);
}
```

References

[1] R. de L. Kronig and W. G. Penney, "Quantum mechanics of electrons in crystal lattices," *Proc. Royal Soc. London A*, vol. 130, no. 814, pp. 499–513, 1931.

[2] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd ed. New York (NY), Wiley, 2006.

[3] P. T. Landsberg, *Recombination in semiconductors*. Cambridge (UK), Cambridge University Press, 1992.

[4] C. De Santi, M. Meneghini, M. La Grassa, B. Galler, R. Zeisel, M. Goano, S. Dominici, M. Mandurrino, F. Bertazzi, D. Robidas, G. Meneghesso, and E. Zanoni, "Role of defects in the thermal droop of ingan-based light emitting diodes," *J. Appl. Phys.*, vol. 119, no. 9, p. 094501, 2016.

[5] B. Delaunay, "Sur la sphère vide. A la mémoire de Georges Voronoï," *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et naturelles*, no. 6, pp. 793–800, 1934.

[6] A. F. Franz, G. A. Franz, S. Selberherr, C. Ringhofer, and P. Markowich, "Finite boxes - A generalization of the finite-difference method suitable for semiconductor device simulation," *IEEE Trans. Electron Devices*, vol. 30, no. 9, pp. 1070–1082, 1983.

[7] A. Rogalski, *Infrared detectors*, 2nd ed. Boca Raton (FL), CRC Press, 2011.

[8] W. Smith, "Effect of light on selenium during the passage of an electric current," *Nature*, vol. 7, p. 303, 1873.

[9] T. W. Case, "Thalofide cell – a new photo-electric substance," *Phys. Rev.*, vol. 15, no. 4, pp. 289–292, 1920.

[10] C. L. Littler and D. G. Seller, "Temperature dependence of the energy gap of InSb using nonlinear optical techniques," *Appl. Phys. Lett.*, vol. 46, pp. 986–988, 1985.

[11] W. D. Lawson, S. Nielsen, E. H. Putley, and A. S. Young, "Preparation and properties of HgTe and mixed crystals of HgTe-CdTe," *J. Phys. Chem. Solids*, vol. 9, no. 3, pp. 325–329, 1959.

[12] M. A. Kinch, *Fundamentals of infrared detector materials*. Bellingham (WA), SPIE Press, 2007.

[13] W. S. Boyle and G. E. Smith, "Charge-coupled semiconductor devices," *Bell Syst. Tech. J.*, vol. 49, pp. 587–593, 1970.

[14] M. F. Tompsett, G. F. Amelio, W. J. Bertram, R. R. Buckley, W. J. McNamara, J. C. Mikkelsen, and D. A. Sealer, "Charge-coupled imaging devices: experimental results," *IEEE Trans. Electron Devices*, vol. 18, no. 11, pp. 992–996, 1971.

[15] A. D. D. Dwivedi, "Analytical modeling and numerical simulation of P^{+}-$Hg_{0.69}$ $Cd_{0.31}Te$/n-$Hg_{0.78}Cd_{0.22}Te$/CdZnTe heterojunction photodetector for a long-wavelength infrared free space optical communication system," *J. Appl. Phys.*, vol. 110, no. 4, p. 043101, 2011.

[16] A. G. Chynoweth, W. L. Feldmann, and R. A. Logan, "Excess tunnel current in silicon Esaki junctions," *Phys. Rev.*, vol. 121, no. 3, pp. 684–694, 1961.

[17] H. S. Sommers, "Degenerate germanium. II. Band gap and carrier recombination," *Phys. Rev.*, vol. 124, no. 4, pp. 1101–1110, 1961.

[18] J. A. del Alamo and R. M. Swanson, "Forward-bias tunneling: a limitation to bipolar device scaling," *IEEE Electron Device Lett.*, vol. 7, no. 11, pp. 629–631, 1986.

[19] D. K. Blanks, J. D. Beck, M. A. Kinch, and L. Colombo, "Band-to-band tunnel processes in HgCdTe: comparison of experimental and theoretical studies," *J. Vac. Sci. Technol. A*, vol. 6, no. 4, pp. 2790–2794, 1988.

[20] Y. Nemirovsky, D. Rosenfeld, R. Adar, and A. Kornfeld, "Tunneling and dark currents in HgCdTe photodiodes," *J. Vac. Sci. Technol. A*, vol. 7, no. 2, pp. 528–535, 1989.

[21] W. Shockley and W. T. Read, Jr., "Statistics of the recombinations of holes and electrons," *Phys. Rev.*, vol. 87, no. 5, pp. 835–842, 1952.

[22] P. Capper and J. Garland, *Mercury Cadmium Telluride: growth, properties and applications*. Chichester, UK, John Wiley & Sons, 2011.

[23] P. Muralidharan, D. Vasileska, and P. S. Wijewarnasuriya, "Modeling of HgCdTe photodetectors in the LWIR region," in *Computational Electronics (IWCE), 2012 15th International Workshop on*, 2012, pp. 1–4.

[24] A. Rogalski, "Heterostructure infrared photovoltaic detectors," *Infrared Phys. Tech.*, vol. 41, no. 4, pp. 213–238, 2000.

[25] A. M. Itsuno, J. D. Phillips, and S. Velicu, "Predicted performance improvement of Auger-suppressed HgCdTe photodiodes and p-n heterojunction detectors," *IEEE Trans. Electron Devices*, vol. 58, no. 2, pp. 501–507, 2011.

[26] J. Chu and A. Sher, *Physics and properties of narrow gap semiconductors*. New York (NY), Springer, 2008.

[27] M. H. Weiler, "Chapter 3. Magnetooptical properties of $Hg_{1-x}Cd_xTe$ alloys," in *Defects, (HgCd)Se, (HgCd)Te*, ser. Semiconductors and Semimetals, R. K. Willardson and A. C. Beer, Eds. Elsevier, 1981, vol. 16, pp. 119–191.

[28] P. W. Bridgman, "Certain physical properties of single crystals of tungsten, antimony, bismuth, tellurium, cadmium, zinc, and tin," *Proc. Am. Acad. Arts Sci.*, vol. 60, no. 6, pp. 305–383, 1925.

[29] F. Bertazzi, M. Moresco, M. Penna, M. Goano, and E. Bellotti, "Full-band Monte Carlo simulation of HgCdTe APDs," *J. Electron. Mater.*, vol. 39, no. 7, pp. 912–917, 2010.

[30] R. E. Longshore, "Chapter 7. MCT properties, growth methods and characterization," in *Handbook of Infra-red Detection Technologies*, M. Henini and M. Razeghi, Eds. Elsevier Science, 2002, pp. 233–267.

[31] G. T. Hess and T. J. Sanders, "HgCdTe double-layer heterojunction detector device," in *Proc. SPIE 4028, Infrared Detectors and Focal Plane Arrays VI, 353*, 2000, pp. 353–364.

[32] G. L. Hansen, J. L. Schmit, and T. N. Casselman, "Energy gap versus alloy composition and temperature in $Hg_{1-x}Cd_xTe$," *J. Appl. Phys.*, vol. 53, no. 10, pp. 7099–7101, 1982.

[33] D. G. Seiler, J. R. Lowney, C. L. Littler, and M. R. Loloee, "Temperature and composition dependence of the energy gap of $Hg_{1-x}Cd_xTe$ by two-photon magnetoabsorption techniques," *J. Vac. Sci. Technol. A*, vol. 8, no. 2, pp. 1237–1244, 1990.

[34] J. R. Lowney, D. G. Seiler, C. L. Littler, and I. T. Yoon, "Intrinsic carrier concentration of narrow-gap mercury cadmium telluride based on the nonlinear temperature dependence of the band gap," *J. Appl. Phys.*, vol. 71, no. 3, pp. 1253–1258, 1992.

[35] W. M. Higgins, G. N. Pultz, R. G. Roy, R. A. Lancaster, and J. L. Schmit, "Standard relationships in the properties of $Hg_{1-x}Cd_xTe$," *J. Vac. Sci. Technol. A*, vol. 7, no. 2, pp. 271–275, 1989.

[36] F. L. Madarasz, F. Szmulowicz, and J. R. McBath, "Intrinsic carrier concentrations and effective masses in $Hg_{1-x}Cd_xTe$," *J. Appl. Phys.*, vol. 58, no. 1, pp. 361–365, 1985.

[37] F. Urbach, "The long-wavelength edge of photographic sensitivity and of the electronic absorption of solids," *Phys. Rev.*, vol. 92, no. 5, pp. 1324–1324, 1953.

[38] W. Martienssen, "Über die excitonenbanden der alkalihalogenidkristalle," *J. Phys. Chem. Solids*, vol. 2, no. 4, pp. 257–267, 1957.

[39] ——, "The optical absorption edge in ionic crystals," *J. Phys. Chem. Solids*, vol. 8, pp. 294–296, 1959.

[40] D. T. F. Marple, "Optical absorption edge in CdTe: experimental," *Phys. Rev.*, vol. 150, no. 2, pp. 728–734, 1966.

[41] C. A. Hougen, "Model for infrared absorption and transmission of liquid-phase epitaxy $Hg_{1-x}Cd_xTe$," *J. Appl. Phys.*, vol. 66, no. 8, pp. 3763–3766, 1989.

[42] E. Finkman and S. E. Schacham, "The exponential optical absorption band tail of $Hg_{1-x}Cd_xTe$," *J. Appl. Phys.*, vol. 56, no. 10, pp. 2896–2900, 1984.

[43] J. Wenus, J. Rutkowski, and A. Rogalski, "Two-dimensional analysis of double-layer heterojunction HgCdTe photodiodes," *IEEE Trans. Electron Devices*, vol. 48, no. 7, pp. 1326–1332, 2001.

[44] K. Jóźwikowski, M. Kopytko, and A. Rogalski, "Numerical estimations of carrier generation-recombination processes and the photon recycling effect in HgCdTe heterostructure photodiodes," *J. Electron. Mater.*, vol. 41, no. 10, pp. 2766–2774, 2012.

[45] M. Mandurrino, "TCAD models for tunneling processes in narrow-gap semiconductors," Master's thesis, Politecnico di Torino, Dec. 2013.

[46] M. Fukuda, *Optical semiconductor devices*, ser. Wiley Interscience publication in Microwave and Optical Engineering. New York (NY), Wiley, 1999, vol. 46.

[47] A. M. White, "Infrared detectors," U.S. Patent 4,679,063, 1987.

[48] S. Maimon and G. W. Wicks, "nBn detector, an infrared detector with reduced dark current and higher operating temperature," *Appl. Phys. Lett.*, vol. 89, no. 15, p. 151109, 2006.

[49] A. M. Itsuno, J. D. Phillips, and S. Velicu, "Mid-wave infrared HgCdTe nBn photodetector," *Appl. Phys. Lett.*, vol. 100, no. 16, p. 161102, 2012.

[50] V. C. Lopes, A. J. Syllaios, and M. C. Chen, "Minority carrier lifetime in mercury cadmium telluride," *Semiconductor Sci. Tech.*, vol. 8, no. 6S, pp. 824–841, 1993.

[51] W. van Roosbroeck and W. Shockley, "Photon-radiative recombination of electrons and holes in germanium," *Phys. Rev.*, vol. 94, no. 6, pp. 1558–1560, 1954.

[52] S. E. Schacham and E. Finkman, "Recombination mechanisms in p-type HgCdTe: freezeout and background flux effects," *J. Appl. Phys.*, vol. 57, no. 6, pp. 2001–2009, 1985.

[53] G. Masetti and S. Solmi, "Relationship between carrier mobility and electron concentration in silicon heavily doped with phosphorus," *IEE J. Solid-State Electron Devices*, vol. 3, no. 3, pp. 65–68, 1979.

[54] D. J. Roulston, N. D. Arora, and S. G. Chamberlain, "Modeling and measurement of minority-carrier lifetime versus doping in diffused layers of n^+-p silicon diodes," *IEEE Trans. Electron Devices*, vol. 29, no. 2, pp. 284–291, 1982.

[55] A. R. Beattie and A. M. White, "An analytic approximation with a wide range of applicability for electron initiated Auger transitions in narrow gap semiconductors," *J. Appl. Phys.*, vol. 79, no. 2, pp. 802–813, 1996.

[56] J. S. Blakemore, *Semiconductor Statistics*, ser. International Series of Monographs on Semiconductors. New York (NY), Pergamon Press, 1962, vol. 3.

[57] A. R. Beattie and P. T. Landsberg, "Auger effect in semiconductors," *Proc. Royal Soc. London A*, vol. 249, no. 1256, pp. 16–29, 1959.

[58] A. R. Beattie and G. Smith, "Recombination in semiconductors by a light hole Auger transition," *Phys. Status Solidi B*, vol. 19, no. 2, pp. 577–586, 1967.

[59] S. Krishnamurthy and T. N. Casselman, "A detailed calculation of the Auger lifetime in p-type HgCdTe," *J. Electron. Mater.*, vol. 29, no. 6, pp. 828–831, 2000.

[60] M. A. Kinch, F. Aqariden, D. Chandra, P.-K. Liao, H. F. Schaake, and H. D. Shih, "Minority carrier lifetime in p-HgCdTe," *J. Electron. Mater.*, vol. 34, no. 6, pp. 880–884, 2005.

[61] J. Y. Tsao, J. J. Wierer, L. E. S. Rohwer, M. E. Coltrin, M. H. Crawford, J. A. Simmons, P.-C. Hung, H. Saunders, D. S. Sizov, R. Bhat, and C.-E. Zah, "Introduction part B. Ultra-efficient Solid-State Lighting: likely characteristics, economic benefits, technological approaches," in *III-nitride based light emitting diodes and applications*, T.-Y. Seong, J. Han, H. Amano, and H. Morkoc, Eds. Springer, 2013, pp. 11–26.

[62] R. N. Hall, G. E. Fenner, J. D. Kingsley, T. J. Soltys, and R. O. Carlson, "Coherent light emission from GaAs junctions," *Phys. Rev. Lett.*, vol. 9, no. 9, pp. 366–368, 1962.

[63] M. I. Nathan, W. P. Dumke, G. Burns, F. H. Dill, and G. Lasher, "Stimulated emission of radiation from GaAs p-n junctions," *Appl. Phys. Lett.*, vol. 1, no. 3, pp. 62–64, 1962.

[64] T. M. Quist, R. H. Rediker, R. J. Keyes, W. E. Krag, B. Lax, A. L. McWhorter, and H. J. Zeigler, "Semiconductor maser of GaAs," *Appl. Phys. Lett.*, vol. 1, no. 4, pp. 91–92, 1962.

[65] N. Holonyak and S. F. Bevacqua, "Coherent (visible) light emission from $Ga(As_{1-x}P_x)$ junctions," *Appl. Phys. Lett.*, vol. 1, no. 4, pp. 82–83, 1962.

[66] R. A. Logan, H. G. White, and W. Wiegmann, "Efficient green electroluminescence in nitrogen-doped GaP p-n junctions," *Appl. Phys. Lett.*, vol. 13, no. 4, pp. 139–141, 1968.

[67] ——, "Efficient green electroluminescent junctions in GaP," *Solid-State Electron.*, vol. 14, no. 1, pp. 55–70, 1971.

[68] W. O. Groves, A. H. Herzog, and M. G. Craford, "The effect of nitrogen doping on $Ga(As_{1-x}P_x)$ electroluminescent diodes," *Appl. Phys. Lett.*, vol. 19, no. 6, pp. 184–186, 1971.

[69] M. G. Craford, R. W. Shaw, A. H. Herzog, and W. O. Groves, "Radiative recombination mechanisms in GaAsP diodes with and without nitrogen doping," *J. Appl. Phys.*, vol. 43, no. 10, pp. 4075–4083, 1972.

[70] J. W. Allen, M. E. Moncaster, and J. Starkiewicz, "Electroluminescent devices using carrier injection in gallium phosphide," *Solid-State Electron.*, vol. 6, no. 2, pp. 95–102, 1963.

[71] H. G. Grimmeiss and H. Scholz, "Efficiency of recombination radiation in GaP," *Phys. Lett.*, vol. 8, no. 4, pp. 233–235, 1964.

[72] R. A. Logan, H. G. White, and F. A. Trumbore, "p-n junction in compensated solution-grown GaP," *J. Appl. Phys.*, vol. 38, no. 6, pp. 2500–2508, 1967.

[73] ——, "p-n junctions in GaP with external electroluminescence efficiency ∼2% at 25°C," *Appl. Phys. Lett.*, vol. 10, no. 7, pp. 206–208, 1967.

[74] R. Juza and H. Hahn, "Über die nitride der metalle der ersten nebengruppen des periodischen systems. Metallamide und metallnitride. X. Mitteilung," *Anorg. Allg. Chem.*, vol. 244, no. 2, pp. 133–148, 1940.

[75] H. P. Maruska and J. J. Tietjen, "The preparation and properties of vapor-deposited single-crystal-line GaN," *Appl. Phys. Lett.*, vol. 15, no. 10, pp. 327–329, 1969.

[76] J. I. Pankove, E. A. Miller, and J. E. Berkeyheiser, "GaN electroluminescent diodes," in *Electron Devices Meeting, 1971 International*, vol. 17, 1971, pp. 78–78.

[77] H. Amano, M. Kito, K. Hiramatsu, and I. Akasaki, "p-type conduction in Mg-doped GaN treated with low-energy electron beam irradiation (LEEBI)," *Japan. J. Appl. Phys.*, vol. 28, no. 12, pp. L2112–L2114, 1989.

[78] I. Akasaki and N. Sawaki, "Process for growing III-V compound semiconductors on sapphire using a buffer layer," U.S. Patent 4,855,249, 1989.

[79] K. Manabe, H. Kato, I. Akasaki, K. Hiramatsu, and H. Amano, "Substrate for growing gallium nitride compound-semiconductor device and light emitting diode," U.S. Patent 5,122,845, 1992.

[80] S. Nakamura, N. Iwasa, and M. Senoh, "Method of manufacturing p-type compound semiconductor," U.S. Patent 5,306,662, 1994.

[81] S. Nakamura, M. Senoh, and T. Mukai, "p-GaN/n-InGaN/n-GaN double-heterostructure blue-light-emitting diodes," *Japan. J. Appl. Phys.*, vol. 32, no. 1A, pp. L8–L11, 1993.

[82] S. Nakamura, T. Mukai, and M. Senoh, "Candela-class high-brightness InGaN/AlGaN double-heterostructure blue-light-emitting diodes," *Appl. Phys. Lett.*, vol. 64, no. 13, pp. 1687–1689, 1994.

[83] X. Guo, J. Graff, and E. F. Schubert, "Photon recycling semiconductor light emitting diode," in *Electron Devices Meeting, 1999. IEDM '99. Technical Digest. International*, 1999, pp. 600–603.

[84] M. Benamara, Z. Liliental-Weber, J. H. Mazur, W. Swider, J. Washburn, M. Iwaya, I. Akasaki, and H. Amano, "The role of the multi buffer layer technique on the structural quality of GaN," in *MRS Proc. Symposium W: GaN and Related Alloys*, vol. 595, 1999, p. F99W5.8.

[85] G. Verzellesi, L. Morassi, G. Meneghesso, M. Meneghini, E. Zanoni, G. Pozzovivo, S. Lavanga, T. Detzel, O. Haberlen, and G. Curatola, "Influence of buffer carbon doping on pulse and ac behavior of insulated-gate field-plated power AlGaN/GaN HEMTs," *IEEE Electron Device Lett.*, vol. 35, no. 4, pp. 443–445, 2014.

[86] A. Chitnis, A. Kumar, M. Shatalov, V. Adivarahan, A. Lunev, J. W. Yang, G. Simin, M. A. Khan, R. Gaska, and M. Shur, "High-quality p-n junctions with quaternary AlInGaN/InGaN quantum wells," *Appl. Phys. Lett.*, vol. 77, no. 23, pp. 3800–3802, 2000.

[87] M. Mandurrino, G. Verzellesi, M. Goano, M. Vallone, F. Bertazzi, G. Ghione, M. Meneghini, G. Meneghesso, and E. Zanoni, "Physics-based modeling and experimental implications of trap-assisted tunneling in InGaN/GaN light-emitting diodes," *Phys. Status Solidi A*, vol. 212, no. 5, pp. 947–953, 2015.

[88] N. I. Bochkareva, E. A. Zhirnov, A. A. Efremov, Y. T. Rebane, R. I. Gorbunov, and Y. G. Shreter, "Tunnel-recombination currents and electroluminescence efficiency in InGaN/GaN LEDs," *Semicond.*, vol. 39, no. 5, pp. 627–632, 2005.

[89] N. I. Bochkareva, V. V. Voronenkov, R. I. Gorbunov, A. S. Zubrilov, Y. S. Lelikov, P. E. Latyshev, Y. T. Rebane, A. I. Tsyuk, and Y. G. Shreter, "Defect-related tunneling mechanism of efficiency droop in III-nitride light-emitting diodes," *Appl. Phys. Lett.*, vol. 96, no. 13, p. 133502, 2010.

[90] E. Taylor, P. R. Edwards, and R. W. Martin, "Colorimetry and efficiency of white LEDs: spectral width dependence," *Phys. Status Solidi A*, vol. 209, no. 3, pp. 461–464, 2012.

[91] M. R. Krames, O. B. Shchekin, R. Mueller-Mach, G. O. Mueller, L. Zhou, G. Harbers, and M. G. Craford, "Status and future of high-power light-emitting diodes for solid-state lighting," vol. 3, no. 2, pp. 160–175, 2007.

[92] A. Zubrilov, "Indium nitride (InN)," in *Properties of advanced semiconductor materials: GaN, AIN, InN, BN, SiC, SiGe*, M. E. Levinshtein, S. L. Rumyantsev, and M. S. Shur, Eds. New York (NY), John Wiley & Sons, 2001, pp. 49–66.

[93] Y. Goldberg, "Aluminum nitride (AlN)," in *Properties of advanced semiconductor materials: GaN, AIN, InN, BN, SiC, SiGe*, M. E. Levinshtein, S. L. Rumyantsev, and M. S. Shur, Eds. New York (NY), John Wiley & Sons, 2001, pp. 31–47.

[94] W. Gian, M. Skowronski, and G. R. Rohrer, "Structural defects and their relationship to nucleation of GaN thin films," in *MRS Proc. Symposium E: III-Nitride, SiC, and Diamond Materials for Electronic*, vol. 423, 1999, pp. 475–486.

[95] A. Bougrov, "Gallium nitride (GaN)," in *Properties of advanced semiconductor materials: GaN, AIN, InN, BN, SiC, SiGe*, M. E. Levinshtein, S. L. Rumyantsev, and M. S. Shur, Eds. New York (NY), John Wiley & Sons, 2001, pp. 1–30.

[96] A. Trampert, O. Brandt, and K. H. Ploog, "Chapter 7. Crystal structure of group III Nitrides," in *Gallium Nitride (GaN) I*, ser. Semiconductors and Semimetals, J. I. Pankove and T. D. Moustakas, Eds. Elsevier, 1997, vol. 50, pp. 167–192.

[97] Z. Dridi, B. Bouhafs, and P. Ruterana, "First-principles investigation of lattice constants and bowing parameters in wurtzite $Al_xGa_{1-x}N$, $In_xGa_{1-x}N$ and $In_xAl_{1-x}N$ alloys," *Semiconductor Sci. Tech.*, vol. 18, no. 9, p. 850, 2003.

[98] G. L. Bir and G. E. Pikus, *Symmetry and strain-induced effects in semiconductors*, ser. A Halsted Press Book. New York (NY), Wiley, 1974.

[99] M. Suzuki and T. Uenoyama, "General remarks on the band structures of group III nitrides," in *Properties, processing and applications of gallium nitride and related semiconductors*, ser. EMIS datareviews series, J. H. Edgar, S. Strite, I. Akasaki, H. Amano, and C. Wetzel, Eds. INSPEC, 1999, vol. 23, pp. 153–158.

[100] M. Suzuki, T. Uenoyama, and A. Yanase, "First-principles calculations of effective-mass parameters of AlN and GaN," *Phys. Rev. B*, vol. 52, no. 11, pp. 8132–8139, 1995.

[101] H. Morkoç, F. Hamdani, and A. Salvador, "Chapter 8. Electronic and optical properties of III-V nitride based quantum wells and superlattices," in *Gallium Nitride (GaN) I*, ser. Semiconductors and Semimetals, J. I. Pankove and T. D. Moustakas, Eds. Elsevier, 1981, vol. 50, pp. 199–257.

[102] Y. P. Varshni, "Temperature dependence of the energy gap in semiconductors," *Physica*, vol. 34, no. 1, pp. 149–154, 1967.

[103] B. Monemar, "Fundamental energy gap of GaN from photoluminescence excitation spectra," *Phys. Rev. B*, vol. 10, no. 2, pp. 676–681, 1974.

[104] Q. Guo and A. Yoshida, "Microhardness of indium nitride single crystal films," *Japan. J. Appl. Phys.*, vol. 33, no. 1R, pp. 90–91, 1994.

[105] I. Vurgaftman and J. R. Meyer, "Band parameters for nitrogen-containing semiconductors," *J. Appl. Phys.*, vol. 94, no. 6, pp. 3675–3696, 2003.

[106] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, "Band parameters for III-V compound semiconductors and their alloys," *J. Appl. Phys.*, vol. 89, no. 11, pp. 5815–5875, 2001.

[107] D. Fritsch, H. Schmidt, and M. Grundmann, "Band-structure pseudopotential calculation of zinc-blende and wurtzite AlN, GaN, and InN," *Phys. Rev. B*, vol. 67, no. 23, p. 235205, 2003.

[108] A. S. Barker and M. Ilegems, "Infrared lattice vibrations and free-electron dispersion in GaN," *Phys. Rev. B*, vol. 7, no. 2, pp. 743–750, 1973.

[109] 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, and J. Hilsenbeck, "Two-dimensional electron gases induced by spontaneous and piezoelectric polarization charges in N- and Ga-face AlGaN/GaN heterostructures," *J. Appl. Phys.*, vol. 85, no. 6, pp. 3222–3233, 1999.

[110] L. Vegard, "Die konstitution der mischkristalle und die raumfüllung der atome," *Zeitschrift für Physik*, vol. 5, no. 1, pp. 17–26, 1921.

[111] V. Fiorentini, F. Bernardini, and O. Ambacher, "Evidence for nonlinear macroscopic polarization in III-V nitride alloy heterostructures," *Appl. Phys. Lett.*, vol. 80, no. 7, pp. 1204–1206, 2002.

[112] F. Bernardini and V. Fiorentini, "Nonlinear macroscopic polarization in III-V nitride alloys," *Phys. Rev. B*, vol. 64, no. 8, p. 085207, 2001.

[113] J. Vaitkus, W. Cunningham, E. Gaubas, M. Rahman, S. Sakai, K. M. Smith, and T. Wang, "Semi-insulating GaN and its evaluation for α particle detection," *Nucl. Instr. Meth. Phys. Res. A*, vol. 509, no. 1-3, pp. 60–64, 2003.

[114] A. Stocco, S. Gerardin, D. Bisi, S. Dalcanale, F. Rampazzo, M. Meneghini, G. Meneghesso, J. Grünenpütt, B. Lambert, H. Blank, and E. Zanoni, "Proton induced trapping effect on space compatible GaN HEMTs," *Microelectron. Reliability*, vol. 54, no. 9-10, pp. 2213–2216, 2014.

[115] C. De Santi, M. Meneghini, N. Trivellin, S. Gerardin, M. Bagatin, A. Paccagnella, G. Meneghesso, and E. Zanoni, "Recoverable degradation of blue ingan-based light emitting diodes submitted to 3 MeV proton irradiation," *Appl. Phys. Lett.*, vol. 105, no. 21, p. 213506, 2014.

[116] P. Kivisaari, J. Oksanen, and J. Tulkki, "Polarization doping and the efficiency of III-nitride optoelectronic devices," *Appl. Phys. Lett.*, vol. 103, no. 21, p. 211118, 2013.

[117] J. Piprek, "AlGaN polarization doping effects on the efficiency of blue LEDs," in *Proc. SPIE 8262, Gallium Nitride Materials and Devices VII*, 2012, pp. 82 620E–82 620E–11.

[118] Z. Li, "Radiation hardness/tolerance of Si sensors/detectors for nuclear and high energy physics experiments," in *Proc. Semiconductor pixel detectors for particles and X-rays. International Workshop, PIXEL2002*, 2002.

[119] M. Moll, "Development of radiation hard sensors for very high luminosity colliders - CERN - RD50 project," *Nucl. Instr. Meth. Phys. Res. A*, vol. 511, no. 1-2, pp. 97–105, 2003.

[120] M. Calciati, M. Goano, F. Bertazzi, M. Vallone, X. Zhou, G. Ghione, M. Meneghini, G. Meneghesso, E. Zanoni, E. Bellotti, G. Verzellesi, D. Zhu, and C. Humphreys, "Correlating electroluminescence characterization and physics-based models of InGaN/GaN LEDs: pitfalls and open issues," *AIP Adv.*, vol. 4, no. 6, p. 067118, 2014.

[121] M. Ilegems and H. C. Montgomery, "Electrical properties of n-type vapor-grown gallium nitride," *J. Phys. Chem. Solids*, vol. 34, no. 5, pp. 885–895, 1973.

[122] J. Piprek, *Optoelectronic devices: Advanced simulation and analysis*. New York (NY), Springer, 2005.

[123] G. Verzellesi, D. Saguatti, M. Meneghini, F. Bertazzi, M. Goano, G. Meneghesso, and E. Zanoni, "Efficiency droop in InGaN/GaN blue light-emitting diodes: physical mechanisms and remedies," *J. Appl. Phys.*, vol. 114, no. 7, p. 071101, 2013.

[124] D. Schiavon, M. Binder, M. Peter, B. Galler, P. Drechsel, and F. Scholz, "Wavelength-dependent determination of the recombination rate coefficients in single-quantum-well GaInN/GaN light emitting diodes," *Phys. Status Solidi B*, vol. 250, no. 2, pp. 1521–3951, 2013.

[125] C. G. Van de Walle, J. Neugebauer, and C. Stampfl, "Native defects, impurities and doping in gan and related compounds: general remarks," in *Properties, processing and applications of gallium nitride and related semiconductors*, ser. EMIS datareviews series, J. H. Edgar, S. Strite, I. Akasaki, H. Amano, and C. Wetzel, Eds. INSPEC, 1999, vol. 23, pp. 274–293.

[126] J. Neugebauer and C. G. Van de Walle, "Defects and doping in GaN," in *Proc. 22nd International Conference on the Physics of Semiconductors*, 1995, p. 2327.

[127] S. Fischer, C. Wetzel, E. E. Haller, and B. K. Meyer, "On p-type doping in GaN-acceptor binding energies," *Appl. Phys. Lett.*, vol. 67, no. 9, pp. 1928–1300, 1995.

[128] C. Stampfl and C. G. Van de Walle, "Doping of $Al_xGa_{1-x}N$," *Appl. Phys. Lett.*, vol. 72, no. 4, pp. 459–461, 1998.

[129] J. Li, T. N. Oder, M. L. Nakarmi, J. Y. Lin, and H. X. Jiang, "Optical and electrical properties of Mg-doped p-type $Al_xGa_{1-x}N$," *Appl. Phys. Lett.*, vol. 80, no. 7, pp. 1210–1212, 2002.

[130] K. B. Nam, M. L. Nakarmi, J. Li, J. Y. Lin, and H. X. Jiang, "Mg acceptor level in AlN probed by deep ultraviolet photoluminescence," *Appl. Phys. Lett.*, vol. 83, no. 5, pp. 878–880, 2003.

[131] M. L. Nakarmi, N. Nepal, J. Y. Lin, and H. X. Jiang, "Photoluminescence studies of impurity transitions in Mg-doped AlGaN alloys," *Appl. Phys. Lett.*, vol. 94, no. 9, p. 091903, 2009.

[132] K. Sakowski, L. Marcinkowski, S. Krukowski, S. Grzanka, and E. Litwin-Staszewska, "Simulation of trap-assisted tunneling effect on characteristics of gallium nitride diodes," *J. Appl. Phys.*, vol. 111, no. 12, p. 123115, 2012.

[133] V. V. Mitin, D. I. Sementsov, and N. Z. Vagidov, *Quantum mechanics for nanostructures*. Cambridge (UK), Cambridge University Press, 2010.

[134] M. Peter, A. Laubsch, W. Bergbauer, T. Meyer, M. Sabathil, J. Baur, and B. Hahn, "New developments in green leds," *Phys. Status Solidi A*, vol. 206, no. 6, pp. 1125–1129, 2009.

[135] R. Charash, P. P. Maaskant, L. Lewis, C. McAleese, M. J. Kappers, C. J. Humphreys, and B. Corbett, "Carrier distribution in InGaN/GaN tricolor multiple quantum well light emitting diodes," *Appl. Phys. Lett.*, vol. 95, no. 15, p. 151103, 2009.

[136] C.-Y. Huang, Q. Yan, Y. Zhao, K. Fujito, D. Feezell, C. G. Van de Walle, J. S. Speck, S. P. DenBaars, and S. Nakamura, "Influence of Mg-doped barriers on semipolar $(20\bar{2}1)$ multiple-quantum-well green light-emitting diodes," *Appl. Phys. Lett.*, vol. 99, no. 14, p. 141114, 2011.

[137] B. Galler, A. Laubsch, A. Wojcik, H. Lugauer, A. Gomez-Iglesias, M. Sabathil, and B. Hahn, "Investigation of the carrier distribution in InGaN-based multi-quantum-well structures," *Phys. Status Solidi C*, vol. 8, no. 7-8, pp. 2372–2374, 2011.

[138] N. K. Dutta and R. J. Nelson, "The case for Auger recombination in $In_{1-x}Ga_xAs_yP_{1-y}$," *J. Appl. Phys.*, vol. 53, no. 1, pp. 74–92, 1982.

[139] A. S. Polkovnikov and G. G. Zegrya, "Auger recombination in semiconductor quantum wells," *Phys. Rev. B*, vol. 58, no. 7, pp. 4039–4056, 1998.

[140] E. Kioupakis, P. Rinke, K. T. Delaney, and C. G. Van de Walle, "Indirect Auger recombination as a cause of efficiency droop in nitride light-emitting diodes," *Appl. Phys. Lett.*, vol. 98, no. 16, p. 161107, 2011.

[141] J. Piprek, "Efficiency droop in nitride-based light-emitting diodes," *Phys. Status Solidi A*, vol. 207, no. 10, pp. 2217–2225, 2010.

[142] W. W. Chow, "Modeling excitation-dependent bandstructure effects on InGaN light-emitting diode efficiency," *Opt. Express*, vol. 19, no. 22, pp. 21 818–21 831, 2011.

[143] D. Saguatti, L. Bidinelli, G. Verzellesi, M. Meneghini, G. Meneghesso, E. Zanoni, R. Butendeich, and B. Hahn, "Investigation of efficiency-droop mechanisms in multi-quantum-well InGaN/GaN blue light-emitting diodes," *IEEE Trans. Electron Devices*, vol. 59, no. 5, pp. 1402–1409, 2012.

[144] J. Piprek, "Introduction to the Special Issue on "Simulation of GaN-based Light-Emitting Diodes"," *J. Comp. Electron.*, vol. 14, no. 2, p. 381, 2015.

[145] M. Mandurrino, M. Goano, M. Vallone, F. Bertazzi, G. Ghione, G. Verzellesi, M. Meneghini, G. Meneghesso, and E. Zanoni, "Semiclassical simulation of trap-assisted tunneling in GaN-based light-emitting diodes," *J. Comp. Electron.*, vol. 14, no. 2, pp. 444–455, 2015.

[146] M. Auf der Maur, "Multiscale approaches for the simulation of InGaN/GaN LEDs," *J. Comp. Electron.*, vol. 14, no. 2, pp. 398–408, 2015.

[147] A. Shedbalkar, Z. Andreev, and B. Witzigmann, "Simulation of an indium gallium nitride quantum well light-emitting diode with the non-equilibrium Green's function method," *Phys. Status Solidi B*, vol. 253, no. 1, pp. 158–163, 2016.

[148] J. Geng, P. Sarangapani, E. Nelson, C. Wordelman, B. Browne, T. Kubis, and G. Klimeck, "Multi-scale, multi-physics NEGF quantum transport for nitride LEDs," in *2016 International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD)*, 2016, pp. 107–108.

[149] G. Wentzel, "Eine verallgemeinerung der quantenbedingungen für die zwecke der wellenmechanik," *Zeitschrift für Physik*, vol. 38, no. 6, pp. 518–529, 1926.

[150] H. A. Kramers, "Wellenmechanik und halbzahlige quantisierung," *Zeitschrift für Physik*, vol. 39, no. 10, pp. 828–840, 1926.

[151] L. N. Brillouin, "Remarques sur la mécanique ondulatoire," *J. Phys. Radium*, vol. 7, no. 12, pp. 353–368, 1926.

[152] S. M. Sze, *Physics of semiconductor devices*, 2nd ed. New York (NY), Wiley, 1981.

[153] J. L. Moll, *Physics of semiconductors*, ser. Physical and quantum electronics. New York (NY), McGraw-Hill, 1964.

[154] E. O. Kane, "Zener tunneling in semiconductors," *J. Phys. Chem. Solids*, vol. 12, no. 2, pp. 181–188, 1960.

[155] K. H. Gundlach, "Zur berechnung des tunnelstroms durch eine trapezförmige potentialstufe," *Solid-State Electron.*, vol. 9, no. 10, pp. 949–957, 1966.

[156] A. Gehring and S. Selberherr, "Tunneling models for semiconductor device simulation," in *Handbook of theoretical and computational nanotechnology*, M. Rieth and W. Schommers, Eds. American Scientific Publishers, 2006, vol. 10, pp. 469–543.

[157] M. A. Kinch, M. J. Brau, and A. Simmons, "Recombination mechanisms in 8–14–μ HgCdTe," *J. Appl. Phys.*, vol. 44, no. 4, pp. 1649–1663, 1973.

[158] J. P. Rosbeck, R. E. Starr, S. L. Price, and K. J. Riley, "Background and temperature dependent current–voltage characteristics of HgCdTe photodiodes," *J. Appl. Phys.*, vol. 53, no. 9, pp. 6430–6441, 1982.

[159] C. Zener, "A theory of the electrical breakdown of solid dielectrics," *Proc. Royal Soc. London A*, vol. 145, no. 855, pp. 523–529, 1934.

[160] A. G. Chynoweth and K. G. McKay, "Internal field emission in silicon p-n junctions," *Phys. Rev.*, vol. 106, no. 3, pp. 418–426, 1957.

[161] G. H. Wannier, "Possibility of a Zener effect," *Phys. Rev.*, vol. 100, no. 4, p. 1227, 1955.

[162] E. N. Adams, "Definition of energy bands in the presence of an external force field," *Phys. Rev.*, vol. 107, no. 3, pp. 698–701, 1957.

[163] W. Franz, "Zener-effekt und stoßionisation," in *Vorträge des Internationalen Kolloquiums Halbleiter und Phosphore*, 1958, pp. 317–328.

[164] L. V. Keldysh, "Behavior of non-metallic crystals in strong electric fields," *Sov. Phys. JETP*, vol. 33, no. 4, pp. 736–770, 1958.

[165] G. A. M. Hurkx, "On the modelling of tunnelling currents in reverse-biased p-n junctions," *Solid-State Electron.*, vol. 32, no. 8, pp. 665–668, 1989.

[166] G. A. M. Hurkx, D. B. M. Klaassen, and M. P. G. Knuvers, "A new recombination model for device simulation including tunneling," *IEEE Trans. Electron Devices*, vol. 39, no. 2, pp. 331–338, 1992.

[167] R. Adar, "Spatial integration of direct band-to-band tunneling currents in general device structures," *IEEE Trans. Electron Devices*, vol. 39, no. 4, pp. 976–981, 1992.

[168] K. Jóźwikowski, M. Kopytko, A. Rogalski, and A. Jóźwikowska, "Enhanced numerical analysis of current-voltage characteristics of long wavelength infrared n-on-p HgCdTe photodiodes," *J. Appl. Phys.*, vol. 108, no. 7, p. 074519, 2010.

[169] T. Yajima and L. Esaki, "Excess noise in narrow germanium p-n junctions," *J. Phys. Soc. Japan*, vol. 13, no. 11, pp. 1281–1287, 1958.

[170] G. A. M. Hurkx, D. B. M. Klaassen, M. P. G. Knuvers, and F. G. O'Hara, "A new recombination model describing heavy-doping effects and low-temperature behaviour," in *1989 IEEE International Electron Devices Meeting (IEDM '89)*, 1989, pp. 307–310.

[171] G. Vincent, A. Chantre, and D. Bois, "Electric field effect on the thermal emission of traps in semiconductor junctions," *J. Appl. Phys.*, vol. 50, no. 8, pp. 5484–5487, 1972.

[172] S. Selberherr, *Analysis and simulation of semiconductor devices*. Wien, Springer-Verlag, 1984.

[173] M. Vallone, M. Mandurrino, M. Goano, F. Bertazzi, G. Ghione, W. Schirmacher, S. Hanna, and H. Figgemeier, "Numerical modeling of SRH and tunneling mechanisms in High-Operating-Temperature MWIR HgCdTe photodetectors," *J. Electron. Mater.*, vol. 44, no. 9, pp. 3056–3063, 2015.

[174] W. Vandenberghe, B. Sorée, W. Magnus, and M. V. Fischetti, "Generalized phonon-assisted zener tunneling in indirect semiconductors with non-uniform electric fields: a rigorous approach," *J. Appl. Phys.*, vol. 109, no. 12, p. 124503, 2011.

[175] A. S. Verhulst, D. Leonelli, R. Rooyackers, and G. Groeseneken, "Drain voltage dependent analytical model of tunnel field-effect transistors," *J. Appl. Phys.*, vol. 110, no. 2, p. 024510, 2011.

[176] K. Ahmed, M. M. M. Elahi, and M. Shofiqul Islam, "A compact analytical model of band-to-band tunneling in a nanoscale p-i-n diode," in *Informatics, Electronics Vision (ICIEV), 2012 International Conference on*, 2012, pp. 521–524.

[177] A. Schenk, "A model for the field and temperature dependence of Shockley-Read-Hall lifetimes in silicon," *Solid-State Electron.*, vol. 35, no. 11, pp. 1585–1596, 1992.

[178] ——, "Rigorous theory and simplified model of the band-to-band tunneling in silicon," *Solid-State Electron.*, vol. 36, no. 1, pp. 19–34, 1993.

[179] K. Huang and A. Rhys, "Theory of light absorption and non-radiative transitions in f-centres," vol. 204, no. 1078, pp. 406–42, 1950.

[180] C. H. Henry and D. V. Lang, "Nonradiative capture and recombination by multiphonon emission in GaAs and GaP," *Phys. Rev. B*, vol. 15, no. 2, pp. 989–1016, 1977.

[181] A. Palma, A. Godoy, J. A. Jiménez-Tejada, J. E. Carceller, and J. A. López-Villanueva, "Quantum two-dimensional calculation of time constants of random telegraph signals in metal-oxide-semiconductor structures," *Phys. Rev. B*, vol. 56, no. 15, pp. 9565–9574, 1997.

[182] A. Alkauskas, J. L. Lyons, D. Steiauf, and C. G. Van de Walle, "First-principles calculations of luminescence spectrum line shapes for defects in semiconductors: the example of GaN and ZnO," *Phys. Rev. Lett.*, vol. 109, no. 26, p. 267401, 2012.

[183] D. Pons and S. Makram-Ebeid, "Phonon assisted tunnel emission of electrons from deep levels in GaAs," *J. Physique*, vol. 40, no. 12, pp. 1161–1172, 1979.

[184] D. V. Lang, "Deep-level transient spectroscopy: a new method to characterize traps in semiconductor," *J. Appl. Phys.*, vol. 45, no. 7, pp. 3023–3032, 1974.

[185] S. Makram-Ebeid and M. Lannoo, "Quantum model for phonon-assisted tunnel ionization of deep levels in a semiconductor," *Phys. Rev. B*, vol. 25, no. 10, pp. 6406–6424, 1982.

[186] E. O. Kane, "Energy band structure in p-type germanium and silicon," *J. Phys. Chem. Solids*, vol. 1, no. 1–2, pp. 82–99, 1956.

[187] W. Kohn, "Shallow impurity states in silicon and germanium," in *Solid state physics. Advances in research and applications*, F. Seitz and D. Turnbull, Eds. New York (NY), Academic Press, 1957, vol. 5, pp. 257–320.

[188] C. M. Chaves, N. Majlis, and M. Cardona, "Electronic energy bands in GaAs for imaginary crystal momentum," *Solid State Commun.*, vol. 4, no. 12, pp. 631–633, 1966.

[189] F. Jiménez-Molinos, F. Gámiz, A. Palma, P. Cartujo, and J. A. López-Villanueva, "Direct and trap-assisted elastic tunneling through ultrathin gate oxides," *J. Appl. Phys.*, vol. 91, no. 8, pp. 5116–5124, 2002.

[190] J. H. Zheng, H. S. Tan, and S. C. Ng, "Theory of non-radiative capture of carriers by multiphonon processes for deep centres in semiconductors," *J. Phys. Condens. Matter*, vol. 6, no. 9, pp. 1695–1706, 1994.

[191] B. K. Ridley, *Quantum processes in semiconductors*, 5th ed. Oxford (UK), Clarendon Press, 2013.

[192] *Sentaurus Device User Guide. Version L-2016.03*, Synopsys, Inc., Mountain View (CA), Mar. 2016.

[193] F. Rossi, A. Di Carlo, and P. Lugli, "Microscopic theory of quantum-transport phenomena in mesoscopic systems: a Monte Carlo approach," *Phys. Rev. Lett.*, vol. 80, no. 15, pp. 3348–3351, 1998.

[194] R. C. Iotti and F. Rossi, "Microscopic theory of semiconductor-based optoelectronic devices," *Rep. Prog. Phys.*, vol. 68, no. 11, pp. 2533–2571, 2005.

[195] E. Wigner, "On the quantum correction for thermodynamic equilibrium," *Phys. Rev.*, vol. 40, no. 5, pp. 749–759, 1932.

[196] N. C. Kluksdahl, A. M. Kriman, D. K. Ferry, and C. Ringhofer, "Self-consistent study of the resonant-tunneling diode," *Phys. Rev. B*, vol. 39, no. 11, pp. 7720–7735, 1989.

[197] N. C. Kluksdahl, A. M. Kriman, and D. K. Ferry, "Wigner function simulation of quantum tunneling," *Superlatt. Microstruct.*, vol. 4, no. 2, pp. 127–131, 1988.

[198] C. Jacoboni and P. Bordone, "The Wigner-function approach to non-equilibrium electron transport," *Rep. Prog. Phys.*, vol. 67, no. 7, pp. 1033–1071, 2004.

[199] P. Bordone, M. Pascoli, R. Brunetti, A. Bertoni, C. Jacoboni, and A. Abramo, "Quantum transport of electrons in open nanostructures with the Wigner-function formalism," *Phys. Rev. B*, vol. 59, no. 4, pp. 3060–3069, 1999.

[200] L. P. Kadanoff and G. Baym, *Quantum statistical mechanics: Green's function methods in equilibrium and non equilibrium problems*. New York (NY), W. A. Benjamin, 1962.

[201] L. V. Keldysh, "Diagram technique for nonequilibrium processes," *Zh. Eksp. Teor. Fiz.*, vol. 47, no. 4, pp. 1515–1527, 1964.

[202] S. Datta, *Electronic conduction in mesoscopic systems*. Cambridge (UK), Cambridge University Press, 1997.

[203] ——, "Nanoscale device modeling: the Green's function method," *Superlatt. Microstruct.*, vol. 28, no. 4, pp. 253–278, 2000.

[204] S. Steiger, "Modelling Nano-LEDs," Ph.D. dissertation, Swiss federal institute of technology, Zurich, 2009.

[205] A. Rogalski, "Infrared detectors: status and trends," *Progress Quantum Electron.*, vol. 27, no. 2-3, pp. 59–210, 2003.

[206] M. B. Reine, A. K. Sood, and T. J. Tredwell, "Chapter 6. Photovoltaic Infrared Detectors," in *Mercury Cadmium Telluride*, R. K. Willardson and A. C. Beer, Eds. New York (NY), Elsevier, 1981, vol. 18, pp. 201–311.

[207] M. B. Reine, "Review of HgCdTe photodiodes for IR detection," in *Proc. SPIE 4028, VI Infrared Detectors and Focal Plane Arrays*, 2000, pp. 320–330.

[208] M. A. Kinch, "Fundamental physics of infrared detector materials," *J. Electron. Mater.*, vol. 29, no. 6, pp. 809–817, 2000.

[209] M. B. Reine, "HgCdTe photodiodes for IR detection: a review," in *Proc. SPIE 4288, VI Photodetectors: Materials and Devices*, 2001, pp. 266–277.

[210] M. Z. Tidrow and W. R. Dyer, "Infrared sensors for ballistic missile defense," *Infrared Phys. Tech.*, vol. 42, no. 3-5, pp. 333–336, 2001.

[211] M. B. Reine, P. W. Norton, R. Starr, M. H. Weiler, M. Kestigian, B. L. Musicant, P. Mitra, T. Schimert, F. C. Case, Lb. Bhat, H. Ehsani, and V. Rao, "Independently accessed back-to-back HgCdTe photodiodes: a new dual-band infrared detector," *J. Electron. Mater.*, vol. 24, no. 5, pp. 669–679, 1995.

[212] M. B. Reine, A. W. Hairston, P. O'Dette, S. P. Tobin, F. T. J. Smith, B. L. Musicant, P. Mitra, and F. C. Case, "Simultaneous MW/LW dual-band MOVPE HgCdTe 64×64 FPAs," in *Proc. SPIE 3379, V Infrared Detectors and Focal Plane Arrays*, 1998, pp. 200–212.

[213] R. D. Rajavel, D. M. Jamba, J. E. Jensen, O. K. Wu, P. D. Brewer, J. A. Wilson, J. L. Johnson, E. A. Patten, K. Kosai, J. T. Caulfield, and P. M. Goetz, "Molecular beam epitaxial growth and performance of HgCdTe-based simultaneous-mode two-color detectors," *J. Electron. Mater.*, vol. 27, no. 6, pp. 747–751, 1998.

[214] R. D. Rajavel, D. M. Jamba, J. E. Jensen, O. K. Wu, J. A. Wilson, J. L. Johnson, E. A. Patten, K. Kosai, P. Goetz, and S. M. Johnson, "Molecular beam epitaxial growth and performance of integrated multispectral HgCdTe photodiodes for the detection of mid-wave infrared radiation," *J. Cryst. Growth*, vol. 184-185, pp. 1272–1278, 1998.

[215] R. D. Rajavel, D. M. Jamba, O. K. Wu, J. E. Jensen, J. A. Wilson, E. A. Patten, K. Kosai, P. Goetz, G. R. Chapman, and W. A. Radford, "High performance HgCdTe two-color infrared detectors grown by molecular beam epitaxy," *J. Cryst. Growth*, vol. 175-176, pp. 653–658, 1997.

[216] R. A. Coussa, A. M. Gallagher, K. Kosai, L. T. Pham, G. K. Pierce, E. P. Smith, G. M. Venzor, T. J. De Lyon, J. E. Jensen, B. Z. Nosho, J. A. Roth, and J. R. Waterman, "Spectral crosstalk by radiative recombination in sequential-mode, dual mid-wavelength infrared band HgCdTe detectors," *J. Electron. Mater.*, vol. 33, no. 6, pp. 517–525, 2004.

[217] E. P. G. Smith, L. T. Pham, G. M. Venzor, E. M. Norton, M. D. Newton, P. M. Goetz, V. K. Randall, A. M. Gallagher, G. K. Pierce, E. A. Patten, R. A. Coussa, K. Kosai, W. A. Radford, L. M. Giegerich, J. M. Edwards, S. M. Johnson, S. T. Baur, J. A. Roth, B. Nosho, T. J. De Lyon, J. E. Jensen, and R. E. Longshore, "HgCdTe focal plane arrays for dual-color mid- and long-wavelength infrared detection," *J. Electron. Mater.*, vol. 33, no. 6, pp. 509–516, 2004.

[218] E. Bellotti and D. D'Orsogna, "Numerical analysis of HgCdTe simultaneous two-color photovoltaic infrared detectors," *IEEE J. Quantum Electron.*, vol. 42, no. 4, pp. 418–426, 2006.

[219] T. Ashley and C. T. Elliott, "Nonequilibrium devices for infra-red detection," *Electron. Lett.*, vol. 21, no. 10, pp. 451–452, 1985.

[220] M. A. Kinch, "HgCdTe: Recent trends in the ultimate IR semiconductor," *J. Electron. Mater.*, vol. 39, no. 7, pp. 1043–1052, 2010.

[221] S. Velicu, C. Grein, P. Emelie, A. Itsuno, J. Philips, and P. Wijewarnasuriya, "MWIR and LWIR HgCdTe infrared detectors operated with reduced cooling requirements," *J. Electron. Mater.*, vol. 39, no. 7, pp. 873–881, 2010.

[222] H. F. Schaake, M. A. Kinch, D. Chandra, P. K. Liao, D. F. Weirauch, C.-F. Wan, and H. D. Shih, "High operating temperature MWIR detectors," in *Proc. SPIE 7608, VII Quantum Sensing and Nanophotonic Devices*, 2010.

[223] J. Wenisch, H. Bitterlich, M. Bruder, P. Fries, R. Wollrab, J. Wendler, R. Breiter, and J. Ziegler, "Large-format and long-wavelength infrared mercury cadmium telluride detectors," *J. Electron. Mater.*, vol. 42, no. 11, pp. 3186–3190, 2013.

[224] J. Ziegler, D. Eich, M. Mahlein, T. Schallenberg, R. Scheibner, J. Wendler, J. Wenisch, R. Wollrab, V. Daumer, R. Rehm, F. Rutz, and M. Walther, "The development of 3rd generation IR detectors at AIM," in *Proc. SPIE 7608, XXXVII Infrared Technology and Applications*, 2011.

[225] R. Wollrab, W. Schirmacher, T. Schallenberg, H. Lutz, J. Wendler, and J. Ziegler, "Recent progress in the development of HOT MWIR detectors," in *6th International Symposium on Optronics in Defense and Security (OPTRO 2014)*, 2014.

[226] T. N. Casselman, "Calculation of the auger lifetime in p-type $Hg_{1-x}Cd_xTe$," *J. Appl. Phys.*, vol. 52, no. 2, pp. 848–854, 1981.

[227] S. Krishnamurthy, M. A. Berding, Z. G. Yu, C. H. Swartz, T. H. Myers, D. D. Edwall, and R. DeWames, "Model for minority carrier lifetimes in doped HgCdTe," *J. Electron. Mater.*, vol. 34, no. 6, pp. 873–879, 2005.

[228] C. T. Elliott, N. T. Gordon, R. S. Hall, and G. Crimes, "Reverse breakdown in long wavelength lateral collection $Cd_xHg_{1-x}Te$ diodes," *J. Vac. Sci. Technol. A*, vol. 8, no. 2, pp. 1251–1253, 1990.

[229] Y. Okuto and C. R. Crowell, "Energy-conservation considerations in the characterization of impact ionization in semiconductors," *Phys. Rev. B*, vol. 6, no. 8, pp. 3076–3081, 1972.

[230] J. Rothman, L. Mollard, S. Goût, L. Bonnefond, and J. Wlassow, "History-dependent impact ionization theory applied to HgCdTe e-APDs," *J. Electron. Mater.*, vol. 40, no. 8, pp. 1757–1768, 2011.

[231] M. A. Kinch, J. D. Beck, C.-F. Wan, F. Ma, and J. Campbell, "HgCdTe electron avalanche photodiodes," *J. Electron. Mater.*, vol. 33, no. 6, pp. 630–639, 2004.

[232] M. Mandurrino, G. Verzellesi, M. Goano, M. Vallone, F. Bertazzi, G. Ghione, M. Meneghini, G. Meneghesso, and E. Zanoni, "Trap-assisted tunneling in InGaN/GaN LEDs: experiments and physics-based simulation," in *Numerical Simulation of Optoelectronic Devices, 2014, 14th International Conference on*, Sep. 2014, pp. 13–14.

[233] M. Mandurrino, G. Verzellesi, M. Goano, S. Dominici, F. Bertazzi, G. Ghione, M. Meneghini, G. Meneghesso, and E. Zanoni, "Defect-related tunneling contributions to subthreshold forward current in GaN-based LEDs," in *Fotonica AEIT Italian Conference on Photonics Technologies, 2015*, May 2015, pp. 1–4.

[234] M. Mandurrino, M. Goano, S. Dominici, M. Vallone, F. Bertazzi, G. Ghione, M. Bernabei, L. Rovati, G. Verzellesi, M. Meneghini, G. Meneghesso, and E. Zanoni, "Trap-assisted tunneling contributions to subthreshold forward

current in InGaN/GaN light-emitting diodes," in *Proc. SPIE 9571, XIV International Conference on Solid State Lighting and LED-based Illumination Systems*, Aug. 2015, p. 95710U.

[235] M. Goano, F. Bertazzi, X. Zhou, M. Mandurrino, S. Dominici, M. Vallone, G. Ghione, A. Tibaldi, M. Calciati, P. Debernardi, F. Dolcini, F. Rossi, G. Verzellesi, M. Meneghini, N. Trivellin, C. De Santi, E. Zanoni, and E. Bellotti, "Challenges towards the simulation of GaN-based LEDs beyond the semi-classical framework," in *Proc. SPIE 9571, XXIV International Conference on Physics and Simulation of Optoelectronic Devices*, Feb. 2016, p. 974202.

[236] N. S. Averkiev, A. E. Chernyakov, M. E. Levinshtein, P. V. Petrov, E. B. Yakimov, N. M. Shmidt, and E. I. Shabunina, "Two channels of non-radiative recombination in InGaN/GaN LEDs," *Physica B*, vol. 404, no. 23-24, pp. 4896–4898, 2009.

[237] K. B. Lee, P. J. Parbrook, T. Wang, J. Ba, F. Ranalli, R. J. Airey, and G. Hill, "The origin of the high ideality factor in AlGaN-based quantum well ultraviolet light emitting diodes," *Phys. Status Solidi B*, vol. 247, no. 7, pp. 1761–1763, 2010.

[238] S. Huang, Y. Xian, B. Fan, Z. Zheng, Z. Chen, W. Jia, H. Jiang, and G. Wang, "Contrary luminescence behaviors of InGaN/GaN light emitting diodes caused by carrier tunneling leakage," *J. Appl. Phys.*, vol. 110, no. 6, p. 064511, 2011.

[239] J.-M. Lee and S.-B. Kim, "Analysis of current components and estimation of internal quantum efficiency in light-emitting diodes," *IEEE Trans. Electron Devices*, vol. 58, no. 9, pp. 3053–3057, 2011.

[240] T.-K. Yang and S.-B. Kim, "Quantitative analysis of initial short-term aging behavior and its implication for the efficiency droop in blue light-emitting diodes," *IEEE Trans. Electron Devices*, vol. 59, no. 10, pp. 2656–2661, 2012.

[241] L. Liu, J. Yang, D. Teng, S. Qi, and G. Wang, "An explanation for invalidity of working currents' derating on improving light-emitting diode devices' reliability," *J. Appl. Phys.*, vol. 114, no. 2, p. 023102, 2013.

[242] B. Kang and S.-B. Kim, "Temperature dependence of the component currents and internal quantum efficiency in blue light-emitting diodes," *IEEE Trans. Electron Devices*, vol. 60, no. 3, pp. 1060–1067, 2013.

[243] D. S. Meyaard, J. Cho, E. F. Schubert, S.-H. Han, M.-H. Kim, and C. Sone, "Analysis of the temperature dependence of the forward voltage characteristics of GaInN light-emitting diodes," *Appl. Phys. Lett.*, vol. 103, no. 4, p. 121103, 2013.

[244] X. A. Cao, E. B. Stokes, P. M. Sandvik, S. F. LeBoeuf, J. Kretchmer, and D. Walker, "Diffusion and tunneling currents in GaN/InGaN multiple quantum well light-emitting diodes," *IEEE Electron Device Lett.*, vol. 23, no. 9, pp. 535–537, 2002.

[245] X. A. Cao, J. M. Teetsov, M. P. D'Evelyn, D. W. Merfeld, and C. H. Yan, "Electrical characteristics of InGaN/GaN light-emitting diodes grown on GaN and sapphire substrates," *Appl. Phys. Lett.*, vol. 85, no. 1, pp. 7–9, 2004.

[246] H. C. Casey Jr., J. Muth, S. Krishnankutty, and J. M. Zavada, "Dominance of tunneling current and band filling in InGaN/AlGaN double heterostructure blue lightemitting diodes," *Appl. Phys. Lett.*, vol. 68, no. 3, pp. 2867–2869, 1996.

[247] P. Perlin, M. Osinski, P. G. Eliseev, V. A. Smagley, J. Mu, M. Banas, and P. Sartori, "Low-temperature study of current and electroluminescence in InGaN/AlGaN/GaN double-heterostructure blue light-emitting diodes," *Appl. Phys. Lett.*, vol. 69, no. 12, pp. 1680–1682, 1996.

[248] L. Hirsch and A. S. Barrière, "Electrical characterization of InGaN/GaN light emitting diodes grown by molecular beam epitaxy," *J. Appl. Phys.*, vol. 94, no. 8, pp. 5014–5020, 2003.

[249] D. Yan, H. Lu, D. Chen, R. Zhang, and Y. Zheng, "Forward tunneling current in GaN-based blue light-emitting diodes," *Appl. Phys. Lett.*, vol. 96, no. 8, p. 083504, 2010.

[250] J. Kim, Y. Tak, J. Kim, S. Chae, J.-K. Kim, and Y. Park, "Analysis of forward tunneling current in InGaN/GaN multiple quantum well light-emitting diodes grown on Si (111) substrate," *J. Appl. Phys.*, vol. 114, no. 1, p. 013101, 2013.

[251] C. L. Reynolds and A. Patel, "Tunneling entity in different injection regimes of InGaN light emitting diodes," *J. Appl. Phys.*, vol. 103, no. 8, p. 086102, 2008.

[252] R. Nana, P. Gnanachchelvi, M. A. Awaah, M. H. Gowda, A. M. Kamto, Y. Wang, M. Park, and K. Das, "Effect of deep-level states on current-voltage characteristics and electroluminescence of blue and UV light-emitting diodes," *Phys. Status Solidi A*, vol. 207, no. 6, pp. 1489–1496, 2010.

[253] C.-T. Sah, R. N. Noyce, and W. Shockley, "Carrier generation and recombination in p-n junctions and p-n junction characteristics," *Proc. IRE*, vol. 45, no. 9, pp. 1228–1243, 1957.

[254] J. M. Shah, Y.-L. Li, Th. Gessmann, and E. F. Schubert, "Experimental analysis and theoretical model for anomalously high ideality factors (n≫2.0) in AlGaN/GaN p-n junction diodes," *J. Appl. Phys.*, vol. 94, no. 4, pp. 2627–2630, 2003.

[255] D. Zhu, J. Xu, A. N. Noemaun, J. K. Kim, E. F. Schubert, M. H. Crawford, and D. D. Koleske, "The origin of the high diode-ideality factors in GaInN/GaN multiple quantum well light-emitting diodes," *Appl. Phys. Lett.*, vol. 94, no. 3, p. 081113, 2009.

[256] M. G. Ancona and H. F. Tiersten, "Macroscopic physics of the silicon inversion layer," *Phys. Rev. B*, vol. 35, no. 15, pp. 7959–7965, 1987.

[257] M. G. Ancona and G. J. Iafrate, "Quantum correction to the equation of state of an electron gas in a semiconductor," *Phys. Rev. B*, vol. 39, no. 13, pp. 9536–9540, 1989.

[258] D. K. Ferry and J.-R. Zhou, "Form of the quantum potential for use in hydrodynamic equations for semiconductor device modeling," *Phys. Rev. B*, vol. 48, no. 11, pp. 7944–7950, 1993.

[259] R. Pinnau, "A review on the quantum drift diffusion model," *Transport Theor. Stat.*, vol. 31, no. 4–6, pp. 367–395, 2002.

[260] M. G. Ancona and K. Lilja, "Multi-dimensional tunneling in density-gradient theory," *J. Comp. Electron.*, vol. 3, no. 3, pp. 189–192, 2004.

[261] R.-C. Chen and J.-L. Liu, "A quantum corrected energy-transport model for nanoscale semiconductor devices," *J. Comp. Phys.*, vol. 204, no. 1, pp. 131–156, 2005.

[262] C. de Falco, E. Gatti, A. L. Lacaita, and R. Sacco, "Quantum-corrected drift-diffusion models for transport in semiconductor devices," *J. Comp. Phys.*, vol. 204, no. 2, pp. 533–561, 2005.

[263] M. Rudan, E. Gnani, S. Reggiani, and G. Baccarani, "A coherent extension of the transport equations in semiconductors incorporating the quantum correction - part I: single-particle dynamics," *IEEE Trans. Nanotech.*, vol. 4, no. 5, pp. 495–502, 2005.

[264] ——, "A coherent extension of the transport equations in semiconductors incorporating the quantum correction - part II: collective transport," *IEEE Trans. Nanotech.*, vol. 4, no. 5, pp. 503–509, 2005.

[265] M. G. Ancona and A. Svizhenko, "Density-gradient theory of tunneling: Physics and verification in one dimension," *J. Appl. Phys.*, vol. 104, no. 7, p. 073726, 2008.

[266] G. Baccarani, E. Gnani, A. Gnudi, S. Reggiani, and M. Rudan, "Theoretical foundations of the quantum drift-diffusion and density-gradient models," *Solid-State Electron.*, vol. 52, no. 4, pp. 526–532, 2008.

[267] M. G. Ancona, "Density-gradient theory: a macroscopic approach to quantum confinement and tunneling in semiconductor devices," *J. Comp. Electron.*, vol. 10, no. 1, pp. 65–97, 2011.