

Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Electrical, Electronics and Communications Engineering (38^{th} cycle)

Quantum transport modeling in III-Nitride optoelectronic devices

By

Francesco Mercinelli

******

Supervisors:
Prof. Francesco Bertazzi, Supervisor
Prof. Alberto Tibaldi, Co-Supervisor
Prof. Michele Goano, Co-Supervisor

Referees:
Prof. Matthias Auf der Maur, Tor Vergata University of Rome
Prof. Giovanni Verzellesi, University of Modena and Reggio Emilia

Politecnico di Torino
2026

Abstract

This work investigates quantum transport and quantum-corrected drift-diffusion models for the simulation and optimization of III-nitride optoelectronic devices. Their operation is strongly influenced by quantum confinement, tunneling through heterostructure barriers, trap-assisted tunneling, polarization fields and interband tunneling. These mechanisms cannot be fully described by conventional semiclassical drift-diffusion simulations, whereas a complete microscopic quantum-transport treatment is often too computationally demanding for realistic device structures and optimization studies. The objective of this thesis is therefore to establish a hierarchy of models ranging from microscopic quantum transport to computationally efficient quantum-corrected approaches, and to determine which physical information must be retained to describe the dominant transport and recombination mechanisms.

The non-equilibrium Green's function formalism is introduced as the microscopic reference, providing a unified description of the electronic structure, open boundaries, spectral carrier distributions, tunneling and scattering processes. A reduced effective-mass NEGF formulation based on Büttiker probes is also considered to lower the computational cost while preserving a spectral description of carrier transport. Two quantum-corrected drift-diffusion models are then formulated as further approximations: Poisson-Schrödinger drift-diffusion and density-gradient drift-diffusion. Both retain semiclassical current relations but modify the carrier densities and effective band edges to account for quantum confinement and tunneling-related charge redistribution. The Poisson-Schrödinger approach explicitly reconstructs the quantum local density of states from the Schrödinger eigenstates, whereas the density-gradient model introduces quantum corrections through differential relations involving the carrier densities.

The different levels of description are first compared for trap-assisted tunneling in a visible InGaN/GaN light-emitting diode. Starting from microscopic multi-phonon defect interactions, a quantum recombination rate with the algebraic structure of the conventional Shockley-Read-Hall expression is derived in terms of spectral capture and emission processes. Under strong electric fields, spectral states extend into the nominal bandgap and provide tunneling paths between the

extended bands and localized defects. Trap-assisted tunneling is therefore described through the quantum spectral charge and multi-phonon capture probabilities without introducing empirical field-enhancement factors, while the conventional SRH rate is recovered in the homogeneous semiclassical bulk limit. The NEGF calculation reproduces the experimental current-voltage characteristic and the increase of the ideality factor in the subthreshold regime over most of the investigated bias range. Among the reduced models, Poisson-Schrödinger drift-diffusion provides the closest agreement with NEGF because it retains the spectral information required by the capture process. Density-gradient drift-diffusion captures the qualitative behavior at a lower computational cost but exhibits larger deviations in the tunneling region.

The Poisson-Schrödinger framework is subsequently applied to the optimization of a 275 nm AlGaN UV-C LED. The electron-blocking region is investigated by comparing a conventional bulk layer with polarization-doped graded and superlattice designs. Both alternative structures improve hole injection while maintaining electron confinement, increasing the peak internal quantum efficiency from approximately 40% for the bulk EBL to 57% and 56% for the graded and superlattice EBLs, respectively. The p-side contact is then optimized to reduce the optical absorption associated with conventional p-GaN. A polarization-doped graded contact and a tunnel-junction contact are considered, with the tunnel junction analyzed through quantum transport because its operation relies on interband tunneling. The graded contact increases the peak wall-plug efficiency from 5.08% to 6.25%, while the optimized tunnel-junction contact reaches 8.78%. The latter provides the largest improvement in light extraction and external quantum efficiency, although its advantage depends on limiting the additional voltage drop introduced by the junction.

Overall, NEGF provides the detailed spectral description required to identify and validate quantum transport mechanisms, while Poisson-Schrödinger drift-diffusion retains the most relevant quantum information at a computational cost compatible with device-level optimization. The resulting modeling strategy provides a practical bridge between microscopic transport physics and the design of more efficient III-nitride optoelectronic devices.

Contents

1 Introduction **1**

2 The Non-equilibrium Green's functions formalism **7**

2.1 Steady-state Green's functions 8

2.2 Dyson and Keldysh equations 11

2.3 General expression for physical observables 12

2.3.1 Carrier density 13

2.3.2 Local density of states 13

2.3.3 Current density and divergence 14

2.4 Spatial discretization 15

2.5 Multiband $\mathbf{k} \cdot \mathbf{p}$ band-structure model 18

2.6 Boundary self-energies 21

2.7 Interaction self-energies 24

2.7.1 Acoustic phonon scattering self-energy 26

2.7.2 Polar optical phonon scattering 26

2.7.3 Multi-phonon defect self-energies 27

2.8 Self-consistent solution of the full NEGF scheme 29

2.9 Effective-mass Büttiker-probe NEGF approximation 34

2.9.1 Effective mass approximation for electron and hole Green's functions . 34

2.9.2 Büttiker-probe self-energies 35

2.9.3 Boundary self-energies in the effective-mass Büttiker-probe formulation 38
2.9.4 Büttiker-probe self-energy scheme 41

3 Quantum-corrected drift-diffusion 45
3.1 General quantum-corrected drift-diffusion formalism 46
3.2 Poisson-Schrödinger approach 49
3.3 Density-gradient approximation 55

4 Trap-assisted tunneling modeling in NEGF and QCDD 58
4.1 Trap-assisted tunneling in the NEGF Framework 61
4.1.1 Multi-phonon defect self-energies and SRH-like rate 62
4.1.2 Numerical results 70
4.2 TAT in quantum-corrected drift-diffusion formalism: Poisson-Schrödinger . . . 75
4.2.1 Multi-phonon SRH formulation with Poisson-Schrödinger quantum densities 75
4.2.2 Numerical results 77
4.3 TAT in quantum-corrected drift-diffusion formalism: density-gradient 82
4.3.1 Multi-phonon capture lifetimes in the density-gradient formalism . . . 82
4.3.2 Numerical results 85
4.4 Comparison between the models 89

5 AlGaN UV-C LED optimization 99
5.1 Electron Blocking Layer (EBL) optimization 102
5.1.1 Polarization-doped electron blocking layer (PD-EBL) 103
5.1.2 Superlattice electron blocking layer (SL-EBL) 106
5.1.3 IQE comparison 114
5.2 p-contact layer optimization 116
5.2.1 Polarization-doped graded contact (GC) 118

5.2.2 Tunnel-junction contact: optimization and application 119
5.2.3 Efficiency comparison 130

6 Conclusions and outlook **135**

References **138**

Appendix A Wurtzite eight-band k · p Hamiltonian **145**

Appendix B Recursive Green's function method **151**

B.1 Recursive formulation of the retarded Green's function 152
B.1.1 Forward recursion for the left-connected Green's function 152
B.1.2 Backward recursion for the diagonal blocks of G^R 153
B.1.3 Reconstruction of the off-diagonal blocks 155
B.2 Recursive formulation of the lesser Green's function 157
B.2.1 Forward recursion for the left-connected lesser Green's function 157
B.2.2 Backward recursion for the diagonal blocks of $G^<$ 159
B.2.3 Reconstruction of the off-diagonal blocks of $G^<$ 160

Appendix C Recovery of the SRH rate in the semiclassical bulk limit **163**

Chapter 1

Introduction

The III-nitride material system has played a central role in the development of modern opto-electronic and electronic devices. Its most visible technological success is the GaN-based blue light-emitting diode, whose development enabled efficient white solid-state lighting and was recognized by the 2014 Nobel Prize in Physics awarded to Isamu Akasaki, Hiroshi Amano, and Shuji Nakamura for their work in the 1980s and 1990s [1–3]. This achievement marked a turning point in lighting technology, making compact, efficient, and long-lasting light sources available for general illumination, displays, automotive lighting, and many other applications.
The importance of III-nitrides for opto-electronics is not limited to blue and white LEDs. The material family, which includes InN, GaN, AlN and their ternary alloys, provides direct optical transitions over an exceptionally broad spectral range. As shown in Fig. 1.1, the bandgap extends from approximately 0.7 eV in InN, corresponding to infrared wavelengths around 1.8 μm, to about 6.1 eV in AlN, corresponding to deep-ultraviolet wavelengths close to 200 nm. GaN, with a bandgap of about 3.4 eV, lies close to the ultraviolet edge of the visible spectrum [4]. By tuning the alloy composition of $In_xGa_{1-x}N$ and $Al_xGa_{1-x}N$, III-nitride heterostructures can therefore be engineered for emission from the infrared and visible range to the deep ultraviolet. This spectral flexibility explains why III-nitrides have become a reference platform for optoelectronics. InGaN/GaN LEDs are widely used in solid-state lighting, displays, automotive lighting, visible-light communication, and high-brightness light sources. Nitride-based laser diodes are also important for projection systems, optical storage, and compact high-brightness emitters. On the short-wavelength side, alloying GaN with AlN makes it possible to realize AlGaN-based ultraviolet emitters covering almost the entire UV range, from about 400 nm down to nearly 210 nm. Depending on the emission wavelength, UV LEDs are attractive for water purification, surface sterilization, environmental and biological sensing, phototherapy, plant-growth lighting,

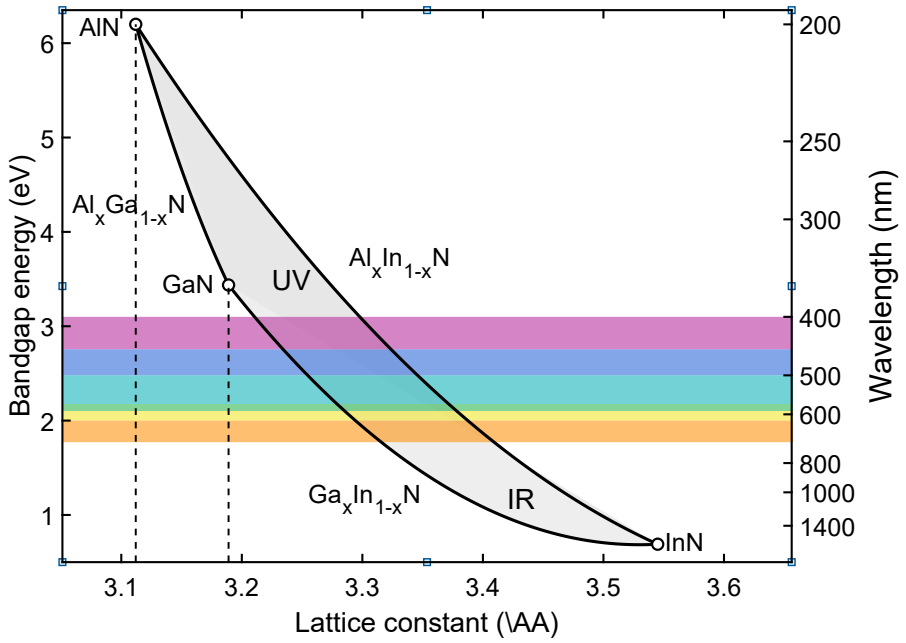


Fig. 1.1 Bandgap energy as a function of lattice constant for the III-nitride material system. The alloy families $In_xGa_{1-x}N$ and $Al_xGa_{1-x}N$ allow the direct bandgap to be tuned from the infrared, close to InN, to the deep ultraviolet, close to AlN.

UV curing, and photo-patterning applications such as resist exposure and printed-circuit-board processing. The interest in UV-C LEDs is particularly strong because photons in the 100–280 nm spectral range can induce photochemical damage in the nucleic acids of microorganisms, preventing their reproduction. The germicidal effectiveness peaks around 265 nm, making AlGaN-based UV-C emitters attractive for compact and mercury-free disinfection systems [5, 6]. Compared with conventional mercury lamps, UV LEDs offer small form factors, moderate driving voltages, fast switching, wavelength selectivity, and environmentally friendlier operation. These properties make them promising for point-of-use water purification, portable sterilization systems, medical instrumentation, and industrial photochemical processes.

Beyond optoelectronics, III-nitrides are also highly relevant for electronic devices. GaN and AlN-rich alloys combine large breakdown fields, high thermal stability, and strong chemical robustness, making them attractive for power conversion, radio-frequency electronics, and high-voltage switching. Therefore, the same material family provides a common platform for efficient light generation, ultraviolet emission, and advanced electronic devices.

Despite this technological success, the performance of III-nitride devices is still limited by several physical mechanisms that are not completely understood. In visible InGaN/GaN LEDs, one of the most studied problems is the efficiency droop, namely the decrease of the internal

quantum efficiency at increasing current density. In AlGaN-based UV-C LEDs, the situation is even more challenging. The high aluminum content required for deep-ultraviolet emission makes p-type doping difficult, increases resistive losses, worsens carrier injection, and enhances the impact of defects and nonradiative recombination. In addition, light extraction becomes more difficult because of the optical properties of Al-rich AlGaN alloys and the polarization of the emitted radiation. As a consequence, UV-C LEDs still exhibit efficiencies much lower than those of visible GaN-based LEDs [7].
These limitations are not independent of each other. Carrier injection, electrostatics, polarization fields, transport across heterostructure barriers, quantum confinement, tunneling, defect-assisted recombination, and heat generation are strongly coupled in III-nitride devices. Numerical simulation is therefore essential not only to reproduce experimental data, but also to identify the dominant loss mechanisms and guide the optimization of device structures.

The standard framework for semiconductor device simulation is the drift-diffusion model. Its success is due to its numerical robustness, low computational cost, and suitability for realistic device geometries. In this model, carriers are described through local densities, quasi-Fermi levels, mobilities, and recombination rates. This description is extremely useful for many semiconductor devices, but it remains semiclassical. Therefore, it cannot natively describe quantum confinement, tunneling, spectral broadening, or transport through classically forbidden regions.
This limitation is particularly important in III-nitride heterostructures. Thin quantum wells, strong polarization fields, high barriers, and large band offsets make quantum effects central to the behavior of many devices. In some cases, quantum effects appear through electrical signatures that cannot be reproduced by the standard semiclassical diode picture. Ideal diffusion current gives an ideality factor close to one, while conventional Shockley-Read-Hall recombination in the depletion region gives an ideality factor close to two. Therefore, ideality factors significantly larger than two, in some structures reaching values of several tens [8, 9], indicate the presence of additional transport or recombination mechanisms, typically involving strong electric fields, defects, and tunneling through classically forbidden regions.
In other cases, quantum effects are not a parasitic contribution, but are part of the intended operating principle of the device. A relevant example is the tunnel junction, where interband tunneling is used to improve carrier injection and to overcome the limitations associated with resistive p-type layers and p-contacts. This approach is particularly attractive for AlGaN-based ultraviolet devices, where the high activation energy of Mg acceptors and the difficulty of forming low-resistance p-type contacts strongly limit the device performance [10, 11].
In practical drift-diffusion simulations, quantum effects are often introduced through phenomeno-

logical corrections. Examples include field-enhanced recombination lifetimes depending on tunneling masses, barrier-lowering factors, or empirical correction functions. Such models can reproduce experimental results with good accuracy, but they often rely on fitting parameters. As a consequence, agreement with measurements does not necessarily guarantee that the underlying microscopic mechanism is correctly described. This becomes a serious limitation when simulations are used predictively, for example to optimize new device designs or to extrapolate toward material systems and geometries that have not yet been experimentally tested.

A more rigorous description is provided by the non-equilibrium Green's function formalism. NEGF is a quantum-kinetic approach in which confinement, tunneling, open-boundary effects, local density of states, spectral current, and scattering processes can be treated within a unified framework. For this reason, it represents a natural reference for the study of quantum transport in semiconductor nanostructures and heterostructure devices. However, its computational cost remains very high. This is especially restrictive in wide-bandgap materials, where large energy windows must be considered in order to include electron and hole transport and tunneling paths. Moreover, realistic LED structures may extend over tens or hundreds of nanometers, making full multiband NEGF simulations impractical for routine device optimization.
For this reason, reduced NEGF formulations are often required when the objective is to retain a quantum-transport description while lowering the computational cost. One possible simplification concerns the band-structure model: when transport is dominated by states close to a single band edge, the multiband Hamiltonian can be replaced by an effective-mass approximation. A second simplification concerns the treatment of scattering. Instead of evaluating fully non-local microscopic interaction self-energies, relaxation processes can be described through Büttiker probes, whose local quasi-Fermi levels are adjusted to enforce current conservation. These approximations preserve the spectral description of carrier density, current, tunneling, and open-boundary transport, but reduce the level of microscopic detail in the band structure and scattering mechanisms.
Quantum-corrected drift-diffusion models provide a further level of simplification between reduced NEGF and semiclassical drift-diffusion. Their goal is to retain the numerical efficiency of drift-diffusion while introducing quantum effects through an effective potential or through a quantum-corrected density of states. In the Poisson-Schrödinger drift-diffusion approach, the Schrödinger equation is solved to obtain quantum-confined states and carrier densities, which are then mapped onto effective band edges for the drift-diffusion solver. In the density-gradient approach, quantum corrections are introduced through a potential depending on the spatial variation of the carrier density. These methods are computationally more accessible than NEGF, but their accuracy is not guaranteed a priori, because transport remains semiclassical and the

spectral nature of quantum transport is only approximately retained.
The models considered in this thesis can therefore be organized as a hierarchy of progressively simplified physical descriptions. At the highest level, full multiband NEGF provides the most detailed microscopic description, but at the largest computational cost. Effective-mass and Büttiker-probe NEGF reduce this cost while retaining a quantum-transport formulation. Poisson-Schrödinger and density-gradient drift-diffusion then provide increasingly efficient quantum-corrected approximations for larger device simulations, while classical drift-diffusion represents the local semiclassical limit, where no quantum effects are considered. This hierarchy is summarized in Table 1.1.

Model	**Description**	**Advantages**	**Limitations**
Multiband $\mathbf{k} \cdot \mathbf{p}$ NEGF with full non local scattering	Quantum transport with multiband electronic structure and microscopic scattering.	Most complete model; includes band coupling, interband tunneling, spectral transport, and scattering.	Very high computational cost; restricted to small structures.
EMA NEGF with Büttiker probes	Quantum transport with parabolic bands and phenomenological local scattering self-energies.	Retains quantum LDOS, spectral observables, and quantum transport at lower cost than multiband SCBA.	Approximate band structure and scattering; still computationally demanding for complete devices.
Poisson-Schrödinger drift-diffusion	Quantum LDOS and charge from the Schrödinger equation; semiclassical transport.	Retains confined states and energy-resolved carrier densities at a much lower cost than NEGF.	Transport remains semiclassical; repeated eigenvalue solutions limit device size and generally restrict the approach to 1D.
Density-gradient drift-diffusion	Semiclassical transport with quantum corrections obtained from spatial variations of the carrier densities.	No eigenvalue problem; low computational cost and suitable for larger devices and 2D/3D simulations.	No explicit quantum LDOS or subband description; less accurate for strongly localized states and sharp spectral features.
Semiclassical drift-diffusion	Semiclassical carrier transport based on local carrier densities, quasi-Fermi levels, and local recombination rates, without explicit quantum corrections.	Lowest computational cost.	Does not describe quantum confinement, tunneling, LDOS broadening, or energy-resolved transport.

Table 1.1 Main advantages and limitations of the transport models considered in this work.

The aim of this thesis is to investigate the use of quantum transport and quantum-corrected drift-diffusion models for the simulation of III-nitride optoelectronic devices. The work begins by introducing the NEGF formalism, which provides the microscopic reference used to describe quantum transport, carrier densities, local density of states, spectral currents, boundary self-

energies, scattering self-energies, and Büttiker-probe approximations. The drift-diffusion formalism is then reviewed together with the quantum-corrected approaches used in the rest of the thesis, namely the density-gradient and Poisson-Schrödinger drift-diffusion models.
All numerical transport models presented in this thesis are implemented using in-house simulation codes, ensuring full control over all aspects of the numerical implementation and simulation framework. The drift-diffusion and quantum-corrected drift-diffusion calculations are carried out in MATLAB, while the more computationally demanding NEGF calculations are implemented in Fortran.
The different levels of description are first compared in the context of trap-assisted tunneling in a visible InGaN/GaN LED. This problem is a useful benchmark because it directly links the electrical signatures of nonclassical transport, such as high ideality factors and subthreshold leakage, to the quantum density of states and to tunneling-assisted defect recombination. In this part of the thesis, NEGF is used as the reference model, while Poisson-Schrödinger and density-gradient drift-diffusion are assessed as practical alternatives. The objective is to determine which quantum-corrected model can reproduce the relevant NEGF results while remaining computationally feasible for larger device simulations.
The insight obtained from this comparison is then used for the study of AlGaN UV-C LEDs. In this case, the objective is to exploit a computationally feasible quantum-corrected model to analyze the dominant loss mechanisms and to guide the optimization of the device structure. The study focuses first on the electron-blocking region, where polarization-doped and superlattice designs are investigated to improve carrier confinement and injection. It then addresses the p-side contact region, comparing conventional p-GaN contacts, polarization-doped graded contacts, and tunnel-junction-based contacts. Since tunnel junctions rely intrinsically on interband tunneling, their analysis requires a quantum-transport description, linking again the device-optimization problem to the central modeling theme of the thesis.
In this way, the thesis connects microscopic quantum-transport modeling with practical simulation strategies for the design of III-nitride optoelectronic devices. NEGF is used to establish a physically grounded reference, quantum-corrected drift-diffusion is evaluated as a computationally efficient approximation, and the resulting modeling framework is applied to the optimization of UV-C LED structures.

Chapter 2

The Non-equilibrium Green's functions formalism

As the dimensions of modern optoelectronic devices are reduced, quantum effects such as carrier confinement and tunneling become increasingly significant and are often deliberately exploited to improve device performance. For instance, in quantum-well structures the degree of confinement depends on the well geometry, enabling precise tuning of the emission spectrum of lasers and light-emitting diodes. Moreover, nanoscale dimensions lead to shorter tunneling distances and stronger electric fields, making both intraband and interband tunneling non-negligible. In some cases, tunneling is not merely a parasitic effect but the fundamental operating mechanism, as in tunnel junctions and resonant tunneling diodes.

Traditional simulation approaches are typically based on semiclassical physics and derived from the Boltzmann Transport Equation (BTE). Among these, the drift-diffusion (DD) model provides an efficient and widely used framework for device simulation. While DD is sufficiently accurate for many applications, it fails to properly describe transport when quantum effects become dominant.

A rigorous treatment of such phenomena requires a microscopic description of the system. In principle, this is provided by the formulation of the system Hamiltonian, from which the Schrödinger equation can be solved. However, in realistic systems involving many interacting particles, solving the Schrödinger equation becomes computationally prohibitive, and only approximate or perturbative approaches are generally feasible.

In this context, Green's functions theory provides a powerful alternative framework. Green's functions play a role analogous to that of the wavefunction, encoding the full microscopic information of the system while allowing for the systematic inclusion of interactions through

perturbation theory. In particular, single-particle Green's functions enable the calculation of key physical quantities such as carrier density and current, while many-body effects are incorporated through self-energies appearing in their equations of motion.
Several formulations of Green's functions theory exist depending on the physical regime of interest, including zero-temperature equilibrium theory [12], the finite-temperature Matsubara formalism [13], and the contour-ordered Green's function approach for non-equilibrium systems. The latter provides the foundation of the Non-Equilibrium Green's Functions (NEGF) formalism, which is particularly suited for the description of quantum transport in semiconductor devices.
The NEGF formalism was developed in the early 1960s through the work of Kadanoff and Baym [14] and Keldysh [15], building on earlier contributions by Martin and Schwinger [16]. Its importance has grown alongside the development of mesoscopic physics and advanced fabrication techniques such as molecular beam epitaxy, which enabled the realization of structures where quantum effects play a central role. For example, NEGF was applied to resonant tunneling diodes by Kim and Arnold [17], demonstrating its capability to describe quantum transport in realistic devices.

This chapter introduces the theoretical framework of the NEGF formalism, including the treatment of particle interactions and the principal approximations required to make the method computationally tractable for extended device geometries. For a more detailed and rigorous treatment, the reader is referred to [18–21].

2.1 Steady-state Green's functions

The concept of a Green's function originates from the theory of differential equations, where it represents the response of a system to a localized source. In analogy with the impulse response used in signal theory, it describes how a disturbance introduced at one space-time point propagates through the system and affects another point. In quantum theory, this concept is extended to describe the propagation and correlations of quantum fields. Green's functions are therefore expressed as expectation values of products of field creation and annihilation operators evaluated at different space-time coordinates. Their dependence on two points contains information about the propagation of excitations through the system and the occupation of the quantum states. The formulation considered in this work is based on one-particle Green's functions, in which the effects of interactions and of the coupling to the external environment are incorporated through appropriate self-energy terms.
For the electronic systems considered in this work, the relevant Green's functions are the retarded,

advanced, lesser, and greater Green's functions. In terms of the fermionic field operators $\hat{\Psi}(\mathbf{r}, t)$ and $\hat{\Psi}^\dagger(\mathbf{r}, t)$, they are defined as [18, 20, 21]

$$G^R(\mathbf{r}, t; \mathbf{r}', t') = -\frac{i}{\hbar}\theta(t - t') \left\langle \left\{ \hat{\Psi}(\mathbf{r}, t), \hat{\Psi}^\dagger(\mathbf{r}', t') \right\} \right\rangle , \tag{2.1a}$$

$$G^A(\mathbf{r}, t; \mathbf{r}', t') = +\frac{i}{\hbar}\theta(t' - t) \left\langle \left\{ \hat{\Psi}(\mathbf{r}, t), \hat{\Psi}^\dagger(\mathbf{r}', t') \right\} \right\rangle , \tag{2.1b}$$

$$G^<(\mathbf{r}, t; \mathbf{r}', t') = +\frac{i}{\hbar} \left\langle \hat{\Psi}^\dagger(\mathbf{r}', t') \hat{\Psi}(\mathbf{r}, t) \right\rangle , \tag{2.1c}$$

$$G^>(\mathbf{r}, t; \mathbf{r}', t') = -\frac{i}{\hbar} \left\langle \hat{\Psi}(\mathbf{r}, t) \hat{\Psi}^\dagger(\mathbf{r}', t') \right\rangle . \tag{2.1d}$$

where $\{\cdot, \cdot\}$ denotes the fermionic anticommutator and $\langle \cdot \rangle$ represents the statistical expectation value of the many-particle system.

The retarded Green's function describes the propagation of a single-particle excitation from $(\mathbf{r}', t')$ to $(\mathbf{r}, t)$ and determines the spectral properties of the system. The Heaviside function enforces causality, since the response at time t can only depend on perturbations occurring at earlier times. Consequently, G^R vanishes for $t < t'$ and represents the causal response of the system. The advanced Green's function provides the complementary anti-causal response: it vanishes for $t > t'$ and satisfies

$$G^A(\mathbf{r}, t; \mathbf{r}', t') = \left[G^R(\mathbf{r}', t'; \mathbf{r}, t) \right]^\dagger . \tag{2.2}$$

The lesser and greater Green's functions are correlation functions of field operators evaluated at different space-time points. They differ in the ordering of the creation and annihilation operators and provide complementary information about the statistical state of the system. The four Green's functions are not independent. From their definitions one obtains

$$G^R - G^A = G^> - G^<. \tag{2.3}$$

In the steady-state regime, the Hamiltonian and the statistical properties of the system are invariant under a common translation of the two time arguments. Consequently, the Green's functions depend only on the time difference

$$\tau = t - t', \tag{2.4}$$

so that

$$G(\mathbf{r}, t; \mathbf{r}', t') = G(\mathbf{r}, \mathbf{r}'; \tau). \tag{2.5}$$

The Green's functions can therefore be transformed from the time domain to the energy domain according to

$$G(\mathbf{r}, \mathbf{r}'; E) = \int_{-\infty}^{+\infty} d\tau \, e^{iE\tau/\hbar} G(\mathbf{r}, \mathbf{r}'; \tau), \tag{2.6}$$

with inverse transformation

$$G(\mathbf{r}, \mathbf{r}'; \tau) = \frac{1}{2\pi\hbar} \int_{-\infty}^{+\infty} dE \, e^{-iE\tau/\hbar} G(\mathbf{r}, \mathbf{r}'; E). \tag{2.7}$$

Under these conventions, the energy-domain Green's functions have the dimensions of an inverse energy, and equal-time correlation functions are obtained by integrating over energy with the measure $dE/(2\pi)$. In the energy representation, the relation between the retarded and advanced Green's functions becomes

$$G^A(E) = \left[G^R(E)\right]^\dagger. \tag{2.8}$$

In the energy representation, it is possible to define the spectral function as

$$A(E) = i\left[G^R(E) - G^A(E)\right] = i\left[G^>(E) - G^<(E)\right]. \tag{2.9}$$

This quantity describes the energy distribution of the available single-particle states. For an isolated system without coupling to external reservoirs or interactions, a discrete eigenstate gives a delta-like contribution at its eigenenergy. Coupling to the contacts or to other degrees of freedom gives the state a finite lifetime and therefore a finite spectral linewidth, so that the corresponding contribution to $A(E)$ is broadened.
At thermodynamic equilibrium, the lesser and greater Green's functions are related to the spectral function through the Fermi-Dirac distribution,

$$G^<(E) = i f_{\mathrm{FD}}(E) A(E), \tag{2.10a}$$

$$G^>(E) = -i\left[1 - f_{\mathrm{FD}}(E)\right] A(E). \tag{2.10b}$$

These expressions show explicitly that the spectral function determines the available states, while the Fermi-Dirac distribution determines their occupation. In a device driven out of equilibrium, however, the electronic population cannot generally be described by a single global Fermi-Dirac distribution. The lesser and greater Green's functions must instead be determined from the carrier injection and scattering processes occurring within the system.

For the layered heterostructures considered in this work, the material parameters and electrostatic potential vary only along the growth direction z. The position vector can therefore

be separated as $\mathbf{r} = (\boldsymbol{\rho}, z)$, where $\boldsymbol{\rho} = (x, y)$ is the transverse coordinate. Since translational invariance is assumed in the transverse plane, the Green's functions depend on the transverse coordinates only through $\boldsymbol{\rho} - \boldsymbol{\rho}'$. A Fourier transform with respect to this transverse coordinate difference gives the mixed representation

$$G(\mathbf{r}, \mathbf{r}'; E) = \int \frac{d^2\mathbf{k}}{(2\pi)^2} e^{i\mathbf{k}\cdot(\rho-\rho')} G(z, z'; \mathbf{k}, E), \tag{2.11}$$

where $\mathbf{k} = (k_x, k_y)$ is the transverse wave vector, perpendicular to the growth direction. In this representation, each transverse wave vector defines an independent one-dimensional quantum transport problem along z, coupled to the other transverse modes only through interaction processes or through macroscopic quantities obtained after transverse integration. Unless otherwise specified, $\mathbf{k}$ denotes the transverse wave vector in the rest of this chapter.
The equations governing the spectral properties and occupations of the nonequilibrium system are the Dyson and Keldysh equations, respectively, which are introduced in the following section.

2.2 Dyson and Keldysh equations

To solve the transport problem, the different components of the Green's function must be determined. At each transverse wave vector, the retarded component satisfies the Dyson equation, which represents its equation of motion in the presence of coupling to external contacts and interactions with the environment,

$$\begin{aligned} G^R(z, z'; \mathbf{k}, E) = {} & G_0^R(z, z'; \mathbf{k}, E) \\ & + \int dz_1 \int dz_2\, G_0^R(z, z_1; \mathbf{k}, E) \Sigma^R(z_1, z_2; \mathbf{k}, E) G^R(z_2, z'; \mathbf{k}, E), \end{aligned} \tag{2.12}$$

where G_0^R is the retarded Green's function of the reference system and Σ^R is the retarded self-energy, which accounts for the coupling to the contacts and for the interactions occurring within the system. In compact operator notation, Eq. (2.12) becomes

$$G^R(\mathbf{k}, E) = G_0^R(\mathbf{k}, E) + G_0^R(\mathbf{k}, E) \Sigma^R(\mathbf{k}, E) G^R(\mathbf{k}, E). \tag{2.13}$$

For a system described by the non-interacting band-structure Hamiltonian $H_0(\mathbf{k})$ and the electrostatic potential $U(z)$, where H_0 contains the kinetic and crystal-periodic contributions

before the device electrostatics is added, one obtains

$$G_0^R(\mathbf{k}, E) = \left[(E + i\eta)\mathbb{1} - H_0(\mathbf{k}) - U(z)\right]^{-1}. \tag{2.14}$$

Here, $\eta \to 0^+$ is a positive infinitesimal that enforces the retarded, i.e. causal, boundary condition. Therefore,

$$G^R(\mathbf{k}, E) = \left[(E + i\eta)\mathbb{1} - H_0(\mathbf{k}) - U(z) - \Sigma^R(\mathbf{k}, E)\right]^{-1}. \tag{2.15}$$

The advanced Green's function is related to the retarded one through

$$G^A(\mathbf{k}, E) = \left[G^R(\mathbf{k}, E)\right]^\dagger. \tag{2.16}$$

Meanwhile, for a steady-state open system, the lesser and greater Green's functions need to satisfy the Keldysh equation

$$\begin{aligned} G^{<,>}(z, z'; \mathbf{k}, E) = &\int dz_1 \int dz_2\, G^R(z, z_1; \mathbf{k}, E) \\ &\times \Sigma^{<,>}(z_1, z_2; \mathbf{k}, E) G^A(z_2, z'; \mathbf{k}, E), \end{aligned} \tag{2.17}$$

where $\Sigma^<(E)$ and $\Sigma^>(E)$ are the lesser and greater self-energies, respectively. They account for the correlations introduced through the coupling to the contacts and through the interactions occurring within the system. In compact operator notation, Eq. (2.17) can be rewritten as:

$$G^{<,>}(\mathbf{k}, E) = G^R(\mathbf{k}, E)\Sigma^{<,>}(\mathbf{k}, E)G^A(\mathbf{k}, E). \tag{2.18}$$

Equations (2.15) and (2.18) constitute the central equations of the steady-state NEGF formalism. A more detailed derivation can be found in Refs. [18, 20, 21].

2.3 General expression for physical observables

In second quantization, a local one-body operator in the layered representation can be written as

$$\hat{O} = \int dz\, \hat{\Psi}^\dagger(z, t)\, O(z, t)\, \hat{\Psi}(z, t). \tag{2.19}$$

Using the definition of the lesser Green's function,

$$G^<(z, t; z', t') = \pm\frac{i}{\hbar}\langle\hat{\Psi}^\dagger(z', t')\hat{\Psi}(z, t)\rangle, \tag{2.20}$$

where the upper sign refers to fermions and the lower to bosons. The expectation value becomes

$$\langle \hat{O}(z,t) \rangle = (\mp i\hbar) \lim_{z' \to z} \lim_{t' \to t^+} O(z,t)\, G^<(z,t;z',t'). \tag{2.21}$$

This relation provides the starting point for the derivation of the main physical observables.

2.3.1 Carrier density

The electron density operator is

$$\hat{n}(z,t) = \hat{\Psi}^\dagger(z,t)\hat{\Psi}(z,t). \tag{2.22}$$

From the general expression above, the carrier density reads

$$n(z,t) = (-i\hbar) \lim_{z' \to z} \lim_{t' \to t^+} G^<(z,t;z',t'). \tag{2.23}$$

Under steady-state conditions, time translational invariance allows Fourier transformation with respect to the time difference. Including the transverse modes, one obtains

$$n(z) = -i \int \frac{dE}{2\pi} \int \frac{d^2\mathbf{k}}{(2\pi)^2}\, G^<(z,z;\mathbf{k},E). \tag{2.24}$$

Similarly, the hole density can be expressed in terms of the greater Green's function as

$$p(z) = +i \int \frac{dE}{2\pi} \int \frac{d^2\mathbf{k}}{(2\pi)^2}\, G^>(z,z;\mathbf{k},E). \tag{2.25}$$

2.3.2 Local density of states

At thermal equilibrium, the lesser Green's function can be expressed in terms of the spectral function $A(E)$ as

$$G^<(E) = i f(E) A(E), \tag{2.26}$$

where $f(E)$ is the Fermi-Dirac distribution and

$$A(E) = i\left[G^R(E) - G^A(E)\right] = i\left[G^>(E) - G^<(E)\right]. \tag{2.27}$$

Substituting into the density expression yields

$$n(z) = \int dE\, f(E) \left[\int \frac{d^2\mathbf{k}}{(2\pi)^2} \frac{1}{2\pi} A(z, z; \mathbf{k}, E) \right]. \tag{2.28}$$

By comparison with the definition

$$n(z) = \int dE\, f(E)\, \mathrm{LDOS}(z, E), \tag{2.29}$$

the local density of states can be identified as

$$\mathrm{LDOS}(z, E) = \int \frac{d^2\mathbf{k}}{(2\pi)^2} \frac{1}{2\pi} A(z, z; \mathbf{k}, E). \tag{2.30}$$

2.3.3 Current density and divergence

The steady-state particle current density along the transport direction can be written in terms of the lesser Green's function as [20]

$$J(z) = -\frac{i}{2} \lim_{z' \to z} [v_z(z) - v_z(z')] \int \frac{dE}{2\pi} \int \frac{d^2\mathbf{k}}{(2\pi)^2} G^<(z, z'; \mathbf{k}, E), \tag{2.31}$$

where

$$v_z = -\frac{i}{\hbar}[z, H_0] \tag{2.32}$$

is the velocity operator associated with the non-interacting Hamiltonian H_0 along the growth direction. Applying the derivative with respect to z to Eq. (2.31), the current divergence can be expressed as

$$\frac{\partial J(z)}{\partial z} = \int \frac{dE}{2\pi\hbar} \int \frac{d^2\mathbf{k}}{(2\pi)^2} \left[H_0 G^<(\mathbf{k}, E) - G^<(\mathbf{k}, E) H_0 \right]_{z,z}. \tag{2.33}$$

In Eq. (2.33), H_0 acts on the first spatial argument in the first term and on the second spatial argument in the second term. The products are therefore operator products and imply integration over intermediate z coordinates. Using the Dyson and Keldysh equations, the same quantity can

be written in terms of the self-energies as

$$\begin{aligned}\frac{\partial J(z)}{\partial z} = \int \frac{dE}{2\pi\hbar} \int \frac{d^2\mathbf{k}}{(2\pi)^2} \int dz' \Big[& \Sigma^R(z,z';\mathbf{k},E) G^<(z',z;\mathbf{k},E) \\ & + \Sigma^<(z,z';\mathbf{k},E) G^A(z',z;\mathbf{k},E) \\ & - G^R(z,z';\mathbf{k},E) \Sigma^<(z',z;\mathbf{k},E) \\ & - G^<(z,z';\mathbf{k},E) \Sigma^A(z',z;\mathbf{k},E) \Big]. \end{aligned} \tag{2.34}$$

Since the last two contributions are related to the first two through the symmetry properties of the Green's functions and self-energies, Eq. (2.34) can equivalently be written as

$$\begin{aligned}\frac{\partial J(z)}{\partial z} = \frac{2}{\hbar} \mathrm{Re} \int \frac{dE}{2\pi} \int \frac{d^2\mathbf{k}}{(2\pi)^2} \int dz' \Big[& \Sigma^R(z,z';\mathbf{k},E) G^<(z',z;\mathbf{k},E) \\ & - G^R(z,z';\mathbf{k},E) \Sigma^<(z',z;\mathbf{k},E) \Big]. \end{aligned} \tag{2.35}$$

The contribution associated with a specific interaction mechanism can be obtained by retaining only the corresponding self-energy in Eq. (2.34). This defines the local net generation-recombination rate associated with that interaction. In particular, when Σ describes transitions mediated by a defect state, the resulting term represents the defect-assisted recombination rate used in the following sections.

2.4 Spatial discretization

To numerically solve the NEGF problem for a layered device, the field operators are expanded in a suitable basis of single-particle states along the z direction. The corresponding single-particle wave functions $\phi_m(z,\sigma) = \langle z,\sigma|m\rangle$ act as expansion coefficients,

$$\hat{\Psi}(z,t,\sigma) = \sum_m \phi_m(z,\sigma)\, \hat{c}_m(t), \tag{2.36a}$$

$$\hat{\Psi}^\dagger(z,t,\sigma) = \sum_m \phi_m^*(z,\sigma)\, \hat{c}_m^\dagger(t). \tag{2.36b}$$

The Green's function can consequently be represented as

$$G(z,t;z',t') = \sum_{mn} \phi_m(z)\, G_{mn}(t,t')\, \phi_n^*(z'), \tag{2.37a}$$

$$G_{mn}(t,t') = \int dz \int dz'\, \phi_m^*(z)\, G(z,t;z',t')\, \phi_n(z'). \tag{2.37b}$$

The coefficients $G_{mn}(t,t')$ therefore form a matrix-valued function in the chosen basis. Under steady-state conditions, the corresponding energy-domain representation is

$$G(z,z';E) = \sum_{mn} \phi_m(z)\, G_{mn}(E)\, \phi_n^*(z'). \tag{2.38}$$

For the following derivation, the basis functions are assumed to be orthonormal,

$$\int dz\, \phi_m^*(z)\phi_n(z) = \delta_{mn}. \tag{2.39}$$

Applying the expansion in Eq. (2.38) to the steady-state Dyson equation, Eq. (2.12), gives

$$\begin{aligned}\sum_{mn} \phi_m(z_1) G^R_{mn}(E) \phi_n^*(z_{1'}) = &\sum_{mn} \phi_m(z_1) \left[G_0^R\right]_{mn}(E) \phi_n^*(z_{1'}) \\ &+ \int dz_2 \int dz_3 \sum_{\substack{m_1,n_1\\ m_2,n_2\\ m_3,n_3}} \phi_{m_1}(z_1) \left[G_0^R\right]_{m_1 n_1}(E) \phi_{n_1}^*(z_2) \\ &\times \phi_{m_2}(z_2) \Sigma^R_{m_2 n_2}(E) \phi_{n_2}^*(z_3) \phi_{m_3}(z_3) G^R_{m_3 n_3}(E) \phi_{n_3}^*(z_{1'}).\end{aligned} \tag{2.40}$$

Using the orthonormality relations

$$\int dz_2\, \phi_{n_1}^*(z_2)\phi_{m_2}(z_2) = \delta_{n_1 m_2}, \tag{2.41a}$$

$$\int dz_3\, \phi_{n_2}^*(z_3)\phi_{m_3}(z_3) = \delta_{n_2 m_3}, \tag{2.41b}$$

the convolution over the intermediate spatial coordinates reduces to a matrix product. Equation (2.40) therefore becomes

$$\sum_{mn} \phi_m(z_1) G^R_{mn}(E) \phi_n^*(z_{1'}) = \sum_{mn} \phi_m(z_1) \left[G_0^R(E) + G_0^R(E)\Sigma^R(E)G^R(E)\right]_{mn} \phi_n^*(z_{1'}). \tag{2.42}$$

Since this equality must hold for every pair of spatial coordinates, the matrix coefficients satisfy

$$G^R(E) = G_0^R(E) + G_0^R(E)\Sigma^R(E)G^R(E). \tag{2.43}$$

The matrix Dyson equation can equivalently be written as

$$\left[\left\{G_0^R(E)\right\}^{-1} - \Sigma^R(E)\right] G^R(E) = \mathbb{1}. \tag{2.44}$$

The matrix elements of the inverse non-interacting Green's function are

$$\begin{aligned}
\left[G_0^R\right]_{mn}^{-1}(E) &= \int dz \int dz'\, \phi_m^*(z) \left[G_0^R\right]^{-1}(z, z'; E)\phi_n(z') \\
&= \int dz \int dz'\, \phi_m^*(z) \left[E + i\eta - H_0(z) - U(z)\right] \delta(z - z')\phi_n(z') \\
&= \int dz\, \phi_m^*(z) \left[E + i\eta - H_0(z) - U(z)\right] \phi_n(z) \\
&= (E + i\eta)\delta_{mn} - H_{mn},
\end{aligned} \tag{2.45}$$

where

$$H_{mn} = \int dz\, \phi_m^*(z) \left[H_0(z) + U(z)\right] \phi_n(z) \tag{2.46}$$

is the Hamiltonian matrix in the chosen basis. Consequently, Eq. (2.44) becomes

$$\left[(E + i\eta)\mathbb{1} - H - \Sigma^R(E)\right] G^R(E) = \mathbb{1}, \tag{2.47}$$

and the retarded Green's function is

$$G^R(E) = \left[(E + i\eta)\mathbb{1} - H - \Sigma^R(E)\right]^{-1}. \tag{2.48}$$

The same projection can be applied to the Keldysh equation. In the chosen basis,

$$G^{<,>}(z_1, z_{1'}; E) = \sum_{mn} \phi_m(z_1) \left[G^R(E)\Sigma^{<,>}(E)G^A(E)\right]_{mn} \phi_n^*(z_{1'}), \tag{2.49}$$

and therefore

$$G^{<,>}(E) = G^R(E)\Sigma^{<,>}(E)G^A(E). \tag{2.50}$$

For a non-orthogonal basis, the overlap matrix is defined as

$$M_{mn} = \int dz\, \phi_m^*(z)\phi_n(z). \tag{2.51}$$

The matrix Dyson equation is then generalized to

$$\left[(E + i\eta)M - H - \Sigma^R(E)\right] G^R(E) = \mathbb{1}. \tag{2.52}$$

For an orthonormal basis, $M_{mn} = \delta_{mn}$ and the overlap matrix reduces to the identity, recovering the expressions derived above.

2.5 Multiband $\mathbf{k} \cdot \mathbf{p}$ band-structure model

The NEGF formalism does not prescribe a specific description of the electronic structure and can be combined with Hamiltonians of different levels of complexity. In the most complete quantum-transport description considered in this chapter, the device Hamiltonian is constructed using a multiband $\mathbf{k} \cdot \mathbf{p}$ model. This Hamiltonian provides a detailed representation of the electronic structure close to the zone center while retaining the coupling between the relevant conduction and valence bands.
The effective mass approximation, used for the reduced Büttiker-probe implementation, is introduced separately in Section 2.9.

The $\mathbf{k} \cdot \mathbf{p}$ method is a continuum description of the electronic structure based on an expansion of the crystal wave functions around a selected point of the Brillouin zone. By retaining several zone-center states in the expansion, the resulting Hamiltonian explicitly accounts for the coupling and mixing between the corresponding energy bands. In the following bulk derivation the full three-dimensional crystal wave vector is denoted by $\boldsymbol{\kappa}$. After the layered geometry is introduced, it is decomposed into the transverse component $\mathbf{k}$ and the longitudinal component k_z, with the latter replaced by a differential operator along the growth direction. Let the Hamiltonian of an electron in a periodic crystal potential be

$$H_{\mathrm{cr}} = \frac{\mathbf{p}^2}{2m_0} + V_{\mathrm{cr}}(\mathbf{r}), \tag{2.53}$$

where $V_{\mathrm{cr}}(\mathbf{r})$ has the periodicity of the crystal lattice. According to Bloch's theorem, the corresponding eigenstates can be written as

$$\psi_{n\boldsymbol{\kappa}}(\mathbf{r}) = e^{i\boldsymbol{\kappa}\cdot\mathbf{r}} u_{n\boldsymbol{\kappa}}(\mathbf{r}), \tag{2.54}$$

where $u_{n\boldsymbol{\kappa}}(\mathbf{r})$ is lattice periodic. Substituting Eq. (2.54) into the Schrödinger equation gives

$$H(\boldsymbol{\kappa})u_{n\boldsymbol{\kappa}} = E_n(\boldsymbol{\kappa})u_{n\boldsymbol{\kappa}}, \tag{2.55}$$

with

$$H(\boldsymbol{\kappa}) = H(\mathbf{0}) + \frac{\hbar}{m_0}\boldsymbol{\kappa} \cdot \mathbf{p} + \frac{\hbar^2\kappa^2}{2m_0}. \tag{2.56}$$

Equation (2.56) shows that the dependence of the electronic structure on $\boldsymbol{\kappa}$ is determined by the momentum operator. To obtain a finite-dimensional representation, the periodic part of the Bloch function is expanded in the zone-center eigenstates,

$$u_{n\boldsymbol{\kappa}}(\mathbf{r}) = \sum_m c_{mn}(\boldsymbol{\kappa})u_{m0}(\mathbf{r}), \tag{2.57}$$

where

$$H(\mathbf{0})u_{m0} = E_m(\mathbf{0})u_{m0}. \tag{2.58}$$

Projecting the Schrödinger equation onto the basis state u_{l0} yields

$$\begin{aligned}\sum_m \left[E_l(\mathbf{0})\delta_{lm} + \frac{\hbar}{m_0}\boldsymbol{\kappa} \cdot \mathbf{p}_{lm} + \frac{\hbar^2\kappa^2}{2m_0}\delta_{lm}\right] c_{mn}(\boldsymbol{\kappa}) \\ = E_n(\boldsymbol{\kappa})c_{ln}(\boldsymbol{\kappa}),\end{aligned} \tag{2.59}$$

where

$$\mathbf{p}_{lm} = \langle u_{l0}|\mathbf{p}|u_{m0}\rangle \tag{2.60}$$

is the momentum matrix element between the zone-center Bloch states. The diagonal matrix elements in Eq. (2.59) describe the uncoupled energy bands, while the off-diagonal momentum matrix elements couple states belonging to different bands. These terms introduce band mixing and produce the non-parabolic dispersion that is neglected in the effective mass approximation. In practice, only a finite number of bands can be retained explicitly. The zone-center states are therefore separated into an active subspace A, containing the bands included in the model, and a remote subspace B, containing all the remaining bands. Eliminating the remote states gives the effective Hamiltonian

$$H_{\text{eff}}(E) = H_{AA} + H_{AB}\left(E - H_{BB}\right)^{-1} H_{BA}. \tag{2.61}$$

The contribution of the remote bands is commonly evaluated perturbatively around the band edges and absorbed into renormalized effective-mass and Luttinger-like parameters. This procedure is

generally referred to as Löwdin renormalization.
For the III-nitride devices considered in this work, the crystal structure is wurtzite and the expansion is performed around the Γ point. The eight-band basis contains two spin-degenerate, predominantly s-like conduction-band states and six predominantly p-like valence-band states,

$$\{|S\uparrow\rangle, |X\uparrow\rangle, |Y\uparrow\rangle, |Z\uparrow\rangle, |S\downarrow\rangle, |X\downarrow\rangle, |Y\downarrow\rangle, |Z\downarrow\rangle\}\,. \tag{2.62}$$

The corresponding Hamiltonian has the block structure

$$H_{\mathbf{k}\cdot\mathbf{p}} = \begin{pmatrix} H_{cc} & H_{cv} \\ H_{vc} & H_{vv} \end{pmatrix}, \qquad H_{vc} = H_{cv}^{\dagger}, \tag{2.63}$$

where H_{cc} describes the first conduction band, H_{vv} describes the three upper valence bands including spin, and H_{cv} contains the conduction-valence coupling.
Because of the anisotropy of the wurtzite crystal of III-Nitrides, the coupling parallel to the crystal c axis differs from that in the transverse plane. The corresponding momentum matrix elements are conventionally represented by the parameters P_1 and P_2, respectively. The valence-band structure is further determined by the crystal-field splitting Δ_{cr}, the spin-orbit splitting Δ_{SO}, and the parameters $A_1, \ldots, A_7$ describing the anisotropic dispersion and mixing of the valence states. The resulting eight-band model includes the first conduction band and the three upper valence bands, conventionally denoted as the A, B, and C bands in wurtzite crystals. It therefore accounts for conduction-valence coupling, valence-band mixing, anisotropic effective masses, spin-orbit interaction, and crystal-field effects. These features are required for an accurate description of the valence-band local density of states and of direct interband tunneling in III-nitride heterostructures [22].
For the layered structures considered here, the material parameters and electrostatic potential vary along the growth direction z, while translational invariance is retained in the transverse plane. The transverse component $\mathbf{k}$ therefore remains a parameter, whereas the longitudinal component is replaced by the differential operator

$$k_z \longrightarrow -i\frac{\partial}{\partial z}. \tag{2.64}$$

The bulk Hamiltonian may be arranged according to its dependence on k_z as

$$H_{\mathbf{k}\cdot\mathbf{p}}(\mathbf{k}, k_z) = H_2(\mathbf{k})k_z^2 + H_1(\mathbf{k})k_z + H_0(\mathbf{k}). \tag{2.65}$$

After the substitution in Eq. (2.64), the heterostructure Hamiltonian becomes an envelope-function differential operator. Since the material parameters depend on position, they do not generally commute with $\partial/\partial z$. The differential operator must therefore be written using an appropriate operator ordering,

$$H_{\mathbf{k}\cdot\mathbf{p}}(\mathbf{k}) = -\frac{\partial}{\partial z}H_2(\mathbf{k},z)\frac{\partial}{\partial z} - iH_{1L}(\mathbf{k},z)\frac{\partial}{\partial z} - i\frac{\partial}{\partial z}H_{1R}(\mathbf{k},z) + H_0(\mathbf{k},z). \tag{2.66}$$

The matrices H_{1L} and H_{1R} define the ordering of the first-order terms at heterointerfaces. A consistent ordering is necessary to preserve the Hermitian character and numerical stability of the Hamiltonian and to avoid nonphysical spurious solutions [23, 22].
The complete wurtzite 8×8 $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian used in this work, together with the adopted basis, the definitions of its matrix elements, and the material parameters employed in the calculations, is reported in Appendix A. The appendix also provides the expressions used to derive the conduction-valence coupling parameters and the renormalized band parameters from tabulated material quantities.
The multiband Hamiltonian is required when conduction-valence coupling, valence-band mixing, anisotropy, or direct interband tunneling must be described explicitly. When transport is instead dominated by states close to a single band edge, the band-structure description can be reduced to the effective mass approximation introduced below.

2.6 Boundary self-energies

A quantum transport device is an open system connected to macroscopic contacts, which supply and absorb carriers. The complete structure can be conceptually divided into a finite device region and two semi-infinite reservoirs. Portions of the semiconductor in which the electrostatic potential varies or scattering must be explicitly resolved are included in the device region, whereas sufficiently far from the active region the contacts are assumed to be homogeneous and in thermal equilibrium. Their carrier populations are therefore described by Fermi-Dirac distributions characterized by the contact chemical potentials μ_L and μ_R.
An explicit treatment of the semi-infinite reservoirs would require Green's functions of infinite dimension. Within the NEGF formalism, their effect can instead be incorporated into the finite device problem through boundary self-energies [20, 19]. Considering first a device coupled to a

single reservoir, the Hamiltonian of the complete system can be partitioned as

$$H = \begin{pmatrix} H_D & \tau \\ \tau^\dagger & H_{\text{res}} \end{pmatrix}, \tag{2.67}$$

where H_D and H_{res} are the Hamiltonians of the isolated device and reservoir, respectively, while τ describes their coupling. The retarded Green's function projected onto the device region is obtained by eliminating the reservoir degrees of freedom,

$$G_D^R(E) = \left[(E + i\eta)\mathbb{1} - H_D - \Sigma_B^R(E)\right]^{-1}, \tag{2.68}$$

where

$$\Sigma_B^R(E) = \tau g_{\text{res}}^R(E)\tau^\dagger, \qquad g_{\text{res}}^R(E) = \left[(E + i\eta)\mathbb{1} - H_{\text{res}}\right]^{-1}. \tag{2.69}$$

Here, g_{res}^R is the retarded Green's function of the isolated semi-infinite reservoir. Equations (2.67)-(2.69) are written for an orthogonal spatial basis, as assumed in the following derivation.
For a device connected to left and right contacts, the retarded Green's function becomes

$$G_D^R(E) = \left[(E + i\eta)\mathbb{1} - H_D - \Sigma_L^R(E) - \Sigma_R^R(E) - \Sigma_{\text{int}}^R(E)\right]^{-1}, \tag{2.70}$$

where Σ_{int}^R contains the interaction self-energies discussed in the following section. The boundary self-energies Σ_L^R and Σ_R^R impose open boundary conditions at the two ends of the finite simulation domain. Their Hermitian components modify the energy of the device states, while their anti-Hermitian components account for the finite lifetime associated with carrier escape into the contacts.
For nearest-neighbour spatial discretizations, the contacts are coupled only to the first and last spatial nodes. Consequently, the left and right boundary self-energies are zero throughout the device matrix except for the first and last band blocks, respectively. Although Eq. (2.69) is completely general, its direct evaluation would still require the surface Green's function of the semi-infinite contact. For the multiband differential Hamiltonian considered here, the same quantity can be obtained by matching the envelope function at the device boundary to the outgoing modes of the corresponding bulk contact [19].
At fixed energy E and transverse wavevector $\mathbf{k}$, the contact modes are solutions of the bulk multiband equation

$$\left[H_0(\mathbf{k}) + H_1(\mathbf{k})k_z + H_2 k_z^2 - E\mathbb{1}\right]\chi = 0, \tag{2.71}$$

where $H_1 = H_{1L} + H_{1R}$ according to the Hamiltonian decomposition introduced in Section 2.5. The solutions generally include both propagating and evanescent states. Propagating modes

are classified according to the direction of their group velocity, whereas evanescent modes are classified according to the direction in which their amplitude decays. The retarded boundary condition retains only modes travelling or decaying away from the finite device region: left-going or left-decaying modes for the left contact and right-going or right-decaying modes for the right contact.

Collecting the selected contact eigenvectors into the matrices

$$\boldsymbol{\chi}^{\alpha} = \left[\boldsymbol{\chi}^{\alpha}_{1}, \ldots, \boldsymbol{\chi}^{\alpha}_{N_b}\right], \qquad \mathbf{K}^{\alpha}_{z} = \operatorname{diag}\left(k^{\alpha}_{z,1}, \ldots, k^{\alpha}_{z,N_b}\right), \qquad \alpha \in \{L, R\}, \tag{2.72}$$

the outgoing modes provide a relation between the envelope function and its normal derivative at the boundary. Using the convention of Ref. [19], this relation can be written as

$$\left.\frac{\partial \mathbf{F}^{\mathrm{out}}_{\mathbf{k}}}{\partial z}\right|_{z=z_\alpha} = -i\boldsymbol{\chi}^{\alpha}\mathbf{K}^{\alpha}_{z}\left[\boldsymbol{\chi}^{\alpha}\right]^{-1}\mathbf{F}^{\mathrm{out}}_{\mathbf{k}}(z_\alpha). \tag{2.73}$$

Equation (2.73) represents the response of the complete semi-infinite reservoir at its interface with the device. Substituting this relation into the boundary terms of the spatially discretized Hamiltonian gives the nonzero blocks of the retarded boundary self-energy,

$$\Sigma^{R}_{L,11}(\mathbf{k}, E) = -iH_2(z_1)\boldsymbol{\chi}^{L}\mathbf{K}^{L}_{z}\left[\boldsymbol{\chi}^{L}\right]^{-1} + iH_{1R}(\mathbf{k}, z_1), \tag{2.74a}$$

$$\Sigma^{R}_{R,NN}(\mathbf{k}, E) = -iH_2(z_N)\boldsymbol{\chi}^{R}\mathbf{K}^{R}_{z}\left[\boldsymbol{\chi}^{R}\right]^{-1} + iH_{1R}(\mathbf{k}, z_N). \tag{2.74b}$$

The signs appearing in Eq. (2.74) depend on the convention adopted for the Bloch factors and longitudinal wavevectors, while the physical requirement is always that the selected modes carry probability away from the device or decay inside the contacts.

The retarded boundary self-energy describes the spectral coupling between the device and the reservoirs, independently of their occupation. Within an energy range supporting propagating contact modes, its anti-Hermitian component is nonzero and describes irreversible carrier escape from the finite device region. In an energy range in which the contact supports only evanescent modes, the boundary self-energy instead describes the penetration of the wave function into the contact without carrying a propagating flux.

The contact broadening matrices are defined as

$$\Gamma_{\alpha}(\mathbf{k}, E) = i\left[\Sigma^{R}_{\alpha}(\mathbf{k}, E) - \Sigma^{A}_{\alpha}(\mathbf{k}, E)\right], \qquad \alpha \in \{L, R\}. \tag{2.75}$$

Since the semi-infinite contacts remain in thermal equilibrium, their lesser and greater self-energies follow directly from the equilibrium fluctuation-dissipation relations,

$$\Sigma_\alpha^<(\mathbf{k}, E) = i f_{\mathrm{FD}}\left(E - \mu_\alpha\right) \Gamma_\alpha(\mathbf{k}, E), \tag{2.76a}$$

$$\Sigma_\alpha^>(\mathbf{k}, E) = -i \left[1 - f_{\mathrm{FD}}\left(E - \mu_\alpha\right)\right] \Gamma_\alpha(\mathbf{k}, E). \tag{2.76b}$$

The retarded component therefore determines which contact states are available for coupling to the device and the rate at which carriers can escape, whereas the lesser and greater components determine the injection and extraction associated with the occupation of those states. Under applied bias, the different chemical potentials of the two reservoirs produce different injection distributions and drive the device out of equilibrium.

2.7 Interaction self-energies

The coherent evolution of the carriers is described by the single-particle Hamiltonian $H = H_0 + U$, where H_0 is the non-interacting Hamiltonian and U is the electrostatic mean-field potential. The carrier-induced contribution to U corresponds to the Hartree approximation, providing the lowest-order mean-field treatment of carrier-carrier interactions. Interactions that produce scattering, energy relaxation, or transitions between different electronic states are instead introduced through the corresponding interaction self-energies.

An exact evaluation of these self-energies would require the complete perturbative expansion in the interaction Hamiltonian and is not feasible for realistic device simulations. In the following, carrier-boson interactions are described by retaining the lowest non-vanishing contribution to the electronic self-energy. For an interaction that is linear in the bosonic field, such as the carrier-phonon or carrier-photon interaction, this contribution contains two interaction vertices and is therefore second order in the coupling matrix element. This approximation is commonly referred to as the Born approximation.

When the non-interacting electronic Green's function G_0 appearing in the self-energy is replaced by the full interacting Green's function G, the self-energy becomes a functional of the Green's function,

$$\Sigma_{\mathrm{int}} = \Sigma_{\mathrm{int}}[G]. \tag{2.77}$$

The Green's functions determine the interaction self-energies, while the self-energies enter the Dyson and Keldysh equations that determine the Green's functions. The resulting equations must therefore be solved self-consistently. This procedure is known as the self-consistent Born

approximation (SCBA). Within this approximation, the carrier-phonon interaction self-energies can be written as[1]

$$\Sigma^{<,>}(\mathbf{k}, E) = \sum_q M^{ep}(q)\big[n_q G^{<,>}(\mathbf{k}-\mathbf{q}_\perp, E \mp \hbar\omega_q) + (n_q+1)G^{<,>}(\mathbf{k}-\mathbf{q}_\perp, E \pm \hbar\omega_q)\big]M^{ep}(-q) \quad , \tag{2.78}$$

$$\begin{aligned}\Sigma^R(\mathbf{k}, E) = \sum_q M^{ep}(q)\Bigg[&(n_q+1)G^R(\mathbf{k}-\mathbf{q}_\perp, E-\hbar\omega_q) + n_q G^R(\mathbf{k}-\mathbf{q}_\perp, E+\hbar\omega_q) \\ &+ \frac{1}{2}(G^<(\mathbf{k}-\mathbf{q}_\perp, E-\hbar\omega_q) - G^<(\mathbf{k}-\mathbf{q}_\perp, E+\hbar\omega_q)) \\ &+ i\mathcal{P}\left\{\int \frac{dE'}{2\pi}\left(\frac{G^<(\mathbf{k}-\mathbf{q}_\perp, E-E')}{E'-\hbar\omega_q} - \frac{G^<(\mathbf{k}-\mathbf{q}_\perp, E-E')}{E'+\hbar\omega_q}\right)\right\}\Bigg]M^{ep}(-q) \\ &= \frac{1}{2}[\Sigma^>(\mathbf{k}, E) - \Sigma^<(\mathbf{k}, E)] + \frac{i}{2\pi}\mathcal{P}\left\{\int dE' \frac{\Sigma^>(E') - \Sigma^<(E')}{E-E'}\right\} \\ &\approx \frac{1}{2}[\Sigma^>(\mathbf{k}, E) - \Sigma^<(\mathbf{k}, E)] \quad ,\end{aligned} \tag{2.79}$$

where $\hbar\omega_q$ is the phonon energy, $M^{ep}(q)$ is the electron-phonon coupling matrix element and n_q is the phonon occupation number given by the Bose-Einstein distribution.
In the full expression of the retarded self-energy an additional principal value integral appears, contributing to the real part of the self-energy. In practice, this contribution is typically found to be very small (on the order of a few percent of the total real part) and can therefore be safely neglected in numerical implementations.

In the following, the bosonic populations are assumed not to be modified by their interaction with the carriers, so that the corresponding free equilibrium propagators can be used. Only bulk phonon modes at thermal equilibrium are considered, while confinement effects enter through the electronic states. Furthermore, the expressions are restricted to orthogonal basis sets.[2] For a generic multiband Hamiltonian, the transverse wave-vector dependence cannot in general be removed analytically. The Green's functions and self-energies must therefore be evaluated on a numerical transverse wave-vector grid, in addition to the energy grid. This transverse integration is one of the factors contributing to the high computational cost of the full NEGF treatment.

[1] For a detailed derivation see [20] App. E or [19] App. B

[2] The extension to non-orthogonal bases is discussed in [24].

2.7.1 Acoustic phonon scattering self-energy

The electron-phonon coupling matrix can be written as [25]

$$M^{ep}_{m,n}(q) = \frac{1}{\sqrt{V}} U_q e^{iq_z(z_m - z_n)}, \tag{2.80}$$

where U_q represents the interaction strength. For acoustic phonons, assuming a linear dispersion relation $\omega_q = qu_l$, the coupling strength is given by

$$|U_q|^2 = \frac{\hbar D_a^2}{2\rho u_l} q, \tag{2.81}$$

where u_l is the sound velocity, D_a is the acoustic deformation potential, $V = \mathcal{A} \times \mathcal{L}$ is the normalization volume and ρ is the mass density. Substituting this expression into Eq. (2.78) and assuming high-temperature conditions such that $n_q \approx k_B T/(\hbar\omega_q) \approx n_q + 1$, the acoustic phonon self-energy becomes

$$\Sigma^{<,>,R}_{ij} = \frac{D_a^2 k_B T}{V \rho u_l^2} \sum_{q_\perp} \left[G^{<,>,R}(q_\perp, E) \right]_{ij} \frac{1}{\Delta_{ij}}, \tag{2.82}$$

where

$$\frac{1}{\Delta_{ij}} = \sum_{q_z} e^{iq_z(z_i - z_j)} \approx \delta_{ij} \frac{\mathcal{L}}{a} \tag{2.83}$$

with a the lattice constant.

2.7.2 Polar optical phonon scattering

For polar optical phonons, the electron-phonon interaction is described by the Fröhlich coupling [25]

$$U_q = \sqrt{\frac{q^2 \hbar\omega_{LO}}{2} \left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_s} \right)} \frac{q}{q^2 + q_0^2}, \tag{2.84}$$

where ϵ_s and ϵ_∞ are the static and high-frequency dielectric constants, and ω_{LO} is the longitudinal optical phonon frequency. The parameter

$$q_0 = \sqrt{\frac{q^2 n}{\epsilon_s k_B T}} \tag{2.85}$$

represents the inverse Debye screening length. After inserting the coupling term into Eq. (2.78) and assuming isotropic transverse dispersion, the lesser and greater self-energies can be written as

$$\Sigma_{ij}^{<,>}(\mathbf{k}, E) = \frac{q^2 \hbar\omega_{LO}}{4\pi^2}\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_s}\right)\int dq\, q\, F(q, s_{ij}, k) \tag{2.86}$$

$$\times \left[n_{LO} G^{<,>}(q, E \mp \hbar\omega_{LO}) + (n_{LO}+1) G^{<,>}(q, E \pm \hbar\omega_{LO})\right]_{ij} \tag{2.87}$$

where $s_{ij} = z_i - z_j$. Here $k = |\mathbf{k}|$ is the magnitude of the initial transverse wave vector, while q is the magnitude of the final transverse wave vector after angular integration. The function $F(q, s_{ij}, k)$ contains the transverse-momentum dependence obtained after the angular integration of the screened Fröhlich interaction and can be written as [19, 20, 25]:

$$F(q, s_{ij}, k) = \int_0^{\pi/a_L} dq_z \cos(q_z s_{ij}) \left[\frac{1}{\sqrt{\left(k^2 + q^2 + q_z^2 + q_0^2\right)^2 - 4k^2q^2}} - \frac{q_0^2\left(k^2 + q^2 + q_z^2 + q_0^2\right)}{\left[\left(k^2 + q^2 + q_z^2 + q_0^2\right)^2 - 4k^2q^2\right]^{3/2}} \right], \tag{2.88}$$

where a_L is the lattice parameter. Only the cosine contribution of the complex exponential remains after the integration over q_z, since the sine contribution vanishes by symmetry.

2.7.3 Multi-phonon defect self-energies

Defect-assisted recombination can be described within the NEGF formalism through interaction self-energies associated with localized defect states. For a deep localized defect, the energy separation between an extended band state and the defect level is generally much larger than the energy of a single phonon. The capture or emission process must therefore be accompanied by the absorption or emission of several phonons and can be treated within the multi-phonon relaxation framework [26]. In NEGF, this mechanism is introduced through defect self-energies coupling the localized defect state to the extended conduction- and valence-band states [27, 28, 19].
Assuming a single effective phonon mode of energy $\hbar\omega_0$, the lesser and greater self-energies

acting on the defect state can be written as

$$\Sigma^{<}_{dc}(E) = \sum_{\mathbf{k}} \sum_{l} M_{dc}(l) G^{<}_{c}(\mathbf{k}, E + l\hbar\omega_0)\,, \tag{2.89a}$$

$$\Sigma^{>}_{dc}(E) = \sum_{\mathbf{k}} \sum_{l} M_{dc}(l) e^{-\beta l\hbar\omega_0} G^{>}_{c}(\mathbf{k}, E + l\hbar\omega_0)\,, \tag{2.89b}$$

$$\Sigma^{<}_{dv}(E) = \sum_{\mathbf{k}} \sum_{l} M_{dv}(l) e^{-\beta l\hbar\omega_0} G^{<}_{v}(\mathbf{k}, E - l\hbar\omega_0)\,, \tag{2.89c}$$

$$\Sigma^{>}_{dv}(E) = \sum_{\mathbf{k}} \sum_{l} M_{dv}(l) G^{>}_{v}(\mathbf{k}, E - l\hbar\omega_0)\,. \tag{2.89d}$$

where $\beta = 1/(k_B T)$. The terms $\Sigma^{\lessgtr}_{dc}$ describe the interaction between the defect and conduction-band states, while $\Sigma^{\lessgtr}_{dv}$ describe the corresponding interaction between the defect and valence-band states. The exponential factors account for detailed balance between phonon-assisted capture and emission processes.
The reciprocal self-energies acting on the extended band states are

$$\Sigma^{<}_{cd}(\mathbf{k}, E) \simeq \sum_{l} M_{dc}(l) e^{-\beta l\hbar\omega_0} G^{<}_{d}(E - l\hbar\omega_0)\,, \tag{2.90a}$$

$$\Sigma^{>}_{cd}(\mathbf{k}, E) = \sum_{l} M_{dc}(l) G^{>}_{d}(E - l\hbar\omega_0)\,, \tag{2.90b}$$

$$\Sigma^{<}_{vd}(\mathbf{k}, E) \simeq \sum_{l} M_{dv}(l) G^{<}_{d}(E + l\hbar\omega_0)\,, \tag{2.90c}$$

$$\Sigma^{>}_{vd}(\mathbf{k}, E) = \sum_{l} M_{dv}(l) e^{-\beta l\hbar\omega_0} G^{>}_{d}(E + l\hbar\omega_0)\,. \tag{2.90d}$$

Equations (2.89)-(2.90) constitute the microscopic self-energy description of carrier capture and emission through a localized defect. They have been written in band-resolved notation for clarity. When a multiband $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian is used, G_c and G_v represent the conduction-like and valence-like projections of the multiband Green's function, and the corresponding self-energies become matrices in the adopted band basis. The same physical formulation can therefore be used with either an effective-mass or a multiband description of the electronic structure.
The multi-phonon coupling elements entering Eqs. (2.89)-(2.90) are

$$M_{dc}(l) = M^{0}_{dc} \frac{(S-l)^2}{S} L(l), \tag{2.91a}$$

$$M_{dv}(l) = M^{0}_{dv} \frac{(S+l)^2}{S} L(l), \tag{2.91b}$$

where S is the Huang-Rhys factor and $L(l)$ is the multi-phonon lineshape function,

$$L(l) = e^{-S(2n_0+1)} \left(\frac{n_0+1}{n_0}\right)^{l/2} I_l\left(2S\sqrt{n_0(n_0+1)}\right), \tag{2.92}$$

with

$$n_0 = \frac{1}{\exp(\beta\hbar\omega_0) - 1}. \tag{2.93}$$

Here, I_l is the modified Bessel function of order l, n_0 is the equilibrium occupation of the effective phonon mode, and M^0_{dc} and M^0_{dv} contain the remaining microscopic information on the coupling of the localized defect with conduction- and valence-band states.

In a complete NEGF treatment, the defect self-energies must be solved self-consistently with the Dyson and Keldysh equations. The conduction- and valence-band Green's functions determine the defect self-energies $\Sigma^{\lessgtr}_{dc}$ and $\Sigma^{\lessgtr}_{dv}$, while the defect Green's function determines the reciprocal self-energies $\Sigma^{\lessgtr}_{cd}$ and $\Sigma^{\lessgtr}_{vd}$ acting on the extended states. These self-energies are then inserted again into the Dyson and Keldysh equations, modifying both the carrier occupations and the spectral properties of the device. In particular, their retarded components introduce an additional renormalization and broadening of the local density of states. The mutual dependence between band Green's functions, defect Green's functions, and defect self-energies introduces an additional self-consistency loop and substantially increases the computational cost.

This self-consistent structure is shown schematically in Fig. 2.1.

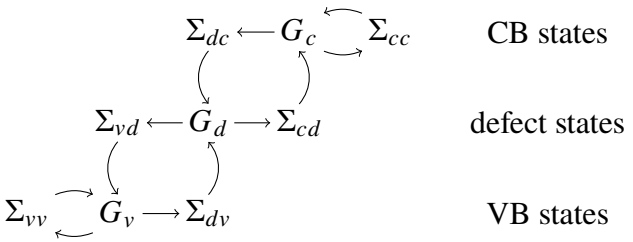


Fig. 2.1 Schematic self-consistent iteration scheme of Green's functions and defect-induced self-energies for nonradiative recombination through localized defect states. The scheme is shown in a band-resolved representation and is most directly applicable to an effective-mass description.

2.8 Self-consistent solution of the full NEGF scheme

The quantum transport problem is solved using the Non-Equilibrium Green's Function (NEGF) formalism self-consistently coupled with the electrostatic Poisson equation. The resulting system of equations consists of the Dyson equation for the retarded Green's function, the Keldysh equation

for the lesser Green's function, and the Poisson equation determining the electrostatic potential. Since the carrier density obtained from the Green's functions determines the electrostatic potential, which in turn modifies the device Hamiltonian entering the Green's functions, the problem is intrinsically nonlinear and must be solved iteratively. The overall computational procedure is illustrated in Fig. 2.2.

The calculation begins with an initial guess for the electrostatic potential $\phi(z)$. In practice, this initial profile is often obtained from a semiclassical drift-diffusion calculation or from an equilibrium Poisson solution. Once the potential is specified, the single-particle Hamiltonian for the device region can be constructed using the Hamiltonian of choice, following the formulation described in Section 2.5.

The device is treated as an open quantum system connected to semi-infinite contacts, whose effect is incorporated through the boundary self-energies $\Sigma_{L,R}^{R,A,<,>}(E)$, which describe carrier injection and extraction and drive the system out of equilibrium. For a given electrostatic potential, the retarded Green's function is obtained by solving the Dyson equation

$$G^R(E) = \left[E\mathbb{1} - H_D - \Sigma^R(E)\right]^{-1}, \tag{2.94}$$

where the total retarded self-energy contains both contact and particle interaction contributions. The latter are the full carrier-phonon interaction self-energies described in Section 2.7. Within this scheme, the interaction self-energies depend on both G^R and $G^{<,>}$, which in turn depend on the self-energies themselves. This mutual dependence makes the system nonlinear and requires a self-consistent iteration scheme to obtain a solution.

As discussed in Section 2.4, in practical numerical implementations the Hamiltonian and Green's functions must be discretized, using either finite-difference or finite-element methods, and represented as matrices whose dimension corresponds to the number of spatial discretization points in the device region. Solving the Dyson equation therefore requires the inversion of the matrix $E\mathbb{1} - H_D - \Sigma^R$ for every energy point considered in the integration grid. For realistic device sizes this matrix can become very large, making its inversion one of the most computationally demanding steps of the NEGF algorithm. In particular, if the device is discretized using N spatial points, direct inversion methods scale approximately as $O(N^3)$, which can significantly increase the computational cost for large systems or fine spatial grids. Once the retarded Green's function and the self-energies are available, the lesser Green's function is determined through the Keldysh equation

$$G^{<,>}(E) = G^R(E)\,\Sigma^{<,>}(E)\,G^A(E), \tag{2.95}$$

where the self-energies $\Sigma^{<,>}$ also depend on $G^{<,>}$, making a self-consistent loop necessary for the solution of the Keldysh equation as well. Moreover, the evaluation of the Keldysh equation involves the multiplication of three matrices of dimension $N \times N$ for each energy point. Although this operation is generally less expensive than matrix inversion, it still represents a considerable computational burden when performed over a large number of energy points and repeated throughout many iterations of the self-consistent loop. Consequently, the Dyson and Keldysh equations together typically dominate the overall runtime of NEGF simulations. After obtaining the lesser Green's function, the electron density inside the device can be calculated from

$$n(z) = -i \int \frac{dE}{2\pi} \int \frac{d^2\mathbf{k}}{(2\pi)^2} G^<(z, z; \mathbf{k}, E). \tag{2.96}$$

The resulting carrier density is then used as input to Poisson's equation

$$\frac{\partial}{\partial z}\left[\varepsilon(z)\frac{\partial \phi}{\partial z}\right] = q\left(n - p + N_{\text{dop}}\right), \tag{2.97}$$

which determines the updated electrostatic potential in the device. The new potential, including the carrier-induced Hartree contribution, modifies the Hamiltonian and therefore the Green's functions, closing the self-consistency loop between quantum transport and electrostatics. Therefore, solving the NEGF-Poisson system requires two nested iterative loops. An inner loop enforces the self-consistency of the Green's functions by iteratively solving the Dyson and Keldysh equations together with the scattering self-energies. An outer loop updates the electrostatic potential by solving Poisson's equation using the carrier density obtained from the Green's functions.

The coupled NEGF-Poisson system is solved using a fixed-point iteration scheme. Starting from the initial potential, the Green's functions are computed, the carrier density is evaluated, and the Poisson equation is solved to obtain an updated electrostatic potential. This procedure is repeated until convergence is reached for both the Green's functions and the electrostatic potential. In practice, convergence is typically monitored through the variation of the potential or the carrier density between successive iterations.

Due to the strong nonlinear coupling between the transport equations and the electrostatics, the fixed-point iteration may require a large number of iterations before reaching convergence. Depending on the device structure, bias conditions, and scattering mechanisms considered, the number of iterations can range from several tens to several thousands. This slow convergence is further exacerbated by the computational cost associated with the repeated evaluation of the Dyson and Keldysh equations, making the use of full self-energies extremely computationally

intensive. As a consequence, simulating realistic devices longer than approximately 50 nm may become impractical unless suitable approximations are introduced.

One possible approximation consists of limiting the full self-energies to a subset of their diagonals, thereby restricting the spatial range of the interaction. In the extreme case, only the main diagonal is retained, corresponding to a purely local interaction. Such approximations reduce the computational effort because only the corresponding elements of the Green's functions are required for the evaluation of the self-energies.

Furthermore, if both the Hamiltonian and the self-energies can be expressed as block tridiagonal matrices, efficient numerical algorithms can be employed. One widely used approach is the recursive Green's function (RGF) method [29], which avoids the explicit inversion of the full matrix and significantly reduces the computational cost. The RGF method is described and derived in detail in Appendix B.

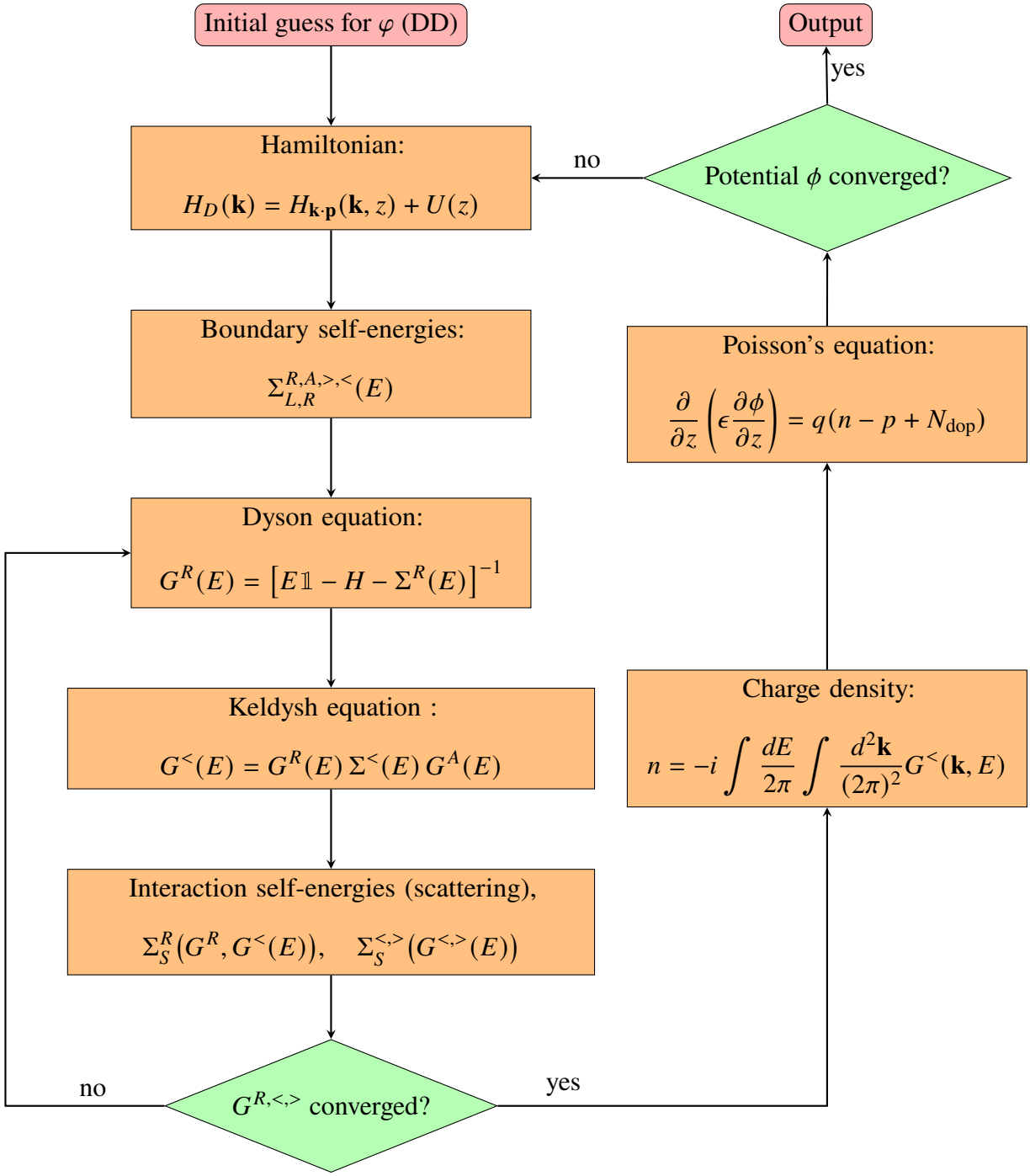


Fig. 2.2 Self-consistent multiband $\mathbf{k}\cdot\mathbf{p}$ NEGF-Poisson solution scheme with full carrier-phonon interaction self-energies. Only electron equations are shown for simplicity

2.9 Effective-mass Büttiker-probe NEGF approximation

The previous sections describe the full multiband NEGF formulation with carrier-phonon interaction self-energies. This approach is useful as a microscopic reference, but it is too expensive for the longer device regions considered in the following chapters. A reduced formulation is therefore introduced by combining an effective-mass band structure with Büttiker-probe self-energies. In this approximation, the electron and hole Green's functions are treated separately, while the electrostatic potential and the recombination terms couple the two carrier systems.

2.9.1 Effective mass approximation for electron and hole Green's functions

The effective mass approximation (EMA) represents the simplest continuum description of the electronic structure. Close to a band extremum, each band is approximated by a parabolic dispersion characterized by an effective mass,

$$E(\mathbf{k}) \simeq E_0 + \frac{\hbar^2 k^2}{2m^*}, \tag{2.98}$$

where E_0 is the energy of the band extremum and m^* is the corresponding effective mass. In a spatially nonuniform structure, the associated real-space Hamiltonian can be written as

$$H_{\mathrm{EMA}} = -\frac{\partial}{\partial z}\frac{\hbar^2}{2m_l^*(z)}\frac{\partial}{\partial z} + \frac{\hbar^2 k^2}{2m_t^*(z)} + V(z), \tag{2.99}$$

where $m_l^*(z)$ and $m_t^*(z)$ are the longitudinal and transverse effective masses, respectively, and $V(z)$ is the potential landscape experienced by the carriers.

EMA is particularly useful for fast and simple simulations in which the complete band structure is not required. Its simplicity allows for a straightforward implementation in transport and electrostatic simulations and provides an intuitive description of carrier dynamics. However, EMA treats the different bands independently and neglects band mixing, non-parabolicity, and the atomistic details of the crystal. Its applicability is therefore limited to systems in which a single-band approximation near the conduction- or valence-band edge is sufficient. In particular, interband tunneling cannot be generated intrinsically within the EMA Hamiltonian and requires a multiband description.

In the reduced NEGF implementation used with Büttiker probes, electron and hole transport

are described with separate effective-mass Hamiltonians. The corresponding retarded Green's functions are

$$G_n^R(\mathbf{k}, E) = \left[(E + i\eta)\mathbb{1} - H_n(\mathbf{k}) - \Sigma_n^R(\mathbf{k}, E)\right]^{-1}, \tag{2.100a}$$

$$G_p^R(\mathbf{k}, E) = \left[(E + i\eta)\mathbb{1} - H_p(\mathbf{k}) - \Sigma_p^R(\mathbf{k}, E)\right]^{-1}. \tag{2.100b}$$

The two Hamiltonians are constructed from the electron and hole band edges and effective masses, while the coupling between the two carrier systems enters through the electrostatic potential and through the generation-recombination terms. This separation is the band-structure approximation used in the Büttiker-probe calculations discussed below.

2.9.2 Büttiker-probe self-energies

After the effective-mass approximation has been introduced, scattering can be incorporated through a reduced Büttiker-probe description. Although the full carrier-phonon self-energies provide a physically detailed description of scattering, their self-consistent energy coupling considerably increases the numerical cost. A widely used phenomenological alternative for incorporating scattering in NEGF simulations is the Büttiker probe method [30], which retains the quantum-transport description and the spectral carrier distribution, but replaces the microscopic interaction self-energies with phenomenological local reservoirs depending on a scattering lifetime. It therefore represents a reduced NEGF model rather than a transition to semiclassical transport. In this framework, virtual contacts, referred to as probes, are attached to each spatial grid point of the device. These probes absorb carriers, randomize their phase and momentum, and subsequently reinject them into the system. The total probe retarded self-energy can be written as

$$\Sigma_{BP}^R(z, z', \mathbf{k}, E) = \sum_{p=1}^{N} \Sigma_p^R(\mathbf{k}, E)\delta(z - z_p)\delta(z' - z_p). \tag{2.101}$$

Because each probe is connected only to a single spatial grid point, the resulting self-energy is purely local. Consequently, the Büttiker probe self-energies are diagonal in real space. Since the probes can be considered locally at equilibrium, the lesser self-energy can be obtained by exploiting the fluctuation-dissipation theorem, yielding

$$\Sigma_{BP}^<(z, z', \mathbf{k}, E) = \sum_{p=1}^{N} i f_p(\mathbf{k}, E)\Gamma_{BP}^p(\mathbf{k}, E)\delta(z - z_p)\delta(z' - z_p), \tag{2.102}$$

where $\Gamma^p_{BP}(\mathbf{k}, E) = i(\Sigma^R_p(\mathbf{k}, E) - \Sigma^A_p(\mathbf{k}, E))$ is the broadening function and f_p is the Fermi-Dirac distribution of the p-th probe,

$$f_p(\mathbf{k}, E) = f_{FD}(E, \mu_p) = \frac{1}{1 + \exp\left(\frac{E-\mu_p}{k_B T}\right)}, \tag{2.103}$$

with μ_p representing the Fermi level of the p-th probe. To ensure charge conservation, no net current is allowed to flow through the virtual contacts. This constraint requires adjusting the probe Fermi levels so that current conservation is satisfied.

In the multi-scattering Büttiker-probe model [31, 32], the scattering lifetime is expressed in terms of parameters associated with the physical scattering mechanisms considered, rather than being introduced as an empirical phenomenological lifetime. In the formulation adopted in the following the probe broadening is constructed from the acoustic and polar-optical phonon scattering mechanisms. Starting from the general expression for optical phonon self-energy in the self-consistent Born approximation (SCBA), the in-scattering (lesser) and out-scattering (greater) functions can be written as

$$\Sigma^<(z, z, \mathbf{k}, E) = D_{op} n_q G^<(z, z, \mathbf{k}, E - \hbar\omega) + D_{op}(n_q + 1) G^<(z, z, \mathbf{k}, E + \hbar\omega), \tag{2.104}$$

$$\Sigma^>(z, z, \mathbf{k}, E) = D_{op}(n_q + 1) G^>(z, z, \mathbf{k}, E - \hbar\omega) + D_{op} n_q G^>(z, z, \mathbf{k}, E + \hbar\omega). \tag{2.105}$$

At thermodynamic equilibrium, the full lesser and greater Green's functions satisfy the fluctuation-dissipation relations $G^< = i f_{\text{eq}} A$ and $G^> = -i(1 - f_{\text{eq}})A$. Since the Büttiker-probe self-energy is local in real space, the local diagonal component entering the probe self-energy is then selected. Eq. (2.104) can therefore be rewritten as

$$\begin{aligned}\Sigma^<_p(z, z, \mathbf{k}, E) = {} & i D_{op} n_q f_{eq}(E - \hbar\omega) A(z, z, \mathbf{k}, E - \hbar\omega) \\ & + i D_{op}(n_q + 1) f_{eq}(E + \hbar\omega) A(z, z, \mathbf{k}, E + \hbar\omega),\end{aligned} \tag{2.106}$$

$$\begin{aligned}\Sigma^>_p(z, z, \mathbf{k}, E) = {} & -i D_{op}(n_q + 1)(1 - f_{eq}(E - \hbar\omega)) A(z, z, \mathbf{k}, E - \hbar\omega) \\ & - i D_{op} n_q (1 - f_{eq}(E + \hbar\omega)) A(z, z, \mathbf{k}, E + \hbar\omega).\end{aligned} \tag{2.107}$$

The corresponding broadening function is then

$$\Gamma_p(z,z,\mathbf{k},E) = i\left[\Sigma_p^{>}(z,z,\mathbf{k},E) - \Sigma_p^{<}(z,z,\mathbf{k},E)\right] \tag{2.108}$$

$$= D_{op}(n_q + 1 - f_{eq}(E-\hbar\omega))A(z,z,\mathbf{k},E-\hbar\omega) \tag{2.109}$$

$$+ D_{op}(n_q + f_{eq}(E+\hbar\omega))A(z,z,\mathbf{k},E+\hbar\omega).$$

It can be shown that the retarded self-energy can be approximated as [32]:

$$\Sigma_p^R(z,z,\mathbf{k},E) \approx -\frac{i}{2}\Gamma_p(z,z,\mathbf{k},E). \tag{2.110}$$

Recalling that the local spectral function satisfies $A(z,z,\mathbf{k},E) = i[G^R(z,z,\mathbf{k},E) - G^A(z,z,\mathbf{k},E)] = -2\,\mathrm{Im}\{G^R(z,z,\mathbf{k},E)\}$, the final probe retarded self-energy becomes

$$\Sigma_p^R(z,z,\mathbf{k},E) = iD_{op}(n_q + 1 - f_{eq}(E-\hbar\omega))\mathrm{Im}\{G^R(z,z,\mathbf{k},E-\hbar\omega)\} \tag{2.111}$$

$$+ iD_{op}(n_q + f_{eq}(E+\hbar\omega))\mathrm{Im}\{G^R(z,z,\mathbf{k},E+\hbar\omega)\}.$$

This formulation would be applicable only if the system is in thermodynamic equilibrium. However, it is possible to still use it as a non-equilibrium self-energy by replacing f_{eq} with an equivalent occupation function $f_{neq} = -iG^<(E)/A(E)$. In our case, we will use the concept of quasi-equilibrium Fermi levels, still using the Fermi-Dirac distribution. The quasi-Fermi levels will be found by imposing the conservation of current of the probes. An analogous construction is applied to the acoustic-phonon self-energy, starting from the elastic high-temperature expression introduced in Section 2.7.1.

Following the approach used in Ref. [33], an additional simplification can be introduced when the transverse dispersion is parabolic and separable from the longitudinal transport problem. In this case, the dependence on the transverse wave vector enters only through the transverse kinetic energy,

$$E_k = \frac{\hbar^2 k^2}{2m_t^*}, \tag{2.112}$$

so that retarded quantities can be written as functions of the shifted energy $E - E_k$. In contrast with the full multiband formulation discussed in Section 2.7, the transverse integration can then be performed analytically and no numerical integration over a transverse wave-vector grid is required.

After the change of variables from k to E_k, the occupation factor entering the lesser Büttiker probe self-energy is replaced by the corresponding integrated occupation over the transverse

modes. For a parabolic in-plane dispersion, this gives

$$\Sigma^{<}_{BP}(z,z,E) = i\frac{m_t^* k_B T}{\pi\hbar^2} F_0\left(\frac{\mu_{BP}-E}{k_B T}\right)\Gamma_{BP}(z,z,E), \tag{2.113}$$

where m_t^* is the transverse effective mass, $\Gamma_{BP} = i(\Sigma^R_{BP} - \Sigma^A_{BP})$, and F_0 is the zero-order Fermi-Dirac integral. In this way, the transverse degrees of freedom are included analytically through F_0, while the NEGF problem remains one-dimensional along the transport direction. Before this transverse integration is performed, the probe self-energies retain their dependence on $\mathbf{k}$; after the analytical integration, the effective one-dimensional quantities entering the solver depend only on z and E. Within the quasi-equilibrium approximation adopted here, the probe chemical potentials are obtained by imposing current conservation at each probe, while the injected carrier population accounts for the full transverse density of states.

In this formulation the scattering potentials depend only on the retarded Green's function and on the probe quasi-Fermi level, allowing the computation to remain efficient while still capturing the main features of different scattering processes. Because only the local diagonal components of G^R are required to evaluate the probe self-energies, the computational cost is significantly reduced compared to a full carrier-phonon self-energy calculation. Moreover, more advanced convergence algorithms can be implemented to ensure current conservation. This is discussed in greater detail in the following section.

2.9.3 Boundary self-energies in the effective-mass Büttiker-probe formulation

The boundary self-energy formulation introduced in Section 2.6 is independent of the Hamiltonian used to describe the device. In the effective-mass Büttiker-probe implementation, electrons and holes are described by separate single-band Hamiltonians. Consequently, each spatial node contains a single degree of freedom for each carrier system, and the band blocks appearing in the general multiband formulation reduce to scalars.
After spatial discretization, the complete operator entering the Dyson equation can be written as

$$\mathbf{A}(E) = E\mathbf{M} - \mathbf{H} - \mathbf{\Sigma}^R_{\text{scatt}}(E) - \mathbf{\Sigma}^R_L(E) - \mathbf{\Sigma}^R_R(E), \tag{2.114}$$

where $\mathbf{M}$ is the overlap matrix associated with the chosen spatial basis, $\mathbf{H}$ is the effective-mass Hamiltonian, and $\mathbf{\Sigma}^R_{\text{scatt}}$ contains all the local scattering self-energies. For the one-dimensional discretizations considered here, the resulting matrix is tridiagonal. To construct the boundary

terms, let $h_{0,\alpha}$ and $h_{1,\alpha}$ denote the diagonal and nearest-neighbour coefficients of the homogeneous coherent contact operator $E\mathbf{M}-\mathbf{H}$. The local scattering contribution is included from the beginning through the dressed diagonal coefficient

$$\widetilde{h}_{0,\alpha}(E) = h_{0,\alpha}(E) - \Sigma^R_{\mathrm{scatt},\alpha}(E) = h_{0,\alpha}(E) - \Sigma^R_{\mathrm{BP},\alpha}(E) - \Sigma^R_{\mathrm{other},\alpha}(E), \tag{2.115}$$

where $\alpha \in \{L, R\}$ identifies the contact. This notation separates the coherent and scattering contributions used to construct the semi-infinite continuation, while all self-energies remain included in the complete Dyson operator of Eq. (2.114).

The surface of a homogeneous semi-infinite contact is equivalent to the same contact after removal of its first node. The corresponding retarded boundary self-energy therefore satisfies

$$\Sigma^R_\alpha(E) = h_{1,\alpha}\left[\widetilde{h}_{0,\alpha}(E) - \Sigma^R_\alpha(E)\right]^{-1} h_{1,\alpha}. \tag{2.116}$$

Since the coefficients are scalars, Eq. (2.116) reduces to

$$\left[\Sigma^R_\alpha(E)\right]^2 - \widetilde{h}_{0,\alpha}(E)\Sigma^R_\alpha(E) + h^2_{1,\alpha} = 0, \tag{2.117}$$

whose solutions are

$$\Sigma^R_\alpha(E) = \frac{\widetilde{h}_{0,\alpha}(E) \pm \sqrt{\widetilde{h}^2_{0,\alpha}(E) - 4h^2_{1,\alpha}}}{2}. \tag{2.118}$$

The physical solution is selected by imposing the retarded boundary condition. In particular, the contact broadening

$$\Gamma_\alpha(E) = i\left[\Sigma^R_\alpha(E) - \Sigma^A_\alpha(E)\right] = -2\,\mathrm{Im}\left[\Sigma^R_\alpha(E)\right] \tag{2.119}$$

must be non-negative. This choice ensures that the contact absorbs outgoing carriers rather than acting as an unphysical source in the retarded problem. Equation (2.118) is the scalar counterpart of the outgoing-mode boundary condition derived for the multiband Hamiltonian in Section 2.6. Including $\Sigma^R_{\mathrm{scatt},\alpha}$ in $\widetilde{h}_{0,\alpha}$ prevents an abrupt connection between a scattering device region and a perfectly coherent semi-infinite contact. In particular, because $\Sigma^R_{\mathrm{BP},\alpha}$ depends on the retarded Green's function and on the corresponding probe quasi-Fermi level, the retarded boundary self-energy must be updated within the self-consistent Dyson loop.

The retarded contact self-energy determines the available contact states and the escape rate from the device, but it does not determine their occupation. The injection from contact α is introduced

through the lesser boundary self-energy,

$$\Sigma_\alpha^<(E) = i\Gamma_\alpha(E)F_\alpha(E,\mu_\alpha), \tag{2.120}$$

where μ_α is the Fermi level associated with the external reservoir. In the effective-mass implementation, the transverse degrees of freedom are integrated analytically. Consequently, F_α denotes the transverse-integrated occupation introduced previously, rather than the dimensionless Fermi-Dirac function of a single transverse mode. For a parabolic transverse dispersion,

$$F_\alpha(E,\mu_\alpha) = \frac{m_t^* k_B T}{\pi\hbar^2} F_0\left(\frac{\mu_\alpha - E}{k_B T}\right), \tag{2.121}$$

with the appropriate finite transverse-energy cutoff included when required by the numerical integration.

It is important to distinguish the external reservoir Fermi levels μ_L and μ_R from the local scattering quasi-Fermi levels associated with the Büttiker probes. At the first and last device nodes, the total self-energy contains both contributions. For example, at the left boundary,

$$\Sigma_{\text{tot},1}^{R,<} = \Sigma_L^{R,<} + \Sigma_{\text{BP},1}^{R,<} + \Sigma_{\text{other},1}^{R,<}. \tag{2.122}$$

The local probe describes incoherent scattering inside the semiconductor, whereas the contact self-energy describes exchange with the macroscopic external reservoir. The internal probes do not represent additional physical terminals. In the absence of generation and recombination, their quasi-Fermi levels are determined so that the net current exchanged with each probe vanishes. More generally, their current balance is combined with the local generation-recombination terms through the carrier continuity equation.

The external contacts instead carry the physical terminal current. The current associated with contact α can be evaluated from its boundary self-energies as

$$J_\alpha = s_\alpha \frac{q}{\hbar}\int \frac{dE}{2\pi}\,\mathrm{Re}\left\{G_{\alpha\alpha}^R\Sigma_\alpha^< - \Sigma_\alpha^< G_{\alpha\alpha}^A - \Sigma_\alpha^R G_{\alpha\alpha}^< + G_{\alpha\alpha}^<\Sigma_\alpha^A\right\}, \tag{2.123}$$

where $s_L = -1$ and $s_R = +1$ account for the orientation of the two interfaces. The current evaluated from Eq. (2.123) must be continuous with the current computed inside the device. The corresponding closure conditions can be written as

$$J_L^{\text{dev}} - J_L^{\text{cont}} = 0, \tag{2.124a}$$

$$J_R^{\text{dev}} - J_R^{\text{cont}} = 0, \tag{2.124b}$$

with the signs adapted to the adopted current convention.
Equations (2.124) may be implemented using different but algebraically equivalent choices of nonlinear unknowns. The contact Fermi levels may be adjusted directly until the contact and device currents coincide. Alternatively, the common terminal currents may be retained as explicit unknowns and constrained to be equal to both the reservoir current and the current entering the device. Eliminating the auxiliary current variables recovers the same current-continuity equations. The distinction is therefore numerical rather than physical.
The complete Büttiker-probe problem consequently combines three types of conditions: Poisson's equation determines the electrostatic potential, the internal probe equations enforce the local carrier balance, and Eqs. (2.124) connect the scattering device to the external reservoirs. The numerical solution of this coupled system is discussed in the following subsection.

2.9.4 Büttiker-probe self-energy scheme

As discussed above, the solution of the full multiband NEGF-Poisson system with carrier-phonon interaction self-energies requires the simultaneous self-consistent solution of the Dyson and Keldysh equations inside an inner iteration loop. The interaction self-energies depend explicitly on the lesser (greater) Green's functions, which makes the evaluation of both equations necessary at every iteration until convergence is reached.
A simplification of this scheme can be obtained by exploiting the Büttiker probe formulation introduced in Section 2.9.2. In this approach the retarded interaction self-energies depend on the retarded Green's function and on the probe quasi-Fermi levels, but not on the lesser or greater Green's functions. As a consequence, the internal self-consistency loop can be restricted to the Dyson equation alone, which is iterated until the retarded Green's function converges. At each iteration, the retarded Büttiker-probe self-energies and the boundary self-energies depending on them are updated before the Dyson equation is evaluated again. Once G^R has been obtained, the corresponding lesser self-energy of the probes can be computed exploiting the fluctuation-dissipation theorem, and the Keldysh equation can be evaluated to determine the carrier distribution.
The overall algorithm therefore differs from the full carrier-phonon self-energy scheme described previously. As illustrated in Fig. 2.3, the inner loop is devoted exclusively to the convergence of the retarded Green's function through the Dyson equation, while the outer self-consistency loop enforces both electrostatic self-consistency and current conservation within the device.
In this loop, the carrier density obtained from the lesser Green's function is used to update the electrostatic potential through Poisson's equation, which in turn modifies the Hamiltonian

entering the Dyson equation. At the same time, the quasi-Fermi levels associated with the Büttiker probes must be adjusted in order to enforce current conservation.
In the probe model, each fictitious reservoir exchanges carriers with the device, and the probe chemical potentials are determined such that the net current flowing into each probe vanishes. This condition ensures that the probes only emulate inelastic scattering processes without acting as additional sources or sinks of particles. The electrostatic potential and the probe quasi-Fermi levels are therefore determined simultaneously within the outer self-consistent loop. The current conservation condition can be expressed through the continuity equation

$$\frac{\partial J_n}{\partial z} = q(R - G), \tag{2.125}$$

which relates the divergence of the electron current density to the local generation and recombination rates. Within the NEGF framework the energy-resolved current density divergence can be written as

$$r(z, E) = \frac{q}{h}\left[H(E)G^{<}(E) - G^{<}(E)H(E)\right]_{z,z}, \tag{2.126}$$

The total current-density divergence is obtained by integrating r(z,E) over energy. The probe quasi-Fermi levels are therefore adjusted so that the current exchanged with each probe satisfies the continuity relation. One advantage of this formulation is that by explicitly enforcing current conservation, generation and recombination phenomena can be incorporated in a manner analogous to drift-diffusion models, so that DD-like recombination and generation rates can be used to account for SRH or radiative recombination, without necessarily needing to compute the appropriate self-energies.
An additional advantage of the Büttiker probe formulation concerns the numerical solution of the self-consistency loop. While the full carrier-phonon self-energy scheme typically relies on a fixed-point iteration and may require a very large number of iterations before convergence is reached, the determination of the probe quasi-Fermi levels can instead be formulated as a system of nonlinear equations. This system can be solved using a Newton-Raphson method, which generally exhibits much faster convergence properties. In practice, while the full carrier-phonon self-energy algorithm may require hundreds or even thousands of iterations, the Büttiker probe scheme often converges in only a few tens of iterations and in favorable cases in as few as ~ 10 iterations.
The resulting reduction in the number of iterations, combined with the simplified inner loop involving only the Dyson equation, leads to a significant decrease in the overall computational cost of the simulation. However, the inner iteration loop is still required, meaning that the Dyson equation must be solved repeatedly until convergence of the retarded Green's function

is achieved. In practice, this involves the repeated inversion of large matrices over the energy grid, which remains one of the most computationally demanding steps of the algorithm. As a consequence, even with the simplifications introduced by the Büttiker probe approach, the method still requires significant computational resources. In practical implementations, this typically limits the simulation domain to devices of at most a few hundred nanometers.

Therefore, although the NEGF formalism represents one of the most physically accurate frameworks for describing quantum transport in electronic devices, its direct application to realistic device structures remains challenging due to these computational limitations. Nevertheless, NEGF simulations are extremely valuable as a reference tool. In particular, they can be used to benchmark simplified models or to extract effective material parameters, which can then be employed in less computationally demanding approaches. In the following applications, the Büttiker probe formulation will be mainly used as a practical NEGF implementation, since it allows the essential quantum-transport effects to be retained with a reduced computational cost. One important example of a simplified approach to be benchmarked against NEGF is the quantum-corrected drift-diffusion model, which retains part of the quantum physics while allowing the simulation of much larger device structures.

The Büttiker probe and drift-diffusion formulations share a common mathematical structure based on the coupled solution of the electrostatic and carrier-continuity equations. They differ, however, in the constitutive relations used to evaluate the carrier charge and current. In the Büttiker probe formulation, these quantities are obtained from the Green's functions and retain a quantum-mechanical spectral description. In drift-diffusion, they are instead determined from local carrier densities and semiclassical transport relations. Quantum-corrected drift-diffusion models introduce an intermediate description by retaining quantum corrections in the evaluation of the carrier density and density of states, while preserving the semiclassical treatment of transport. These models are introduced and discussed in the following chapter.

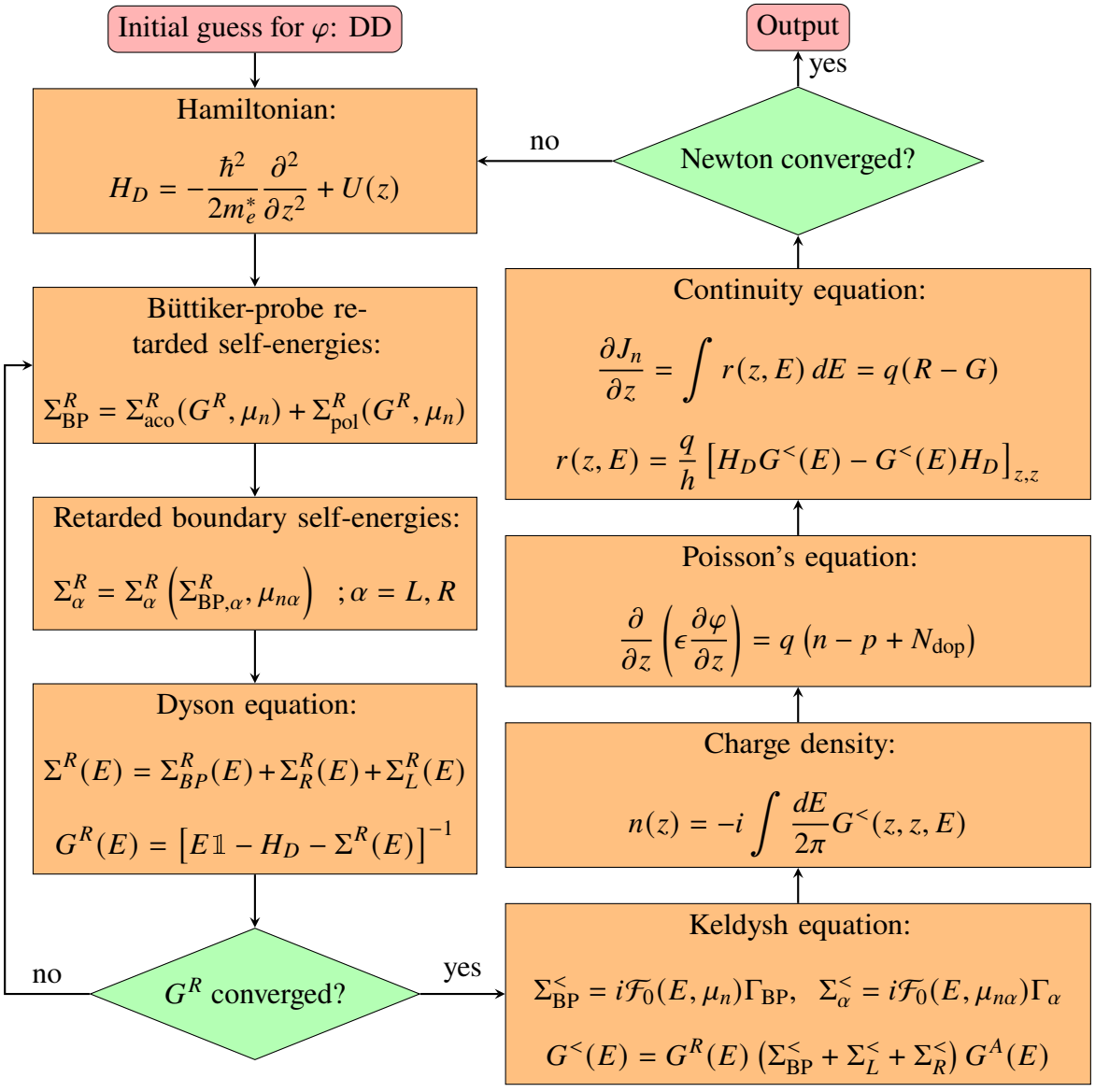


Fig. 2.3 Self-consistent NEGF-Poisson solution scheme with Büttiker-probe self-energies. Only the electron transport system is shown for simplicity.

Chapter 3

Quantum-corrected drift-diffusion

The drift-diffusion (DD) formalism, originally introduced by van Roosbroeck [34], constitutes the standard macroscopic framework for the simulation of carrier transport in semiconductor devices. From a theoretical perspective, it can be obtained as a low-order moment closure of the Boltzmann Transport Equation under near-equilibrium assumptions. Owing to its favorable balance between physical accuracy, numerical robustness, and computational efficiency, it is widely employed in both academic and industrial environments and forms the core of most commercially available simulation tools, such as Synopsys Sentaurus and COMSOL Multiphysics. Detailed mathematical treatments of the classical formulation can be found in [35, 36], while its practical implementation relies on robust numerical schemes, among which the Scharfetter-Gummel discretization [37] plays a central role in ensuring numerical stability and accuracy.

As discussed at the end of the previous chapter, the Büttiker-probe NEGF formulation and the QCDD methods considered in this chapter share a common mathematical structure based on the coupled solution of the electrostatic and carrier-continuity equations. Their main difference lies in the constitutive relations used to evaluate the carrier charge and current. In the Büttiker-probe formulation, both quantities are obtained from the Green's functions and retain a quantum-mechanical spectral description. In QCDD, quantum effects enter the carrier charge through quantum-corrected effective band edges, while the current is still evaluated through semiclassical drift-diffusion relations.

However, the semiclassical nature of the drift-diffusion formalism inherently limits its applicability when quantum-mechanical effects become significant. In particular, phenomena such as carrier confinement and tunneling, which play a central role in nanoscale and heterostructure devices, cannot be fully described within the classical framework. Although phenomenological extensions can be introduced to reproduce some of these effects, their parameters are often not directly

accessible and must be calibrated [38]. Fully quantum approaches, such as the Non-Equilibrium Green's Function formalism discussed in the previous chapter, provide a more rigorous description, but their computational cost remains a major limitation for large-scale and device-level simulations. For this reason, intermediate modeling approaches have been developed to incorporate quantum effects within the drift-diffusion framework. Quantum-corrected drift-diffusion methods aim to retain the numerical efficiency of the classical model while introducing physically motivated corrections for the most relevant quantum phenomena. In this chapter, the common quantum-corrected drift-diffusion formulation is first introduced directly in terms of effective band edges. The Poisson-Schrödinger approach is then presented as a model that retains an explicitly quantum-mechanical description of the local density of states, followed by the density-gradient approximation, which provides a simpler and computationally less demanding quantum correction.

3.1 General quantum-corrected drift-diffusion formalism

Within a quantum-corrected drift-diffusion (QCDD) formulation, quantum effects are incorporated by replacing the classical band edges with effective band edges that reproduce a quantum-corrected carrier distribution. Assuming parabolic bands, the corresponding electron and hole local densities of states can be written as

$$\begin{aligned} LDOS_n(E,\mathbf{r}) &= \frac{1}{2\pi^2}\left(\frac{2m_n^*}{\hbar^2}\right)^{3/2}\sqrt{E-\widetilde{E}_c(\mathbf{r})}, \\ LDOS_p(E,\mathbf{r}) &= \frac{1}{2\pi^2}\left(\frac{2m_p^*}{\hbar^2}\right)^{3/2}\sqrt{\widetilde{E}_v(\mathbf{r})-E}. \end{aligned} \tag{3.1}$$

where $\widetilde{E}_c$ and $\widetilde{E}_v$ are the quantum-corrected conduction- and valence-band edges. The carrier densities are obtained from the corresponding LDOS as

$$\begin{aligned} n(\mathbf{r}) &= \int_{\widetilde{E}_c}^{\infty} f_n(E) LDOS_n(E,\mathbf{r})\, dE, \\ p(\mathbf{r}) &= \int_{-\infty}^{\widetilde{E}_v} \left[1-f_p(E)\right] LDOS_p(E,\mathbf{r})\, dE. \end{aligned} \tag{3.2}$$

where $f_n(E)$ and $f_p(E)$ are the Fermi-Dirac occupation functions associated with the electron and hole quasi-Fermi levels, respectively. By substituting Eq. (3.1) into Eq. (3.2), the electron

and hole densities can be expressed in terms of Fermi-Dirac integrals of order $1/2$:

$$\begin{aligned} n(\mathbf{r}) &= N_C \mathcal{F}_{1/2}\left(\frac{E_{F_n} - \widetilde{E}_c(\mathbf{r})}{k_B T}\right), \\ p(\mathbf{r}) &= N_V \mathcal{F}_{1/2}\left(\frac{\widetilde{E}_v(\mathbf{r}) - E_{F_p}}{k_B T}\right). \end{aligned} \tag{3.3}$$

where N_C and N_V are the effective densities of states in the conduction and valence bands, respectively, and $\mathcal{F}_{1/2}$ denotes the Fermi-Dirac integral of order $1/2$, defined as

$$\mathcal{F}_\mu(x) \equiv \frac{1}{\Gamma(\mu+1)} \int_0^\infty \frac{t^\mu \, dt}{1 + \exp(t - x)}. \tag{3.4}$$

In the non-degenerate limit, when $(E_{F_n} - \widetilde{E}_c) \ll -k_B T$ and $(\widetilde{E}_v - E_{F_p}) \ll -k_B T$, the Fermi-Dirac distribution can be approximated by the Boltzmann distribution. Equation (3.3) then reduces to

$$\begin{aligned} n(\mathbf{r}) &= N_C \exp\left(\frac{E_{F_n} - \widetilde{E}_c(\mathbf{r})}{k_B T}\right), \\ p(\mathbf{r}) &= N_V \exp\left(\frac{\widetilde{E}_v(\mathbf{r}) - E_{F_p}}{k_B T}\right). \end{aligned} \tag{3.5}$$

These expressions provide a simplified description of carrier statistics and are widely used when degeneracy effects can be neglected. The effective band edges are related to the classical band edges through the electron and hole quantum corrections Λ_n and Λ_p:

$$\begin{aligned} \widetilde{E}_c &= E_c - q\Lambda_n = -q\phi + \Delta E_c - q\Lambda_n, \\ \widetilde{E}_v &= E_v + q\Lambda_p = -q\phi + \Delta E_v + q\Lambda_p. \end{aligned} \tag{3.6}$$

Here, Λ_n and Λ_p are expressed in volts. The same quantum corrections therefore enter the constitutive current relations, which become

$$\begin{aligned} \mathbf{J}_n &= -qn\mu_n \nabla\left(\phi + \Lambda_n\right) + qD_n \nabla n, \\ \mathbf{J}_p &= -qp\mu_p \nabla\left(\phi - \Lambda_p\right) - qD_p \nabla p. \end{aligned} \tag{3.7}$$

When Fermi-Dirac statistics are retained, D_n and D_p are understood as generalized diffusion coefficients consistent with the corresponding mobilities. The usual Einstein relation $D = \mu k_B T/q$ is recovered in the non-degenerate limit. The common steady-state electrostatic and transport

equations are consequently written as

$$\nabla \cdot (\varepsilon \nabla \phi) = q \left(n - p + N_{\text{dop}} \right), \tag{3.8a}$$

$$\mathbf{J}_n = -qn\mu_n \nabla \left(\phi + \Lambda_n \right) + qD_n \nabla n, \tag{3.8b}$$

$$\mathbf{J}_p = -qp\mu_p \nabla \left(\phi - \Lambda_p \right) - qD_p \nabla p, \tag{3.8c}$$

$$\nabla \cdot \mathbf{J}_n = qU, \tag{3.8d}$$

$$\nabla \cdot \mathbf{J}_p = -qU. \tag{3.8e}$$

Here, ϕ is the electrostatic potential, n and p are the electron and hole densities, ε is the dielectric permittivity, μ_n and μ_p are the carrier mobilities, and D_n and D_p are the corresponding diffusion coefficients. With the sign convention used in Eq. (3.8), the net doping is defined as $N_{\text{dop}} = N_A^- - N_D^+$. Equations (3.3), (3.6), and (3.8) define the common structure of the QCDD formulation. The classical drift-diffusion limit is recovered by setting $\Lambda_n = \Lambda_p = 0$, or equivalently $\widetilde{E}_{c,v} = E_{c,v}$. A specific QCDD model is obtained by providing an additional closure for the effective band edges, and the resulting coupled system must be solved self-consistently.

The term $U = R - G$ represents the net recombination rate, defined as the difference between recombination and generation processes. In semiconductor devices, several recombination mechanisms contribute to U. The most common ones are Shockley-Read-Hall (SRH), radiative, and Auger recombination. Neglecting optical generation for the moment, $G = 0$, the net recombination rate is written as

$$U = U_{\text{SRH}} + U_{\text{rad}} + U_{\text{Auger}}. \tag{3.9}$$

The SRH recombination describes carrier capture and emission through defect states in the bandgap and is given by [39]

$$U_{\text{SRH}} = \frac{np - n_i^2}{\tau_p (n + n_1) + \tau_n (p + p_1)}, \tag{3.10}$$

where τ_n and τ_p are the electron and hole lifetimes, while n_1 and p_1 depend on the trap energy level. Radiative recombination corresponds to band-to-band electron-hole recombination with photon emission and is typically modeled as

$$U_{\text{rad}} = B \left(np - n_i^2 \right), \tag{3.11}$$

where B is the radiative recombination coefficient. Auger recombination involves a three-particle interaction, where the recombination energy is transferred to a third carrier, and is described by

$$U_{\text{Auger}} = \left(C_{nnp}n + C_{ppn}p\right)\left(np - n_i^2\right), \tag{3.12}$$

where C_{nnp} and C_{ppn} are the Auger coefficients.

In the following, two closures of the common QCDD formulation are presented: the Poisson-Schrödinger approach [33] and the density-gradient approximation [40]. In the Poisson-Schrödinger model, the quantum states and the corresponding local density of states are calculated explicitly from the Schrödinger equation, and the resulting quantum charge is mapped onto the effective band edges used in the common transport system. The density-gradient model introduces a further approximation by determining the same effective band-edge corrections from differential equations depending on the spatial variation of the carrier densities. It therefore reduces the computational cost at the expense of a less detailed representation of the quantum charge.

3.2 Poisson-Schrödinger approach

The Poisson-Schrödinger approach closes the generic QCDD formulation by explicitly solving the Schrödinger equation and extracting the corresponding quantum correction from its solution. The carrier density is first computed from a quantum-mechanical density of states and is then mapped onto the effective band edge entering the common QCDD equations. For simplicity, let us first consider the electron problem in the effective-mass approximation. The electrostatic potential obtained from Poisson's equation defines the conduction-band edge entering the Schrödinger equation,

$$\left[-\frac{\hbar^2}{2m_n^*}\nabla^2 + E_c(\mathbf{r})\right]\psi_i(\mathbf{r}) = E_i\psi_i(\mathbf{r}), \tag{3.13}$$

where

$$E_c(\mathbf{r}) = -q\phi(\mathbf{r}) + \Delta E_c. \tag{3.14}$$

Here E_i and ψ_i are the eigenvalues and eigenfunctions of the confined states. More advanced Hamiltonians, for example a $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian, can be used in place of the effective-mass approximation, but the basic structure of the method remains the same. In a layered structure, the Schrödinger equation is solved along the confinement direction, while the transverse directions are treated as free. Therefore, each eigenvalue E_i represents the bottom of a two-dimensional

subband. The corresponding quantum local density of states can be written as

$$LDOS_n^{\mathrm{PS}}(z,E) = \frac{m_n^*}{\pi\hbar^2}\sum_i |\psi_i(z)|^2\Theta(E-E_i), \tag{3.15}$$

where the prefactor comes from the integration over the transverse momentum considering spin degeneracy and Θ is the Heaviside step function. The step-like behavior reflects the two-dimensional density of states associated with each subband: each confined state starts contributing to the density of states only for energies above its eigenvalue. The electron spectral density is then obtained by weighting the quantum LDOS by the occupation probability,

$$n_Q^{\mathrm{PS}}(E,z) = LDOS_n^{\mathrm{PS}}(z,E) f_n(E), \tag{3.16}$$

where $f_n(E)$ is the electron Fermi-Dirac distribution. The total quantum electron density is recovered by integrating over energy. Substituting Eq. (3.15) gives

$$n_Q^{\mathrm{PS}}(z) = \frac{m_n^* k_B T}{\pi\hbar^2}\sum_i |\psi_i(z)|^2 \mathcal{F}_0\left(\frac{E_{F_n}(z)-E_i}{k_B T}\right), \tag{3.17}$$

where E_{F_n} is the electron quasi-Fermi level and $\mathcal{F}_0$ is the Fermi-Dirac integral of order zero, which can be evaluated analytically as:

$$\mathcal{F}_0(\eta) = \int_0^{+\infty} \frac{dx}{1+\exp(x-\eta)} = \ln\left(1+\exp(\eta)\right). \tag{3.18}$$

Therefore, Eq. (3.17) can equivalently be written as

$$n_Q^{\mathrm{PS}}(z) = \frac{m_n^* k_B T}{\pi\hbar^2}\sum_i |\psi_i(z)|^2 \ln\left[1+\exp\left(\frac{E_{F_n}(z)-E_i}{k_B T}\right)\right]. \tag{3.19}$$

An analogous construction applies to holes:

$$p_Q^{\mathrm{PS}}(z,E) = LDOS_p^{\mathrm{PS}}(z,E)\left[1-f_p(E)\right], \tag{3.20}$$

where the hole LDOS is built from the valence-band Schrödinger eigenstates. Depending on the energy convention used for the valence band, the corresponding step functions are reversed with respect to the electron case, since hole subbands contribute below the valence-band eigenvalues. The total quantum hole density is obtained by integrating Eq. (3.20) over energy.

In the formulation presented so far, the Schrödinger eigenstates are treated as infinitely sharp spectral states. After integration over the transverse wave vector, each discrete eigenvalue therefore gives rise to a sharp two-dimensional subband onset, represented by the Heaviside function in Eq. (3.15). This is the ideal Poisson-Schrödinger limit. In a real device, however, the spectral weight associated with each state is not perfectly sharp: scattering, finite lifetime effects, coupling to contacts, disorder, and alloy fluctuations broaden the subband edges. In a full NEGF treatment, this broadening is generated naturally by the corresponding self-energies and by the open-boundary conditions.

Within the Poisson-Schrödinger framework, the same effect can only be introduced phenomenologically. A simple way to do this is to convolve each sharp subband onset with a normalized Gaussian of width σ. Equivalently, the sharp Heaviside function in Eq. (3.15) is replaced by the Gaussian-broadened step

$$\Theta_\sigma(E - E_i) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{E - E_i}{\sqrt{2}\sigma}\right)\right], \tag{3.21}$$

where

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt \tag{3.22}$$

is the error function. In the limit $\sigma \to 0^+$, the sharp step function is recovered. The parameter σ should therefore be interpreted as a phenomenological spectral broadening, not as a quantity obtained self-consistently from the Poisson-Schrödinger model itself. Figure 3.1 reports an LDOS cut around the quantum well of an InGaN/GaN LED, showing the smoothing of the subband onsets. With this approximation, the electron LDOS becomes

$$LDOS^{\mathrm{PS}}_{n,\sigma}(z, E) = \frac{m_n^*}{\pi\hbar^2} \sum_i |\psi_i(z)|^2 \Theta_\sigma(E - E_i). \tag{3.23}$$

Including this broadened LDOS directly in the computation of the integrated charge density would require evaluating

$$n^{\mathrm{PS}}_{Q,\sigma}(z) = \int_{-\infty}^{+\infty} LDOS^{\mathrm{PS}}_{n,\sigma}(z, E) f_n(E)\, dE, \tag{3.24}$$

which no longer reduces to the simple analytical expression of Eq. (3.19). Therefore, using the broadened LDOS self-consistently inside the Poisson-Schrödinger loop would significantly complicate the evaluation of the quantum charge and the extraction of the effective potential. In practice, for the structures considered in this work, numerical tests showed that including this

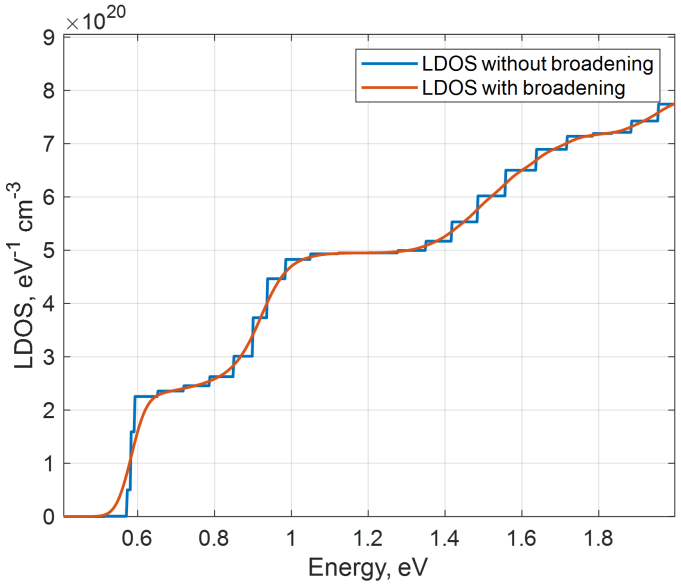


Fig. 3.1 LDOS cut of an InGaN/GaN LED near the QW, with 20 meV broadening (orange line) or without broadening (blue line)

broadening modifies the integrated quantum charge by less than 1%, since it mainly redistributes the spectral weight over a narrow energy range around each subband edge. For this reason, the ideal subband expression of Eq. (3.19) is retained for the self-consistent computation of the integrated quantum charge, while the broadened LDOS can be used, when needed, for energy-resolved spectral quantities, such as spectral charge densities.

Once the quantum carrier density has been computed, the electron part of the QCDD system is closed by defining an effective quantum-corrected conduction-band edge $\widetilde{E}_c$. This is done by requiring that the bulk-like carrier-statistics relation introduced in Eq. (3.3) reproduces the quantum density:

$$n_Q^{\mathrm{PS}}(z) = N_C \mathcal{F}_{1/2}\left(\frac{E_{F_n}(z) - \widetilde{E}_c(z)}{k_B T}\right). \tag{3.25}$$

By inverting this relation, one obtains

$$\widetilde{E}_c(z) = E_{F_n}(z) - k_B T \mathcal{F}_{1/2}^{-1}\left(\frac{n_Q^{\mathrm{PS}}(z)}{N_C}\right). \tag{3.26}$$

The effective band edge $\widetilde{E}_c$ can therefore be interpreted as the conduction-band profile that, when inserted into the common QCDD carrier-statistics relation, gives the same charge density as the quantum Schrödinger solution. Equivalently, one may introduce a quantum correction Λ_n

through

$$\widetilde{E}_c(z) = E_c(z) - q\Lambda_n(z). \tag{3.27}$$

The same procedure can be applied to the valence band, defining an effective valence-band edge and a corresponding hole quantum correction. The physical meaning of the method is therefore the following. The Schrödinger equation provides the quantum density of states and the associated quantum carrier density. This density is then mapped back onto the common QCDD description by defining an effective potential that reproduces the same charge distribution. In this way, confinement effects are included through the quantum LDOS, while the transport problem remains formulated through Eq. (3.8).

The Poisson-Schrödinger approach has several advantages. Since the carrier density is computed from Schrödinger eigenstates, quantum confinement is naturally included, and the energy-resolved quantities $n_Q^{\mathrm{PS}}(z, E)$ and $p_Q^{\mathrm{PS}}(z, E)$ can be used to evaluate spectral recombination rates. This feature will be particularly important in the following chapter, where trap-assisted recombination is written in terms of spectral carrier densities. At the same time, the method is significantly less demanding than a full quantum-transport approach such as NEGF.
However, important limitations remain. The transport model is still semiclassical, since current is described by the constitutive relations in Eq. (3.7) and not by a spectral current. Therefore, tunneling is included only indirectly through the effective quantum potential and the modified carrier density. Moreover, the Schrödinger equation is not directly embedded into the Newton loop used for the coupled QCDD transport and Poisson equations. A nested iterative scheme is required: the Schrödinger equation is solved for a given electrostatic potential, the quantum correction is extracted, the common QCDD system is solved with the updated effective potential, and the resulting electrostatic potential is fed back into the Schrödinger problem. The self-consistent scheme of the PSDD method is shown in Fig. 3.2.
Despite this apparent complexity, in practical applications only a limited number of outer iterations between the Schrödinger and QCDD solvers is usually required to reach convergence.

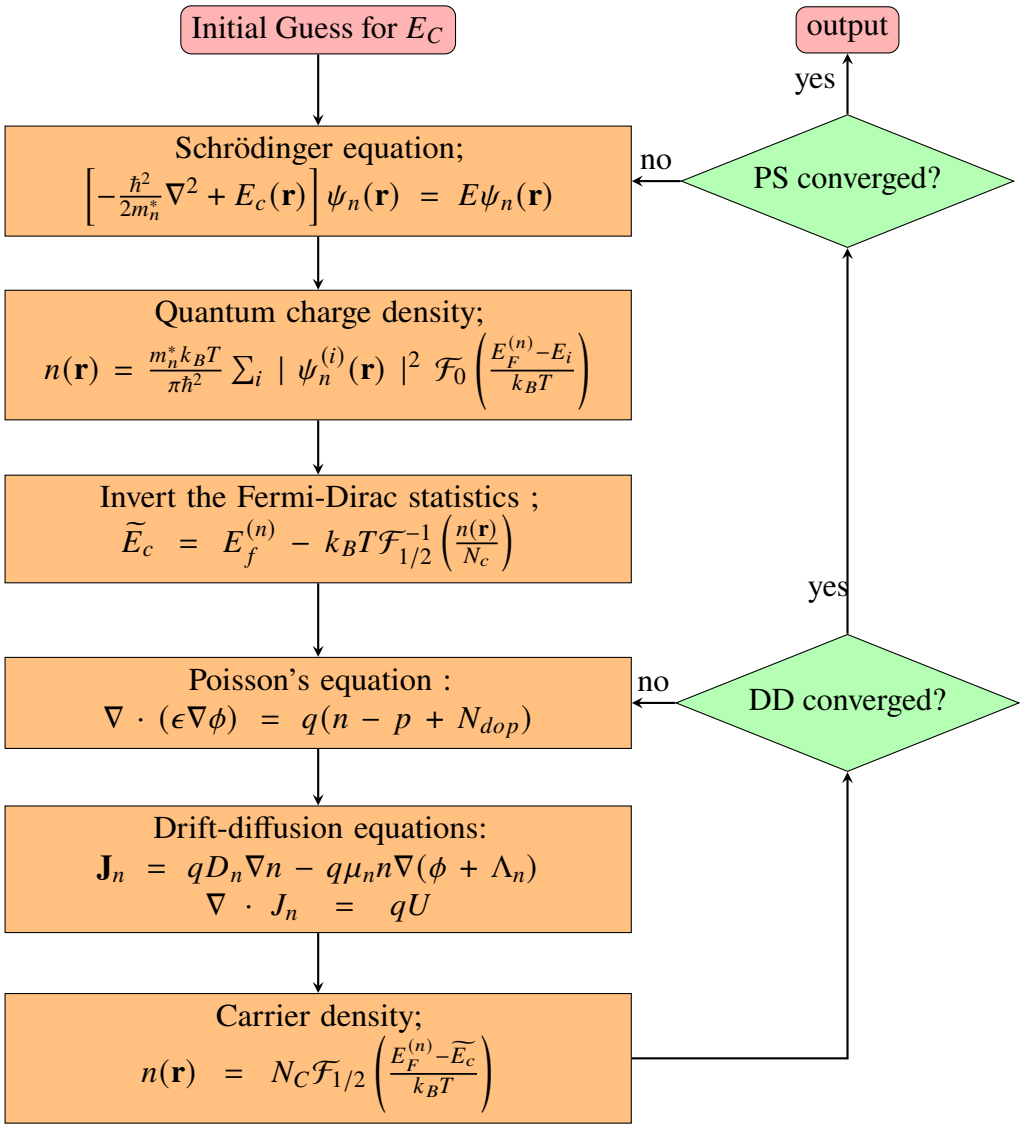


Fig. 3.2 Self-consistent solution scheme of the Poisson-Schrödinger drift-diffusion (PSDD) model, where quantum corrections are extracted from the Schrödinger equation and incorporated into the drift-diffusion framework through an effective potential. Only the electron equations are shown for simplicity.

3.3 Density-gradient approximation

The density-gradient (DG) approach, also referred to as quantum drift-diffusion, provides an alternative closure of the generic QCDD formulation and a further simplification of the Poisson-Schrödinger description [40]. Instead of explicitly solving a Schrödinger eigenvalue problem and reconstructing the quantum local density of states, the effective band-edge corrections are obtained from differential relations depending on the spatial variation of the carrier density.
The quantum-hydrodynamic origin of the density-gradient correction can be traced to moment expansions of the Wigner-Boltzmann equation, which represents the quantum counterpart of the Boltzmann Transport Equation [41]. The zeroth-order moment leads to the usual carrier continuity equations, while the first-order moment produces a momentum-balance equation containing an additional quantum contribution. Under near-equilibrium and diffusive closure assumptions, this contribution can be expressed as the gradient of an effective quantum potential. To close the QCDD formulation, a relation between the quantum contribution and the carrier densities must be introduced. Early quantum-correction formulations expressed this contribution through spatial derivatives of the potential energy. In heterostructures, however, the band edges and material parameters may exhibit abrupt variations at interfaces, making such expressions difficult to evaluate and potentially ill-defined. The density-gradient formulation avoids this difficulty by rewriting the quantum correction in terms of the carrier densities, which generally vary more smoothly. The electron and hole quantum potentials are defined as

$$\Lambda_n = \frac{\hbar^2}{6q} \frac{\nabla \cdot \left[(m_n^*)^{-1} \nabla \sqrt{n} \right]}{\sqrt{n}}, \qquad \Lambda_p = \frac{\hbar^2}{6q} \frac{\nabla \cdot \left[(m_p^*)^{-1} \nabla \sqrt{p} \right]}{\sqrt{p}}. \tag{3.28}$$

The quantum potentials can equivalently be interpreted as corrections to the classical band edges,

$$\widetilde{E}_c = E_c - q\Lambda_n, \qquad \widetilde{E}_v = E_v + q\Lambda_p. \tag{3.29}$$

Equations (3.28) and (3.29) provide the density-gradient closure of the generic formulation introduced in Section 3.1. They must be solved consistently with the carrier-statistics relations in Eq. (3.3) and with the electrostatic, current, and continuity equations in Eq. (3.8). Since the quantum potentials contain spatial derivatives of $\sqrt{n}$ and $\sqrt{p}$, the resulting system is more nonlinear than conventional drift-diffusion and requires the simultaneous self-consistent determination of the electrostatic potential, carrier densities, and quantum corrections.

Unlike the Poisson-Schrödinger approach, the density-gradient model does not require the solution of an eigenvalue problem and therefore avoids the explicit calculation of confined states.

This substantially reduces the computational cost and facilitates its application to extended device structures. However, this method does not provide an energy-resolved quantum LDOS or a direct description of individual subbands. Quantum confinement and tunneling-related charge redistribution are instead represented approximately through the modification of the effective band edges and carrier densities. The density-gradient approximation is therefore expected to be most accurate when the quantum correction varies smoothly and when a detailed spectral description is not required. Its accuracy may be reduced in structures characterized by strongly localized states, sharp resonances, or transport dominated by a small number of discrete confined levels. The self-consistent solution scheme of the DGDD model is summarized in Fig. 3.3.

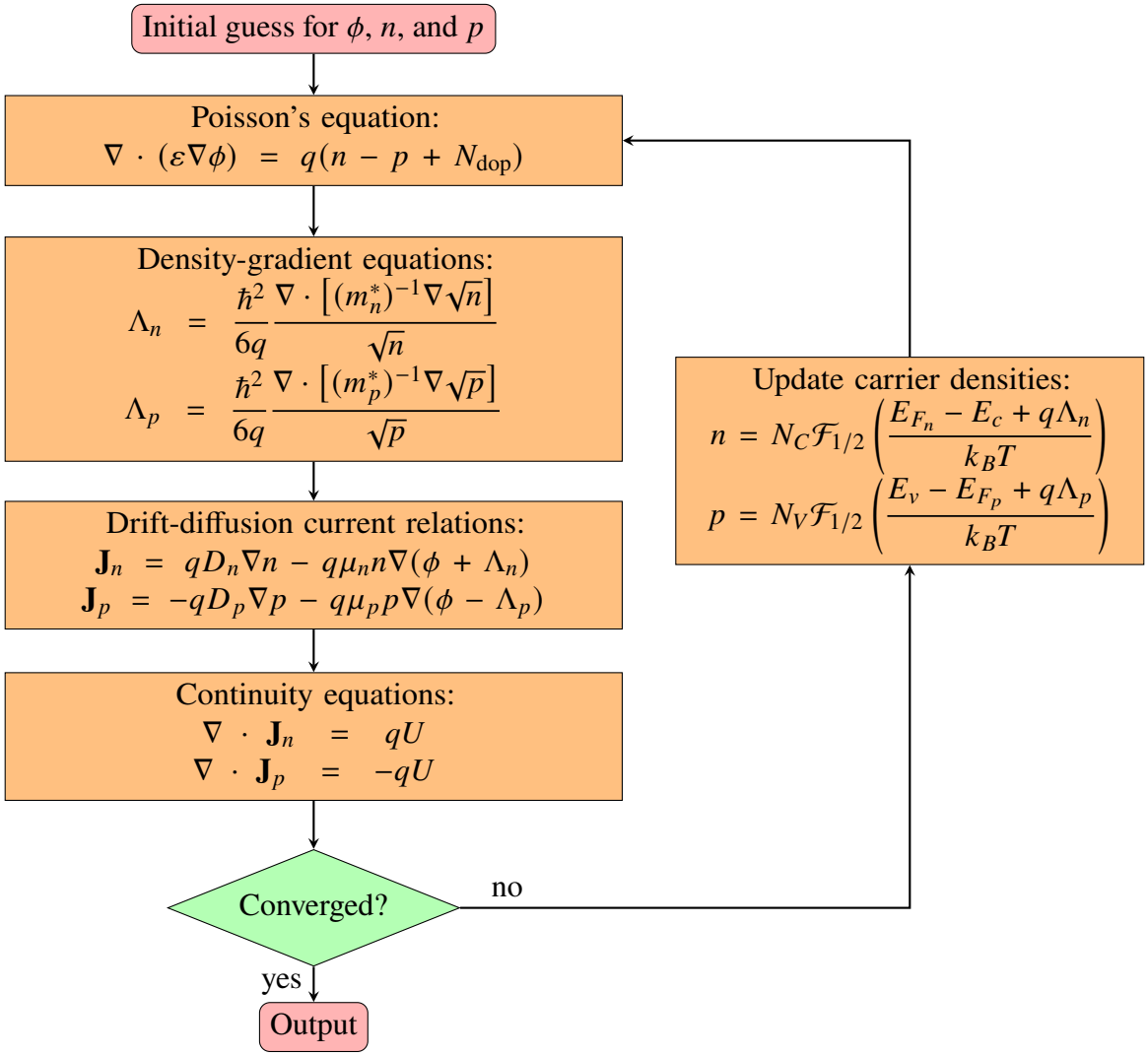


Fig. 3.3 Self-consistent solution scheme of the density-gradient drift-diffusion model, including the Poisson, density-gradient, drift-diffusion, continuity, and carrier-statistics equations.

This chapter introduced the common quantum-corrected drift-diffusion formulation and the two closures used in this work. The Poisson-Schrödinger and density-gradient approaches account for quantum confinement and tunneling-related charge redistribution through quantum-corrected carrier densities and effective band edges, while preserving the semiclassical structure of the transport equations.

Together with the NEGF formulations introduced in Chapter 2, the Poisson-Schrödinger and density-gradient approaches complete the hierarchy of models considered in this work. They progressively reduce the computational cost by replacing microscopic quantum transport with quantum-corrected carrier densities and, ultimately, differential quantum potentials. This reduction allows larger structures to be simulated, at the expense of a less detailed spectral description.

In the following chapter, the Büttiker-probe NEGF, Poisson-Schrödinger drift-diffusion, and density-gradient drift-diffusion models are applied to the same InGaN/GaN LED structure and compared within a common physical problem. In particular, the analysis focuses on the high ideality factors and subthreshold leakage associated with trap-assisted tunneling (TAT). The Büttiker-probe formulation is used as the practical quantum-transport reference, while the two quantum-corrected drift-diffusion methods are assessed according to their ability to reproduce the quantum carrier densities, recombination rates, and electrical characteristics obtained from NEGF at a reduced computational cost. This comparison defines the hierarchy of transport models adopted throughout the remainder of this work.

Chapter 4

Trap-assisted tunneling modeling in NEGF and QCDD

GaN-based LEDs are among the most important technologies for high-efficiency solid-state lighting and optoelectronic applications. Despite their technological maturity, their performance is still limited by several non-ideal recombination mechanisms. One of the most relevant open issues is the efficiency droop, namely the decrease of the internal quantum efficiency when the current density is increased [42]. Since droop has often been correlated with the presence of defects, understanding defect-assisted nonradiative recombination is essential for improving the predictive capability of LED simulations and for optimizing device performance.

The effect of defects is particularly evident in the subthreshold regime, below the optical turn-on of the diode. In this voltage range, GaN-based LEDs often show ideality factors significantly larger than two, which cannot be explained by ideal diffusion or semiclassical recombination alone. This anomalous electrical behavior is commonly interpreted as a signature of trap-assisted tunneling (TAT), where carriers tunnel through the barrier and are captured by localized defect states before recombining nonradiatively. A schematic representation of the process is shown in Fig. 4.1. In this picture, the electric field enhances the access of carriers to the trap position by enabling tunneling into the gap, while multi-phonon relaxation provides the energy dissipation required for capture into the localized state. In conventional drift-diffusion simulations, TAT is usually included by modifying the Shockley-Read-Hall recombination rate through field-enhanced carrier lifetimes. Examples of these models are those based on the Hurkx [43] or Schenk [26] formulations or based on non-local path algorithms [38], which have been successfully used to reproduce experimental current-voltage characteristics of InGaN/GaN LEDs [44, 45]. However, these approaches remain essentially semiclassical. The quantum-mechanical tunneling process is

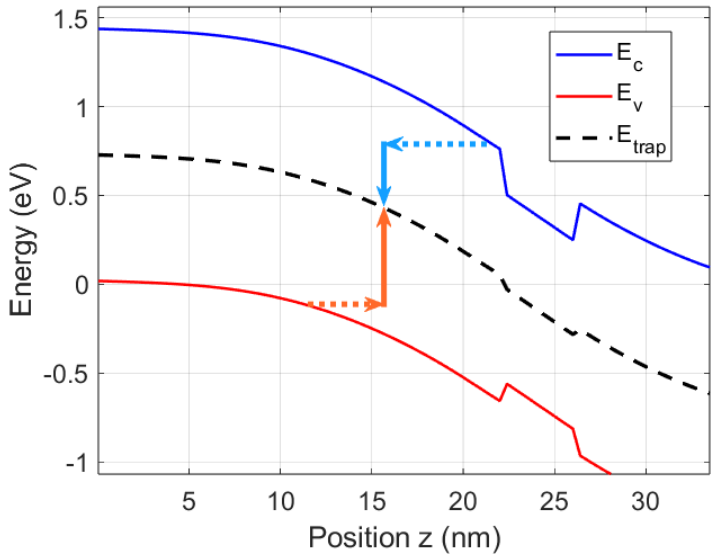


Fig. 4.1 Schematic representation of trap-assisted tunneling. Carriers can tunnel through the energy gap and relax into a localized defect state, opening a nonradiative recombination path that is not captured by a purely semiclassical SRH description.

condensed into empirical enhancement factors, which often depend on poorly known parameters such as tunneling masses. As a consequence, an agreement with experiments does not necessarily guarantee that the microscopic physics of the tunneling and recombination process is being correctly captured.

A more rigorous description can be obtained within the non-equilibrium Green's function (NEGF) formalism. In NEGF, confinement, tunneling, virtual tail states, and scattering-induced broadening can be described directly through Green's functions and self-energies. In particular, previous NEGF studies of TAT in InGaN/GaN LEDs have shown that the local density of states is the key quantity controlling the tunneling-assisted recombination process, and that the conventional SRH expression can reproduce the quantum recombination rate only if the classical carrier densities are replaced by quantum carrier densities evaluated from the spectral functions [28]. This result provides the physical basis for quantum-corrected drift-diffusion approaches: if the quantum density of states is evaluated consistently, an SRH-like recombination rate can still be used in a semiclassical continuity-equation framework.

The purpose of this chapter is to develop and compare different modeling strategies for TAT in GaN/InGaN LEDs. First, the microscopic NEGF formulation of defect-assisted recombination is reviewed in terms of multi-phonon defect self-energies, and its reduction to an SRH-like recombination rate suitable for the Büttiker-probe formulation described in Sec. 2.9.4 is discussed.

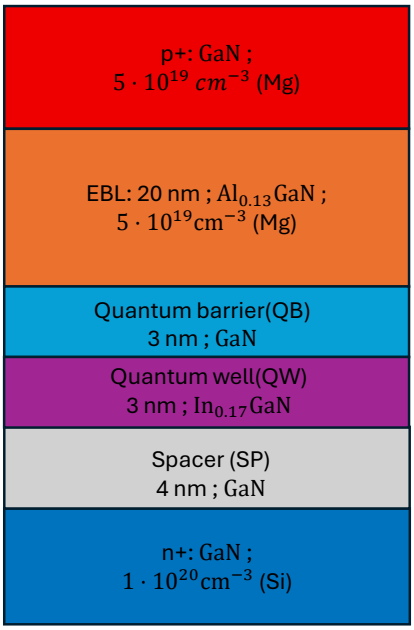


Fig. 4.2 Schematic layer structure of the visible GaN/InGaN LED used as benchmark for the comparison between NEGF, Poisson-Schrödinger drift-diffusion, and density-gradient drift-diffusion models.

The chapter then introduces trap-assisted tunneling and multi-phonon recombination into two quantum-corrected drift-diffusion formulations. In the Poisson-Schrödinger drift-diffusion model, the Schrödinger equation is solved self-consistently with the electrostatic problem, allowing the quantum carrier density and the spectral charge distribution to be reconstructed from the eigenstates. This makes it possible to define a spectral SRH recombination rate within a drift-diffusion framework. In the density-gradient drift-diffusion model, quantum effects are instead included through an effective quantum potential. In this case, the spectral charge is not directly available, but multi-phonon capture lifetimes can be derived by evaluating the capture integrals with a density-gradient reconstruction of the local density of states.
The three approaches are finally compared using a visible GaN/InGaN single-quantum-well LED, similar to the one described in [45]. The structure consists of an n+-GaN region doped with Si 10^{20}cm^{-3}, a 4 nm GaN spacer (SP), a 3 nm $In_{0.17}GaN$ quantum well (QW), a 3 nm GaN quantum barrier (QB), a 20 nm $Al_{0.13}GaN$ electron blocking layer (EBL), and a p-GaN region, Mg doped $5 \cdot 10^{19}\text{cm}^{-3}$. A schematic stack of the device can be found in Fig. 4.2. The main simulation parameters used are summarized in Table 4.1. The same parameters are used in the NEGF, Poisson-Schrödinger drift-diffusion, and density-gradient drift-diffusion calculations in order to make the comparison between the models meaningful. Spontaneous and piezoelectric polarization charges are included according to the Bernardini-Fiorentini model

for III-nitride materials [46]. The band offset ratio, carrier mobilities, radiative coefficient, Auger coefficient, Huang-Rhys factor, and phonon energy are chosen consistently with values commonly used in GaN/InGaN LED simulations [45, 38, 44]. The SRH bulk-like lifetimes are set to $\tau_n = \tau_p = 1$ ns [28]. The trap level is placed at midgap and a Gaussian broadening $E_\sigma = 50$ meV is used as a phenomenological description of energy distribution of the trap.

Table 4.1 Main material, transport, and recombination parameters used in the TAT simulations.

Parameter	**Value**
Electron SRH lifetime	$\tau_n = 1$ ns
Hole SRH lifetime	$\tau_p = 1$ ns
Trap energy position	Midgap
Trap density	$N_t = 10^{15}$ cm^{-3}
Trap-energy broadening	$E_\sigma = 50$ meV
Radiative coefficient	$B_{\text{rad}} = 2 \times 10^{-11}$ cm^{3}s^{-1}
Auger coefficient	$C = 2 \times 10^{-30}$ cm^{6}s^{-1}
Huang-Rhys factor	$S = 10$
Phonon energy	$\hbar\omega = 90$ meV
Electron mobility	$\mu_n = 300$ cm^{2}V^{-1}s^{-1}
Hole mobility	$\mu_p = 10$ cm^{2}V^{-1}s^{-1}

This device is used as a benchmark to clarify which physical information is retained or lost in each approach. NEGF provides the most complete description, being a genuine quantum kinetic method with native quantum transport and spectral currents, but it is computationally intensive. The Poisson-Schrödinger model retains an explicit spectral and quantum representation of the quantum charge and is therefore expected to reproduce the NEGF TAT behavior more accurately, even though transport is still semiclassical in nature. The density-gradient model is computationally cheaper, but relies on an approximate reconstruction of the quantum density of states.

4.1 Trap-assisted tunneling in the NEGF Framework

This section presents the treatment of trap-assisted tunneling within the Büttiker-probe NEGF framework. The multi-phonon defect interaction is first reduced to an SRH-like quantum recombination rate, after which the resulting formulation is applied to the InGaN/GaN LED.

4.1.1 Multi-phonon defect self-energies and SRH-like rate

For a deep-level defect, the energy separation between an extended state and the localized level is generally much larger than the energy of a single phonon. The transition must therefore be accompanied by the emission or absorption of several phonons and can be described through multi-phonon relaxation theory [26]. Within NEGF, this mechanism can be incorporated through the defect self-energy formulation developed in Refs. [27, 28].
The microscopic defect and reciprocal self-energies, together with the multi-phonon coupling elements and lineshape function, were introduced in Section 2.7.3. They provide a band-resolved description of carrier capture and emission through a localized defect state, and their derivation and physical interpretation are discussed in Refs. [27, 28, 19]. Their direct implementation is, however, computationally demanding. In a complete NEGF treatment, the conduction- and valence-band Green's functions determine the defect self-energies, while the defect Green's function determines the reciprocal self-energies acting on the extended states. These reciprocal self-energies must then be included again in the Dyson and Keldysh equations, modifying both the carrier occupations and the spectral properties of the device. In particular, their retarded components introduce an additional renormalization and broadening of the local density of states. The mutual dependence between the band Green's functions, the defect Green's function, and the corresponding self-energies therefore introduces an additional self-consistency loop and substantially increases the computational cost.
Nevertheless, it can be shown that the microscopic self-energy formulation can be reduced to a local recombination rate. This connection makes it possible to retain the quantum spectral information and the multi-phonon capture process while introducing defect-assisted recombination directly into the Büttiker-probe continuity equations,

$$\nabla \cdot \mathbf{J}_n = qU_{\mathrm{SRH},Q}, \tag{4.1a}$$

$$\nabla \cdot \mathbf{J}_p = -qU_{\mathrm{SRH},Q}. \tag{4.1b}$$

The Green's functions and quantum spectral densities are still obtained from the NEGF-Büttiker-probe calculation, but the feedback of the defect self-energies on the Dyson and Keldysh equations is neglected. This approximation therefore omits the additional renormalization and broadening produced specifically by the localized defects. For the dilute defect concentrations considered here, this correction is assumed to be secondary compared with the effect of the quantum spectral carrier distribution on the recombination rate. The comparison with the full self-energy calculations reported in Ref. [28] supports this approximation, since the microscopic NEGF recombination rate was reproduced well by an SRH-like expression evaluated using NEGF carrier

densities.

To establish the recombination rate entering Eq. (4.1), the relation between the microscopic defect self-energies and a local rate equation can be derived explicitly. All quantities are local in space, and the spatial dependence is initially omitted to simplify the notation. The net electron exchange rate between the conduction band and a single defect is

$$R_{dc} = \int_{-\infty}^{+\infty} \frac{dE}{2\pi\hbar} \left[\Sigma_{dc}^{<}(E) G_d^{>}(E) - \Sigma_{dc}^{>}(E) G_d^{<}(E) \right], \tag{4.2}$$

while the corresponding exchange rate between the defect and the valence band is

$$R_{dv} = \int_{-\infty}^{+\infty} \frac{dE}{2\pi\hbar} \left[\Sigma_{dv}^{>}(E) G_d^{<}(E) - \Sigma_{dv}^{<}(E) G_d^{>}(E) \right]. \tag{4.3}$$

For a localized defect state, the lesser and greater defect Green's functions can be written as

$$G_d^{<}(E) = i f_d A_d(E), \tag{4.4a}$$

$$G_d^{>}(E) = -i \left(1 - f_d\right) A_d(E), \tag{4.4b}$$

where f_d is the stationary defect occupation and $A_d(E)$ is the defect spectral function. For a sufficiently narrow defect level, the latter can be approximated as a Dirac delta function, $A_d(E) = 2\pi\delta(E - \varepsilon_d)$. The exchange rates then become

$$R_{dc} = \frac{1}{i\hbar} \left[\Sigma_{dc}^{<}(\varepsilon_d)(1 - f_d) + \Sigma_{dc}^{>}(\varepsilon_d) f_d \right], \tag{4.5a}$$

$$R_{dv} = -\frac{1}{i\hbar} \left[\Sigma_{dv}^{>}(\varepsilon_d) f_d + \Sigma_{dv}^{<}(\varepsilon_d)(1 - f_d) \right]. \tag{4.5b}$$

At steady state, the defect occupation is stationary and the two exchange rates are equal,

$$R_{dc} = R_{dv} = U_d, \tag{4.6}$$

where U_d is the net recombination rate associated with a single defect. For a defect density N_t, the corresponding volumetric rate is $U = N_t U_d$.

The four terms entering the exchange process can be identified as electron capture, electron

emission, hole capture, and hole emission contributions,

$$I_n^c \equiv \frac{N_t}{i\hbar}\Sigma_{dc}^{<}(\varepsilon_d), \qquad I_n^e \equiv -\frac{N_t}{i\hbar}\Sigma_{dc}^{>}(\varepsilon_d), \tag{4.7a}$$

$$I_p^c \equiv -\frac{N_t}{i\hbar}\Sigma_{dv}^{>}(\varepsilon_d), \qquad I_p^e \equiv \frac{N_t}{i\hbar}\Sigma_{dv}^{<}(\varepsilon_d). \tag{4.7b}$$

The rates can consequently be written as

$$U = I_n^c(1 - f_d) - I_n^e f_d = I_p^c f_d - I_p^e(1 - f_d). \tag{4.8}$$

Solving this relation for the stationary occupation gives

$$f_d = \frac{I_n^c + I_p^e}{I_n^c + I_n^e + I_p^c + I_p^e}, \tag{4.9}$$

and therefore the recombination rate can be written as

$$U = \frac{I_n^c I_p^c - I_n^e I_p^e}{I_n^c + I_n^e + I_p^c + I_p^e}. \tag{4.10}$$

Equation (4.10) is the local multi-phonon recombination rate obtained directly from the NEGF defect self-energies. Its algebraic structure follows from the stationary balance of the localized defect occupation, while the specific transport model enters through the four capture and emission contributions.

To express these contributions in terms of the carrier spectral distributions, the occupied electron and hole spectral densities and their complementary distributions are introduced as

$$n_c(z, E) = -\frac{i}{2\pi}\sum_{\mathbf{k}} G_c^{<}(z, z; \mathbf{k}, E), \qquad \overline{n}_c(z, E) = \frac{i}{2\pi}\sum_{\mathbf{k}} G_c^{>}(z, z; \mathbf{k}, E), \tag{4.11a}$$

$$p_v(z, E) = \frac{i}{2\pi}\sum_{\mathbf{k}} G_v^{>}(z, z; \mathbf{k}, E), \qquad \overline{p}_v(z, E) = -\frac{i}{2\pi}\sum_{\mathbf{k}} G_v^{<}(z, z; \mathbf{k}, E). \tag{4.11b}$$

Here, n_c describes occupied conduction-like states, $\overline{n}_c$ empty conduction-like states, p_v empty valence-like states, namely holes, and $\overline{p}_v$ occupied valence-like states. The energy-dependent electron and hole capture coefficients associated with the same multi-phonon interaction appearing

in the microscopic self-energies are

$$c_n(E,z) = \frac{2\pi}{\hbar}\sum_{l\geq 0} M_{dc}(l,z)\delta\left(E - \varepsilon_d(z) - l\hbar\omega_0\right), \tag{4.12a}$$

$$c_p(E,z) = \frac{2\pi}{\hbar}\sum_{l\geq 0} M_{dv}(l,z)\delta\left(E - \varepsilon_d(z) + l\hbar\omega_0\right), \tag{4.12b}$$

where $\varepsilon_d(z)$ is the local defect energy. The multi-phonon coupling elements, introduced in Section 2.7.3, are

$$M_{dc}(l,z) = M^0_{dc}(z)\frac{(S-l)^2}{S}L(l), \tag{4.13a}$$

$$M_{dv}(l,z) = M^0_{dv}(z)\frac{(S+l)^2}{S}L(l). \tag{4.13b}$$

Here, S is the Huang-Rhys factor, $L(l)$ is the multi-phonon lineshape function, and $M^0_{dc}(z)$ and $M^0_{dv}(z)$ are the local coupling prefactors, depending on the defect wavefunction, its localization radius, the impurity potential, and the electron-phonon coupling strength.
By substituting the microscopic self-energies and the spectral-density definitions in Eq. (4.11), and using the capture coefficients in Eq. (4.12), the sums over the phonon order can be recast as energy integrals using the Dirac delta functions. The four capture and emission contributions then become

$$I^c_n(z) = \int_{-\infty}^{+\infty} N_t(z)c_n(E,z)n_c(z,E)\,dE, \tag{4.14a}$$

$$I^e_n(z) = \int_{-\infty}^{+\infty} N_t(z)c_n(E,z)\exp\left(\frac{\varepsilon_d(z)-E}{k_BT}\right)\overline{n}_c(z,E)\,dE, \tag{4.14b}$$

$$I^c_p(z) = \int_{-\infty}^{+\infty} N_t(z)c_p(E,z)p_v(z,E)\,dE, \tag{4.14c}$$

$$I^e_p(z) = \int_{-\infty}^{+\infty} N_t(z)c_p(E,z)\exp\left(\frac{E-\varepsilon_d(z)}{k_BT}\right)\overline{p}_v(z,E)\,dE. \tag{4.14d}$$

Equations (4.14) contain the complete microscopic information on the capture and emission processes and make the connection with the SRH formulation explicit. Unlike the carrier densities entering the semiclassical bulk SRH formulation, they are not constructed only from states above the local conduction-band edge or below the local valence-band edge. The NEGF spectral functions describe an open quantum system and, in the presence of a sufficiently strong electric field, can contain virtual tail states extending into the nominal bandgap due to tunneling. Although these states lie in a classically forbidden energy range, they contribute to $n_c(z,E)$

and $p_v(z, E)$ and provide a spectral path between the extended bands and the localized defect. Their inclusion in the capture integrals in Eq. (4.14) represents the tunneling contribution to trap-assisted tunneling, while the energy-dependent coefficients $c_n(E, z)$ and $c_p(E, z)$ determine the corresponding multi-phonon capture probability.

The capture contributions can now be written in the same lifetime form appearing in the conventional SRH rate,

$$I_n^c(z) = \frac{n_c(z)}{\tau_{n,Q}(z)}, \qquad I_p^c(z) = \frac{p_v(z)}{\tau_{p,Q}(z)}, \tag{4.15}$$

provided that the effective quantum capture lifetimes are defined directly from the microscopic capture integrals,

$$\frac{1}{\tau_{n,Q}(z)} = \frac{I_n^c(z)}{n_c(z)} = \frac{1}{n_c(z)} \int_{-\infty}^{+\infty} N_t(z) c_n(E, z) n_c(z, E)\, dE, \tag{4.16a}$$

$$\frac{1}{\tau_{p,Q}(z)} = \frac{I_p^c(z)}{p_v(z)} = \frac{1}{p_v(z)} \int_{-\infty}^{+\infty} N_t(z) c_p(E, z) p_v(z, E)\, dE, \tag{4.16b}$$

where, $n_c(z)$ and $p_v(z)$ denote the total quantum electron and hole densities obtained by integrating $n_c(z, E)$ and $p_v(z, E)$ over energy. Equations (4.15) and (4.16) do not introduce an additional approximation, but simply rewrite the microscopic capture integrals in terms of the total quantum carrier densities and the corresponding energy-averaged capture probabilities. Since the multi-phonon capture coefficients depend strongly on energy, the effective lifetimes depend not only on the total carrier densities but also on how the spectral charge is distributed with respect to the defect level. The integration is performed over the complete quantum spectrum, so the lifetimes include capture from the usual extended states and from the virtual tail states extending into the gap. They are therefore local and bias-dependent quantities controlled by the electrostatic field, the LDOS, the carrier distribution, and the position of the defect level, and should not be identified directly with the zero-field bulk lifetimes used in conventional SRH models. The emission contributions can be related to the same effective lifetimes through the local nondegenerate detailed-balance relations. In particular,

$$\frac{I_n^e}{I_n^c} = \frac{n_1}{n_{B,\mathrm{cl}}}, \qquad \frac{I_p^e}{I_p^c} = \frac{p_1}{p_{B,\mathrm{cl}}}, \tag{4.17}$$

where n_1 and p_1 are the usual SRH auxiliary densities associated with the defect energy, while $n_{B,\mathrm{cl}}$ and $p_{B,\mathrm{cl}}$ are the corresponding local semiclassical Boltzmann densities. Introducing

$$\gamma_n(z) = \frac{n_c(z)}{n_{B,\mathrm{cl}}(z)}, \tag{4.18a}$$

$$\gamma_p(z) = \frac{p_v(z)}{p_{B,\mathrm{cl}}(z)}, \tag{4.18b}$$

the four capture and emission contributions take the SRH-like form

$$I_n^c = \frac{n_c}{\tau_{n,Q}}, \qquad I_n^e = \frac{\gamma_n n_1}{\tau_{n,Q}}, \tag{4.19a}$$

$$I_p^c = \frac{p_v}{\tau_{p,Q}}, \qquad I_p^e = \frac{\gamma_p p_1}{\tau_{p,Q}}. \tag{4.19b}$$

Substituting these expressions into the general recombination rate in Eq. (4.10) gives

$$U_{\mathrm{SRH},Q} = \frac{n_c p_v - \gamma_n n_1 \gamma_p p_1}{\tau_{p,Q}\,(n_c + \gamma_n n_1) + \tau_{n,Q}\,\left(p_v + \gamma_p p_1\right)}. \tag{4.20}$$

Equation (4.20) has the same algebraic structure as the conventional Shockley-Read-Hall expression, but its physical quantities are obtained from the complete quantum spectral distributions. In the homogeneous nondegenerate semiclassical bulk limit, the electron spectral density is restricted to states above E_C, the hole spectral density is restricted to states below E_V, and the virtual in-gap tails are absent. The full-spectrum capture integrals in Eqs. (4.16a) and (4.16b) then reduce to the corresponding band-limited bulk averages, n_c and p_v reduce to the semiclassical carrier densities, γ_n and γ_p tend to unity, and the effective quantum lifetimes reduce to the bulk SRH lifetimes. Equation (4.20) consequently reduces to the conventional recombination rate derived in the original Shockley-Read-Hall theory [39]. The detailed derivation connecting the band-limited capture and emission integrals to the conventional SRH lifetimes and auxiliary densities is reported in Appendix C.

This comparison also clarifies why replacing only the semiclassical densities by quantum carrier densities while retaining constant bulk lifetimes is not, in general, equivalent to the complete multi-phonon formulation. The virtual tail states modify the total charge entering the SRH-like numerator and denominator, but their contribution to capture is additionally weighted by the strongly energy-dependent multi-phonon coefficients. The full formulation therefore requires the carrier densities and the effective lifetimes to be evaluated consistently from the same quantum spectral distributions. Using n_c and p_v with fixed bulk lifetimes isolates the effect of the quantum charge, whereas using $\tau_{n,Q}$ and $\tau_{p,Q}$ retains the energy selectivity of the multi-phonon capture

process.

One unknown remaining is the microscopic prefactors M^0_{dc} and M^0_{dv}. These quantities are generally not known with sufficient accuracy to determine the prefactors directly. Following the procedure used in Ref. [28], they are instead calibrated by requiring the multi-phonon model to reproduce the bulk SRH lifetimes in the semiclassical zero-field limit, which are available experimentally for the materials.

For electrons, the calibration condition is

$$\int_{E_C(z)}^{+\infty} N_t(z) c_n(E,z) n_{B,\mathrm{cl}}(z,E)\, dE = \frac{n_{B,\mathrm{cl}}(z)}{\tau_{n,0}}, \tag{4.21}$$

where $\tau_{n,0}$ is the prescribed zero-field bulk electron lifetime. Substitution of Eq. (4.12a) gives

$$M^0_{dc}(z) = \frac{\hbar n_{B,\mathrm{cl}}(z)}{2\pi\tau_{n,0} N_t(z)} \left[\sum_{\substack{l\geq 0 \\ \varepsilon_d(z)+l\hbar\omega_0 \geq E_C(z)}} \frac{(S-l)^2}{S} L(l) n_{B,\mathrm{cl}}\left(z, \varepsilon_d(z) + l\hbar\omega_0\right) \right]^{-1}. \tag{4.22}$$

Analogously, the hole prefactor is determined from

$$\int_{-\infty}^{E_V(z)} N_t(z) c_p(E,z) p_{B,\mathrm{cl}}(z,E)\, dE = \frac{p_{B,\mathrm{cl}}(z)}{\tau_{p,0}}, \tag{4.23}$$

which gives

$$M^0_{dv}(z) = \frac{\hbar p_{B,\mathrm{cl}}(z)}{2\pi\tau_{p,0} N_t(z)} \left[\sum_{\substack{l\geq 0 \\ \varepsilon_d(z)-l\hbar\omega_0 \leq E_V(z)}} \frac{(S+l)^2}{S} L(l) p_{B,\mathrm{cl}}\left(z, \varepsilon_d(z) - l\hbar\omega_0\right) \right]^{-1}. \tag{4.24}$$

The restrictions on the sums arise because the calibration is performed in the semiclassical bulk limit: electron capture is restricted to states above E_C, while hole capture is restricted to states below E_V. Once M^0_{dc} and M^0_{dv} have been calibrated, however, the same capture coefficients are evaluated over the complete quantum spectral densities in Eqs. (4.16a) and (4.16b). The extension of the integration domain from the classical bands to the complete quantum spectrum includes the virtual in-gap spectral weight generated by tunneling and allows the calibrated bulk interaction to describe trap-assisted tunneling without introducing an empirical field-enhancement factor.

The same formulation provides an energy-resolved decomposition of the recombination process.

Defining the integrated electron and hole capture weights as

$$A_n(z) = \int_{-\infty}^{+\infty} N_t(z) c_n(E,z) n_c(z,E)\, dE = \frac{n_c(z)}{\tau_{n,Q}(z)}, \tag{4.25a}$$

$$A_p(z) = \int_{-\infty}^{+\infty} N_t(z) c_p(E,z) p_v(z,E)\, dE = \frac{p_v(z)}{\tau_{p,Q}(z)}, \tag{4.25b}$$

and introducing

$$\mathcal{D}(z) = \left[1 + \frac{n_1(z)}{n_{B,\mathrm{cl}}(z)}\right] A_n(z) + \left[1 + \frac{p_1(z)}{p_{B,\mathrm{cl}}(z)}\right] A_p(z), \tag{4.26}$$

the electron spectral recombination rate can be written as

$$U_{cn}(z,E) = \left[1 - \frac{n_1(z)p_1(z)}{n_{B,\mathrm{cl}}(z)p_{B,\mathrm{cl}}(z)}\right] \frac{N_t(z)c_n(E,z)n_c(z,E)A_p(z)}{\mathcal{D}(z)}, \tag{4.27}$$

while the corresponding hole spectral rate is

$$U_{cp}(z,E) = \left[1 - \frac{n_1(z)p_1(z)}{n_{B,\mathrm{cl}}(z)p_{B,\mathrm{cl}}(z)}\right] \frac{N_t(z)c_p(E,z)p_v(z,E)A_n(z)}{\mathcal{D}(z)}. \tag{4.28}$$

The local rate entering the continuity equation is obtained by integrating either spectral contribution,

$$U_{\mathrm{SRH},Q}(z) = \int_{-\infty}^{+\infty} U_{cn}(z,E)\, dE = \int_{-\infty}^{+\infty} U_{cp}(z,E)\, dE. \tag{4.29}$$

The equality of the two integrals follows from the stationary balance of the localized defect occupation. Substituting Eqs. (4.25a) and (4.25b) into Eq. (4.29) recovers the compact SRH-like expression in Eq. (4.20).

The complete multi-phonon defect self-energies therefore provide the microscopic reference description of trap-assisted recombination. In the NEGF-Büttiker-probe implementation used in this work, their feedback on the Dyson and Keldysh equations is neglected and recombination is introduced through Eq. (4.20) in the local continuity condition. Nevertheless, the carrier densities, effective capture lifetimes, and spectral recombination rates are evaluated from the quantum NEGF spectral functions using capture coefficients derived from the same multi-phonon interaction. The model therefore retains both ingredients of trap-assisted tunneling: virtual spectral states produced by tunneling in the nominal bandgap and their energy-selective capture through multi-phonon relaxation. At the same time, it avoids the additional self-consistency loop required by the full defect self-energy formulation.

4.1.2 Numerical results

In the previous subsection, it was shown that the microscopic multi-phonon self-energy description of defect-assisted recombination can be reduced to an SRH-like recombination rate, provided that the carrier densities and effective lifetimes are evaluated from the quantum spectral functions. This makes it possible to introduce defect-assisted recombination as a local rate term in the Büttiker-probe continuity equation, instead of solving the full defect self-energies self-consistently. We now apply this formulation to the GaN/InGaN visible LED in order to assess the effectiveness of the model and to clarify the microscopic origin of the TAT contribution. In particular, the NEGF calculation gives access to spectral quantities, such as the local density of states, which allow the tunneling paths associated with trap-assisted recombination to be identified directly. Figure 4.3 shows the local density of states computed at transverse wave vector $k_\perp = 0$ for two representative bias conditions, namely $V = 0.10$ V and $V = 2.55$ V. For each voltage, the two-dimensional LDOS map is reported together with a vertical cut at $z = 26$ nm, corresponding to the region close to the EBL/quantum-barrier side where trap-assisted recombination is particularly relevant. The two-dimensional LDOS maps highlight the band profile and the confined states in the active region, while the logarithmic horizontal scale used for the cuts makes it possible to resolve the weak spectral tails extending inside the nominal bandgap. At low bias, these in-gap spectral contributions are small but already visible around and beyond the trap energy, indicating the presence of virtual tail states that provide a tunneling-assisted path to the defect position. At higher forward bias, the modification of the band profile changes the energy distribution of these tails and their separation from the trap level. The dominant capture processes may therefore involve transitions of higher multi-phonon order, thereby modifying the corresponding effective capture lifetimes. The LDOS identifies the available quantum states, but it does not directly show which parts of the spectrum contribute to the defect-assisted capture process. This information is provided by the energy-resolved recombination rates introduced in the previous subsection. Figure 4.4 shows the spectral electron and hole contributions at low and high forward bias. The positive and negative features correspond to the two carrier channels according to the sign convention adopted for the spectral rate representation, while their energy integrals give the same local net recombination rate. At $V = 0.10$ V, the spectral recombination contribution is strongly localized close to the EBL/quantum-barrier region. The electron and hole side contributions occur on opposite sides of the trap energy and identify the spectral states connected through the multi-phonon defect transition. At $V = 2.55$ V, the spectral rate increases considerably as a consequence of the larger carrier injection, and its energy distribution follows the bias-induced modification of the band profile. In this regime, the quantum well is sufficiently populated for a

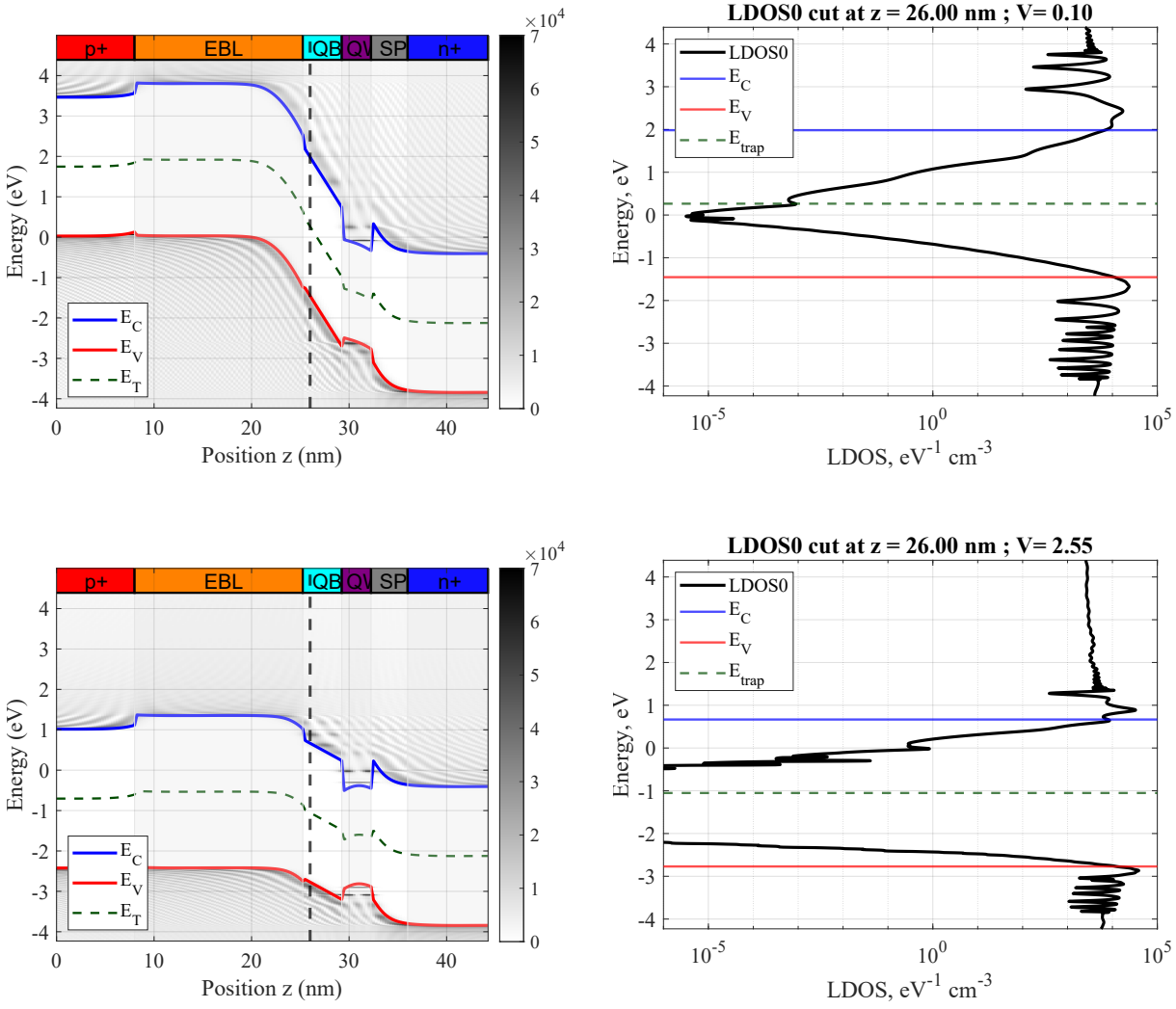


Fig. 4.3 Local density of states computed with NEGF at transverse wave vector $k_\perp = 0$ for two representative bias conditions. For each voltage, the two-dimensional LDOS map is shown together with a vertical cut at $z = 26$ nm. The cuts, plotted on a logarithmic horizontal scale, reveal the weak spectral tails extending inside the nominal bandgap. The conduction-band edge E_C (blue line), valence-band edge E_V (red line), and trap energy E_T (green dashed line) are also reported.

conventional SRH contribution to become visible inside the well. At the same time, the TAT contribution remains present inside the nominal bandgap, but its spatial maximum moves closer to the QB/QW region as the electric field and band bending are reduced.

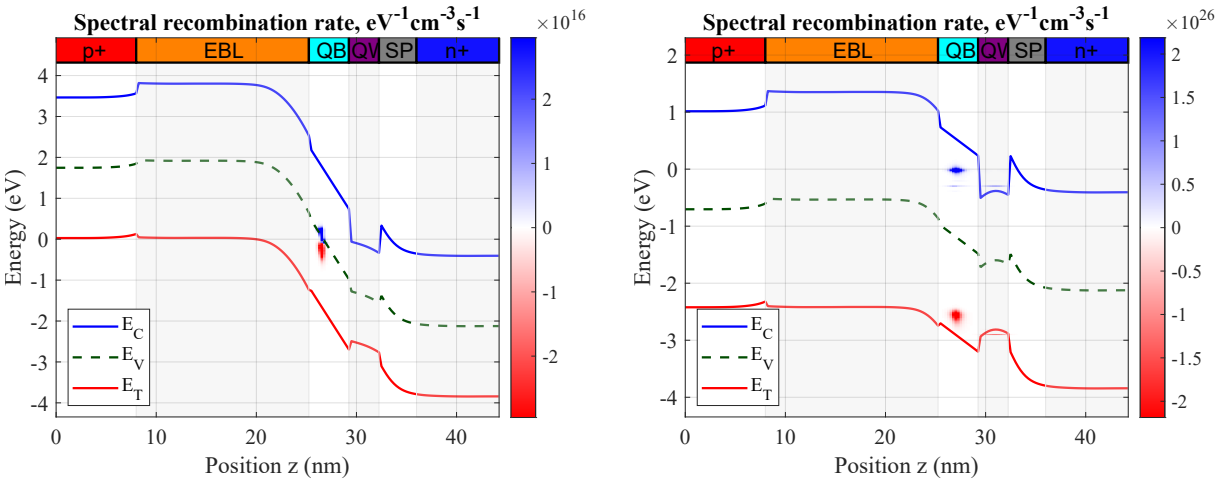


Fig. 4.4 Energy and position resolved multi-phonon recombination rate obtained from the NEGF spectral carrier densities. The blue regions represent the electron recombination contribution, while the red regions represent the hole recombination contribution. Left: $V = 0.10$ V. Right: $V = 2.55$ V.

The spatial recombination profiles provide a direct way to connect the spectral information contained in the LDOS with the local TAT rate. Figure 4.5 compares three recombination rates: the SRH rate evaluated with semiclassical carrier densities from a pure drift-diffusion simulation, in which conventional local SRH recombination is evaluated without quantum corrections to the carrier densities or any tunneling-enhancement model, the SRH rate evaluated with NEGF quantum carrier densities and constant bulk-like lifetimes, and the multi-phonon NEGF recombination rate. The pure drift-diffusion result therefore represents the conventional, non-tunneling SRH reference. This comparison separates the effect of the quantum carrier densities from the additional energy dependence introduced by the multi-phonon capture process. At low bias, the recombination associated with the quantum calculations is localized close to the EBL/QB side of the active region. In this regime, the strong electric field allows carriers to tunnel into the gap and reach the defect-assisted recombination channel. As a result, both quantum approaches give a recombination rate much larger than the semiclassical SRH rate, which does not include tunneling-induced carrier populations. The difference between the SRH rate with quantum charges and bulk-like lifetimes and the multi-phonon NEGF rate shows that the total quantum charge is not the only relevant quantity. In the multi-phonon formulation, the capture

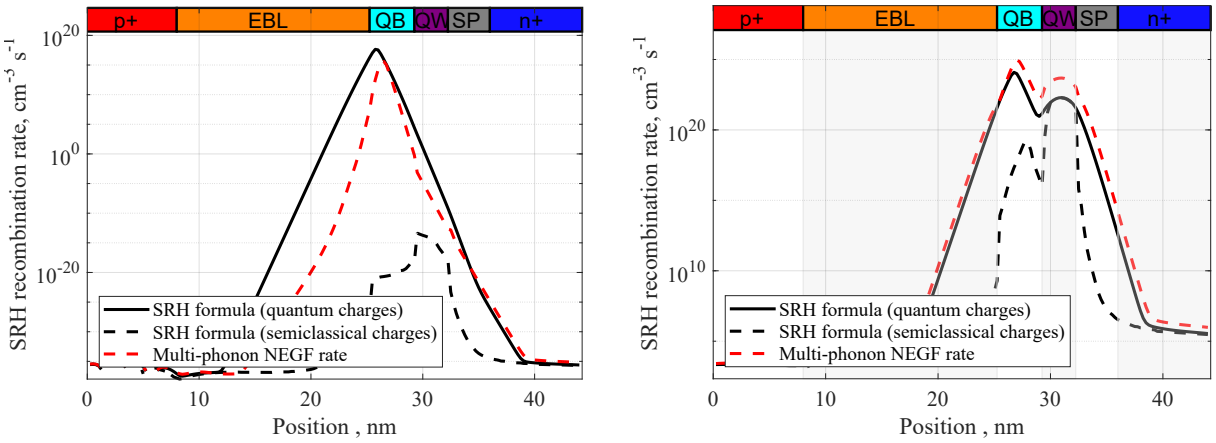


Fig. 4.5 Comparison between the multi-phonon NEGF recombination rate (red dashed lines) and SRH rates evaluated either with quantum carrier densities and constant bulk-like lifetimes (black solid lines) or with semiclassical carrier densities from a pure drift-diffusion simulation (black dashed lines). Left: $V = 0.10$ V. Right: $V = 2.55$ V.

process also depends on the energy alignment with the trap level and on the phonon-assisted transition probabilities. Therefore, part of the tunneling-related spectral weight may not contribute to the selected trap-assisted capture process, even though it is present in the quantum LDOS.
At high forward bias, shown here for $V = 2.55$ V, carrier injection into the quantum well becomes much stronger and the recombination peak shifts toward the well region. In this regime, the SRH expression evaluated with quantum carrier densities and constant bulk-like lifetimes reproduces the multi-phonon NEGF recombination rate with good accuracy. This indicates that, around the LED working point, the dominant effect of the quantum description is the correction of the carrier densities entering the recombination rate. In contrast, the semiclassical SRH rate still underestimates the recombination because it does not include the tunneling-induced quantum charge. The impact of these recombination rates on the electrical characteristics is shown in Fig. 4.6. The current density and ideality factor are computed using the semiclassical SRH rate, the SRH rate evaluated with NEGF quantum carrier densities and constant bulk-like lifetimes, and the multi-phonon NEGF recombination rate. The semiclassical SRH current is several orders of magnitude lower than the quantum IV over most of the subthreshold voltage range. This is the direct consequence of the absence of tunneling-related carrier populations in the pure drift-diffusion description. Since the semiclassical carrier densities are confined by the classical band edges, the overlap between electrons, holes, and trap states in the forbidden region remains strongly suppressed. The corresponding ideality factor remains close to the value expected from conventional depletion-region SRH recombination over a broad bias range.

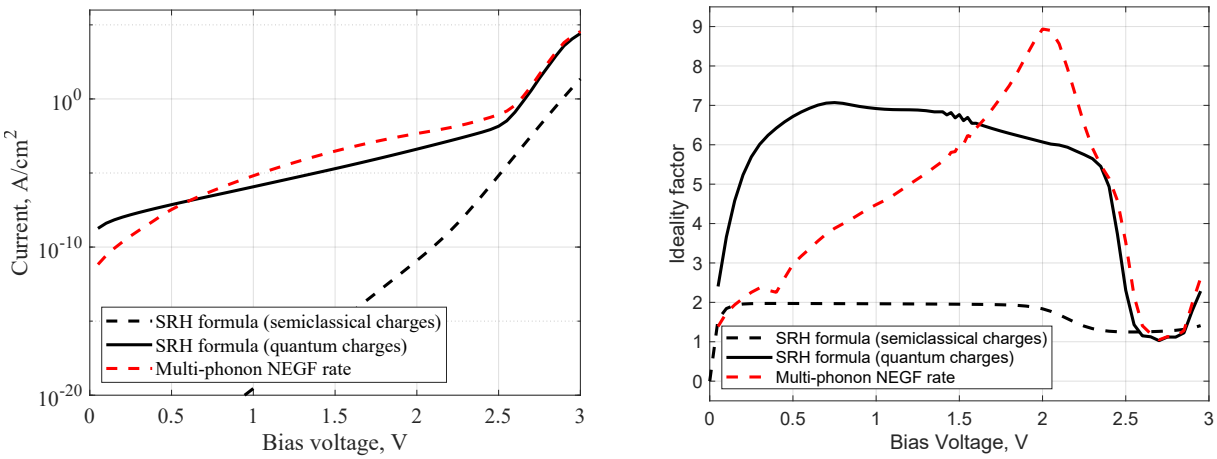


Fig. 4.6 Current-voltage characteristics (left) and ideality factor (right) obtained from the semiclassical SRH rate (black dashed lines), from the SRH formula evaluated with NEGF quantum carrier densities and constant bulk-like lifetimes (black solid lines), and from the multi-phonon NEGF recombination rate (red dashed lines).

When the SRH formula is evaluated with NEGF quantum carrier densities and constant bulk-like lifetimes, the current increases strongly at low and intermediate bias. This confirms that a large part of the TAT enhancement originates from the quantum carrier densities themselves. The corresponding ideality factor is much larger than two in the subthreshold regime, reflecting the transition from conventional SRH recombination to a tunneling-assisted recombination process. At these voltages, the dominant recombination region is located close to the EBL/QB side of the active region, where the strong electric field promotes tunneling into the gap.

The multi-phonon NEGF rate produces a similar current enhancement, but with a different voltage dependence. Since the effective capture rates are obtained from spectral multi-phonon transition integrals, the recombination process depends not only on the total quantum charge but also on how this charge is distributed in energy with respect to the trap level. This introduces an additional field dependence in the effective lifetimes. Consequently, the ideality factor obtained from the multi-phonon rate is not constant: it increases through the low-bias TAT-dominated regime and then decreases when forward injection becomes strong enough to populate the quantum well.

This change of regime occurs when hole injection into the active region becomes significant. Once the quantum well is sufficiently populated, recombination inside the well becomes increasingly dominant with respect to the TAT contribution localized near the EBL/QB region. The decrease of the ideality factor at higher forward bias is therefore associated with the transition from field-assisted trap recombination in the barrier region to recombination dominated by the injected carrier population in the quantum well. In this regime, the SRH expression with quantum

carrier densities and constant bulk-like lifetimes and the multi-phonon NEGF rate become closer, consistently with the spatial recombination profiles shown in Fig. 4.5.
Overall, these NEGF results support the use of an SRH-like recombination rate within the Büttiker-probe continuity equation, provided that the carrier densities are evaluated from the quantum spectral functions. The LDOS shows the formation of virtual tail states, and the rate comparison shows that the SRH formula with quantum carrier densities and constant bulk-like lifetimes can reproduce the multi-phonon NEGF recombination rate in the forward-bias regime where quantum-well recombination dominates. At the same time, the multi-phonon formulation provides a more detailed description of the low-bias TAT regime, where the energy dependence of the capture process and the field dependence of the effective lifetimes become important.

4.2 TAT in quantum-corrected drift-diffusion formalism: Poisson-Schrödinger

This section extends the same multi-phonon recombination framework presented in the previous section to Poisson-Schrödinger drift-diffusion. The formulation is first expressed in terms of the Poisson-Schrödinger quantum spectral densities and is then assessed through the corresponding device simulations.

4.2.1 Multi-phonon SRH formulation with Poisson-Schrödinger quantum densities

The previous section showed that the multi-phonon defect self-energy formulation can be reduced to an SRH-like recombination rate expressed in terms of quantum carrier densities and effective capture lifetimes. The same formulation can be applied within Poisson-Schrödinger drift-diffusion, since this model provides an energy-resolved quantum LDOS from which the spectral electron and hole densities can be reconstructed. The multi-phonon capture coefficients, their calibration from the zero-field bulk lifetimes, and the expressions derived for the effective lifetimes and recombination rates remain unchanged. The difference lies exclusively in the quantum spectral densities used to evaluate them.
In the Poisson-Schrödinger model, the conduction- and valence-band LDOS are obtained from the eigenvalues and eigenfunctions of the corresponding Schrödinger problems, as discussed in

Section 3.2. The energy-resolved electron and hole densities can therefore be written as

$$n_Q^{\mathrm{PS}}(z,E) = LDOS_c^{\mathrm{PS}}(z,E) f\left(\frac{E - E_{F_n}(z)}{k_B T}\right), \tag{4.30a}$$

$$p_Q^{\mathrm{PS}}(z,E) = LDOS_v^{\mathrm{PS}}(z,E) \left[1 - f\left(\frac{E - E_{F_p}(z)}{k_B T}\right)\right], \tag{4.30b}$$

where E_{F_n} and E_{F_p} are the local electron and hole quasi-Fermi levels determined by the drift-diffusion solution. Their integration over energy gives the total quantum carrier densities,

$$n_Q^{\mathrm{PS}}(z) = \int_{-\infty}^{+\infty} n_Q^{\mathrm{PS}}(z,E)\, dE, \qquad p_Q^{\mathrm{PS}}(z) = \int_{-\infty}^{+\infty} p_Q^{\mathrm{PS}}(z,E)\, dE. \tag{4.31}$$

The discrete eigenstate spectrum is represented on the numerical energy grid using the normalized broadening function introduced in Section 3.2. The normalization ensures that Eq. (4.31) recovers the quantum charge calculated directly from the Schrödinger eigenstates.

The effective electron and hole capture lifetimes are then obtained from Eqs. (4.16a) and (4.16b) by replacing the NEGF spectral densities with their Poisson-Schrödinger counterparts, together with the corresponding replacement of the integrated densities. The energy-dependent capture coefficients $c_n(E,z)$ and $c_p(E,z)$ remain those derived from the multi-phonon interaction in the previous section. Similarly, the prefactors M_{dc}^0 and M_{dv}^0 are determined through the same calibration procedure based on the prescribed zero-field semiclassical lifetimes.

The local recombination rate is consequently obtained from Eq. (4.20) using the Poisson-Schrödinger carrier densities, effective lifetimes, and correction factors. The algebraic structure of the SRH expression is unchanged, since the localized-defect occupation and the balance between electron and hole capture processes do not depend on the method used to calculate the quantum carrier spectrum.

The same substitution can be applied to the energy-resolved electron and hole recombination rates derived in Eqs. (4.27) and (4.28). The Poisson-Schrödinger spectral rates are therefore obtained by combining the multi-phonon capture coefficients with $n_Q^{\mathrm{PS}}(z,E)$ and $p_Q^{\mathrm{PS}}(z,E)$. Their integration over energy gives the local recombination rate:

$$U_{\mathrm{SRH},Q}^{\mathrm{PS}}(z) = \int_{-\infty}^{+\infty} U_{cn}^{\mathrm{PS}}(z,E)\, dE = \int_{-\infty}^{+\infty} U_{cp}^{\mathrm{PS}}(z,E)\, dE. \tag{4.32}$$

The difference between the NEGF and Poisson-Schrödinger implementations of multi-phonon SRH and TAT is therefore entirely contained in the description of the quantum spectrum. In NEGF, the spectral densities are obtained from the Green's functions of an open transport

problem and naturally include the coupling to the contacts, coherent coupling between different spatial regions, and the broadening produced by the interaction self-energies retained in the calculation. In the Poisson-Schrödinger approach, the LDOS is instead reconstructed from the discrete eigenstates of a Hermitian Schrödinger problem, while their occupation is determined locally through the drift-diffusion quasi-Fermi levels.
The Poisson-Schrödinger formulation therefore retains quantum confinement and the penetration of the eigenfunctions into classically forbidden spatial regions, which can modify both the local carrier densities and the effective defect-capture lifetimes. However, it does not reproduce the complete open-system spectral broadening obtained within NEGF. The spectral recombination rates calculated with the two models are based on the same multi-phonon defect theory, but may differ because of the different LDOS and carrier distributions predicted by the two quantum descriptions.

4.2.2 Numerical results

The spectral SRH formulation described in the previous subsection is now applied to the InGaN/GaN visible LED benchmark. The goal is to verify whether the Poisson-Schrödinger drift-diffusion model can reproduce the main features observed in the NEGF analysis while remaining within a computationally lighter drift-diffusion framework. As in the previous section, three recombination descriptions are compared: the SRH rate evaluated with semiclassical carrier densities obtained from a pure drift-diffusion simulation, the SRH rate evaluated with quantum carrier densities and constant bulk-like lifetimes, and the multi-phonon recombination rate obtained from the Poisson-Schrödinger spectral carrier densities.
Figure 4.7 shows the $k_\perp = 0$ LDOS obtained from the Poisson-Schrödinger calculation at $V = 0.10$ V and $V = 2.55$ V, together with the semiclassical and quantum-corrected band edges. A phenomenological Gaussian broadening of 20 meV has been considered. With this broadening, the PSDD LDOS provides a spectral picture comparable to the NEGF LDOS discussed in Sec. 4.1.2. The confined states in the quantum well are clearly visible, and the broadened spectral weight extends into the nominally forbidden region where tunneling-assisted capture can occur. At low bias, the electric field across the junction is large and the quantum-corrected band edges differ strongly from the semiclassical ones. This produces a pronounced extension of the effective quantum bands inside the gap, especially across the EBL/QB side of the active region, where the TAT contribution is expected to be most relevant. From the LDOS and the corresponding quantum potential, one can observe how the Poisson-Schrödinger model introduces tunneling-related states inside the gap, which are then included in the drift-diffusion simulation

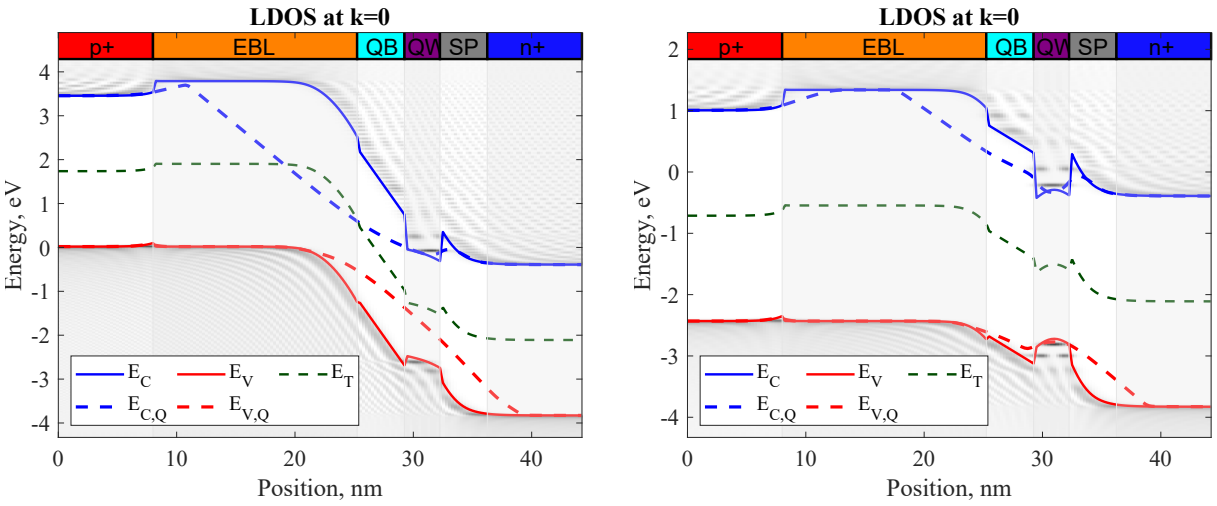


Fig. 4.7 $k_\perp = 0$ local density of states obtained from the Poisson-Schrödinger calculation at $V = 0.10$ V (left) and $V = 2.55$ V (right). The semiclassical band edges (solid lines) and the quantum-corrected band edges (dashed lines) are also reported. A Gaussian broadening of 20 meV is used to phenomenologically include scattering and contact-induced broadening effects that are not present in the closed-boundary Schrödinger solution.

through the effective quantum band edges. At $V = 2.55$ V, the forward bias reduces the electric field across the junction and the tunneling contribution inside the gap becomes less pronounced. The LDOS is then dominated more clearly by the confined states around the quantum well and by the local quantum correction near the QB/QW/SP region. This is consistent with the transition, discussed before, from a low-bias regime dominated by field-assisted tunneling near the EBL/QB region to a higher-bias regime where carrier injection fills the quantum well and recombination inside the well becomes dominant. The corresponding recombination profiles are shown in Fig. 4.8. Here, by SRH rate with quantum carrier densities, we mean that the standard SRH expression is retained using constant bulk-like lifetimes, without introducing empirical field-dependent lifetime enhancement factors, while the carrier densities are obtained from the Poisson-Schrödinger spectral charge. This comparison separates the effect of the quantum carrier densities from the additional energy dependence introduced by the multi-phonon capture process. At $V = 0.10$ V, the quantum-corrected rates are strongly enhanced with respect to the semiclassical SRH rate. The recombination is localized close to the EBL/QB side of the active region, where the electric field is large and tunneling through the classically forbidden region is expected to be most relevant. In this regime, the SRH formula with quantum charges and constant bulk-like lifetimes already gives a much larger recombination rate than the semiclassical calculation, showing that the main enhancement originates from the quantum carrier density.

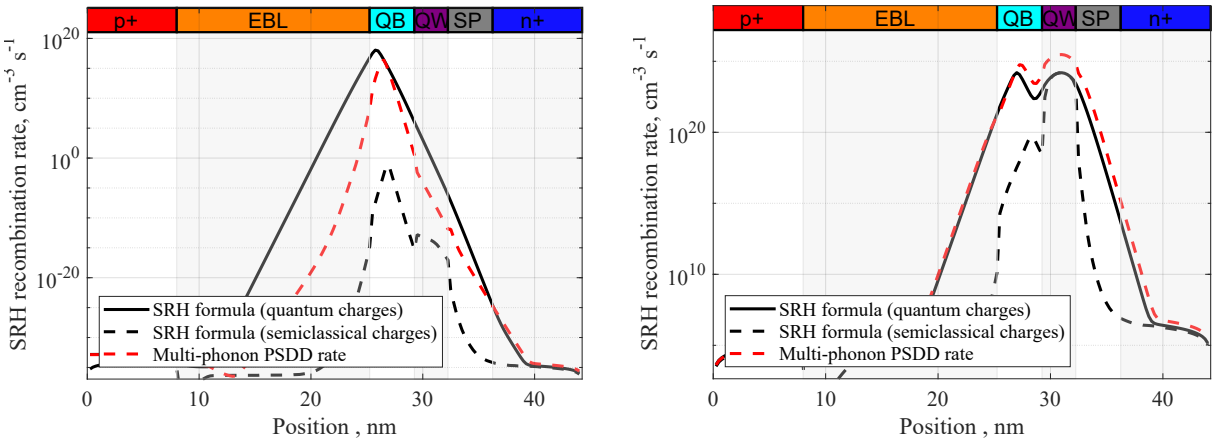


Fig. 4.8 Comparison between the SRH recombination rate evaluated with semiclassical carrier densities (black dashed lines), the SRH recombination rate evaluated with Poisson-Schrödinger quantum-corrected carrier densities and constant bulk-like lifetimes (black solid lines), and the multi-phonon PSDD recombination rate (red dashed lines) at $V = 0.10$ V (left) and $V = 2.55$ V (right).

However, the multi-phonon rate does not simply coincide with the quantum-charge SRH rate. As in the NEGF case, this indicates that the total quantum charge is not the only relevant quantity: the capture process also depends on the energy alignment with the trap level and on the phonon-assisted transition probabilities. At $V = 2.55$ V, carrier injection into the active region becomes much stronger and recombination in the quantum well becomes dominant. In this regime, the SRH rate evaluated with quantum carrier densities and constant bulk-like lifetimes reproduces the multi-phonon recombination rate with good accuracy over most of the active region. This confirms that, also in the Poisson-Schrödinger framework, the dominant effect near the LED working point is the correction of the carrier densities entering the recombination rate. In contrast, the semiclassical SRH rate still underestimates the recombination because it does not include the tunneling-induced quantum charge. The energy-resolved multi-phonon recombination rates obtained from the Poisson-Schrödinger spectral densities are shown in Fig. 4.9. The same spectral rate formulation was previously applied to the NEGF carrier densities in Sec. 4.1.2. The present results therefore provide a direct assessment of how the different quantum descriptions of the LDOS and carrier populations affect the energy and spatial localization of the recombination process. At $V = 0.10$ V, the recombination is concentrated close to the EBL/QB region, where the strong electric field produces spectral weight inside the nominal bandgap. The electron- and hole-side contributions are located on opposite sides of the trap level and identify the states connected by the multi-phonon capture process. This behavior is qualitatively consistent with the NEGF result and confirms that the Poisson-Schrödinger spectral densities reproduce the

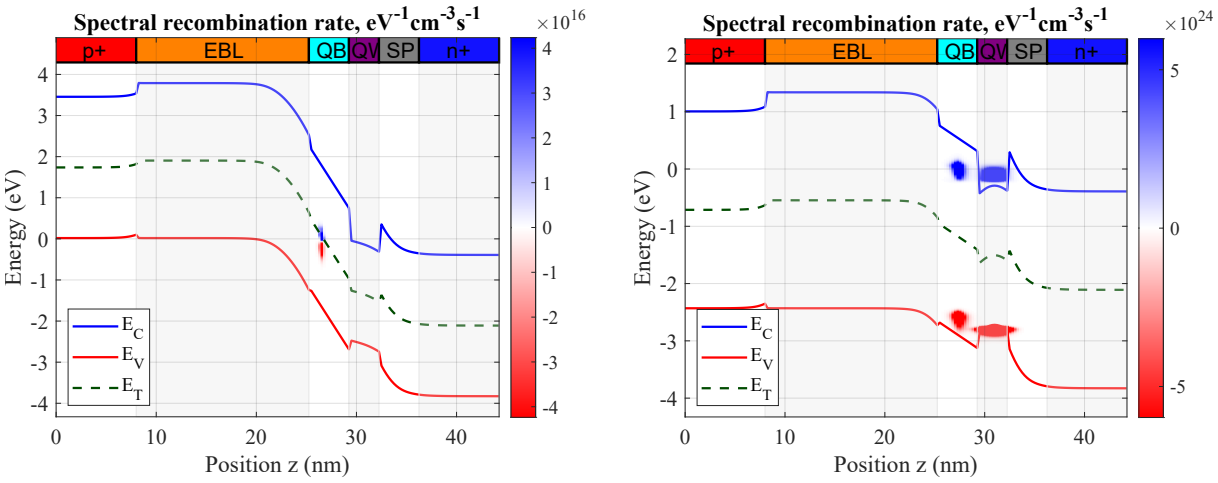
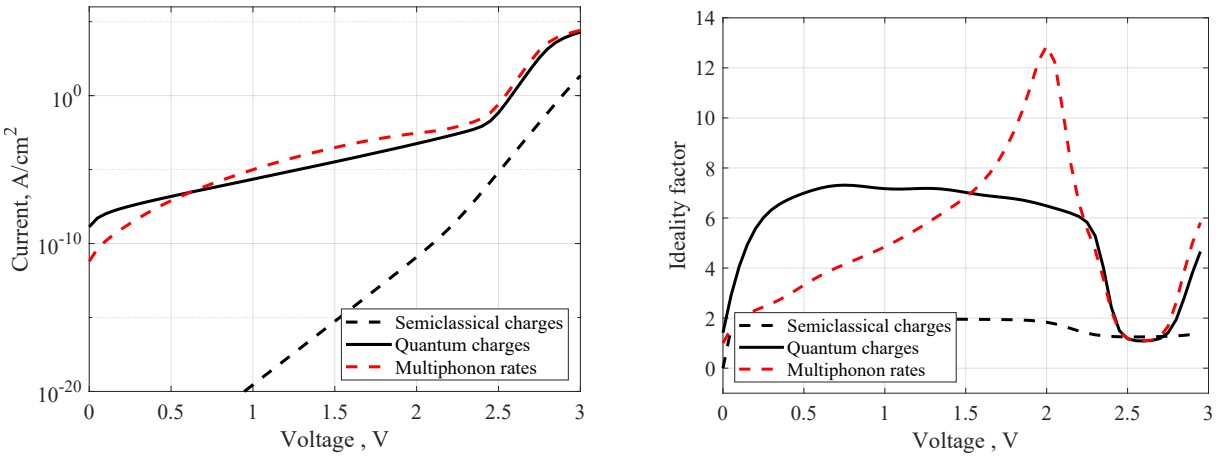


Fig. 4.9 Energy and position resolved multi-phonon recombination rate obtained from the Poisson-Schrödinger spectral carrier densities. The blue regions represent the electron recombination contribution, while the red regions represent the hole recombination contribution. Left: $V = 0.10$ V. Right: $V = 2.55$ V

main tunneling path responsible for the low-bias TAT current. At $V = 2.55$ V, the larger carrier injection produces a strong recombination contribution inside the quantum well. A weaker TAT contribution remains visible inside the gap near the QB/QW region, where the residual band bending still allows carriers to access the defect channel. The impact of the different recombination rates on the integrated electrical characteristics is shown in Fig. 4.10. The current-voltage characteristics and the ideality factor are computed using the semiclassical SRH rate, the SRH rate evaluated with Poisson-Schrödinger quantum carrier densities and constant bulk-like lifetimes, and the multi-phonon Poisson-Schrödinger recombination rate. As in the NEGF section, no comparison with experimental data is included here; the purpose is to isolate the effect of the quantum carrier densities and of the multi-phonon formulation within the PSDD model. The comparison with experiment and with the other modeling approaches is deferred to Sec. 4.4.

The current-voltage curves show that both quantum-corrected descriptions give a much larger current than the semiclassical SRH calculation over most of the subthreshold and intermediate-bias range. This confirms that the main increase of the recombination current is due to the quantum carrier densities, which include tunneling-related contributions absent in the purely semiclassical drift-diffusion solution. The multi-phonon rate gives an additional enhancement with respect to the SRH formula with quantum charges and constant bulk-like lifetimes, especially in the voltage range where trap-assisted tunneling is expected to dominate. The corresponding ideality factor shows the same qualitative behavior already observed in the NEGF analysis. At



(a) Current density.

(b) Ideality factor.

Fig. 4.10 Current-voltage characteristics (left) and ideality factor (right) obtained from the semiclassical SRH rate (black dashed lines), from the SRH formula evaluated with Poisson-Schrödinger quantum carrier densities and constant bulk-like lifetimes (black solid lines), and from the multi-phonon Poisson-Schrödinger recombination rate (red dashed lines).

low and intermediate bias, the recombination is mainly localized near the EBL/QB region, where the electric field allows carriers to tunnel into the gap and recombine through the defect-assisted channel. In this regime, the ideality factor becomes much larger than two, consistently with a recombination process that cannot be described by conventional semiclassical SRH theory. At higher forward bias, the injection of carriers into the quantum well increases and recombination inside the well becomes dominant. This produces a reduction of the ideality factor, marking the transition from a TAT-dominated regime to a regime controlled mainly by carrier injection and recombination in the quantum well.

It should be noted that the ideality-factor peak obtained from the multi-phonon PSDD rate is more pronounced than in the NEGF calculation. Since the ideality factor is extracted from the derivative of the logarithm of the current, it is particularly sensitive to small numerical variations of the current, to the voltage discretization, and to the transition between two recombination regimes. Therefore, the absolute value of the peak should be interpreted with some caution. Nevertheless, the qualitative behavior remains meaningful: the multi-phonon PSDD formulation produces a strong field-dependent TAT contribution at intermediate bias, followed by a decrease of the ideality factor when quantum-well recombination becomes dominant. These results show that, although the Poisson-Schrödinger model remains an approximation to the full NEGF description, it retains the most important information needed to reproduce the main features of TAT. The broadened LDOS gives a physically meaningful representation of confined and tunneling-related states, the integrated rates reproduce the multi-phonon recombination behavior

in the forward-bias regime, and the spectral recombination maps provide direct access to the position and energy of the recombination process. The IV and ideality-factor curves further show that PSDD captures the transition from a low-bias TAT-dominated regime to a higher-bias regime dominated by recombination in the populated quantum well. This supports the use of PSDD as a practical quantum-corrected alternative to NEGF, provided that the phenomenological broadening and the numerical sensitivity of the ideality-factor extraction are kept in mind.

4.3 TAT in quantum-corrected drift-diffusion formalism: density-gradient

This section adapts the multi-phonon TAT model to density-gradient drift-diffusion. Effective capture lifetimes are first derived from auxiliary spectral distributions associated with the quantum-corrected band edges, and the resulting approximation is then evaluated numerically.

4.3.1 Multi-phonon capture lifetimes in the density-gradient formalism

Unlike NEGF and Poisson-Schrödinger drift-diffusion, the density-gradient model does not provide an explicitly calculated energy-resolved LDOS. Quantum effects are instead introduced through differential corrections to the conduction and valence band profiles, which modify the integrated carrier densities but do not directly provide their decomposition over energy. Consequently, the spectral recombination rates introduced in the previous sections cannot be evaluated directly within the density-gradient formalism.
Nevertheless, the same multi-phonon SRH formulation can be retained. The quantum-corrected carrier densities obtained from the density-gradient solution are used in the SRH-like rate derived in Eq. (4.20), while the corresponding effective capture lifetimes are evaluated by constructing auxiliary local spectral distributions from the density-gradient corrected band edges. The effective electron and hole capture lifetimes are defined through the same relations introduced for NEGF:

$$\int_{-\infty}^{+\infty} N_t c_n(E,z) n(E,z)\, dE = \frac{n_Q(z)}{\tau_{n,Q}^{\mathrm{DG}}(z)}, \qquad \int_{-\infty}^{+\infty} N_t c_p(E,z) p(E,z)\, dE = \frac{p_Q(z)}{\tau_{p,Q}^{\mathrm{DG}}(z)}. \quad (4.33)$$

Here, n_Q and p_Q are the quantum-corrected densities obtained self-consistently from the density-gradient drift-diffusion solution. The problem is therefore reduced to the evaluation of $\tau_{n,Q}^{\mathrm{DG}}$ and $\tau_{p,Q}^{\mathrm{DG}}$. Using the convention introduced in Section 3.3, the density-gradient correction for

electrons is expressed as a shift of the conduction-band edge,

$$E_C^{\mathrm{DG}}(z) = E_C(z) - q\Lambda_n(z). \tag{4.34}$$

A local bulk-like density of states can then be associated with the corrected band edge:

$$LDOS_n^{\mathrm{DG}}(E,z) = \frac{1}{2\pi^2}\left(\frac{2m_n^*}{\hbar^2}\right)^{3/2}\sqrt{E - E_C(z) + q\Lambda_n(z)}, \tag{4.35}$$

for $E \geq E_C^{\mathrm{DG}}$, while it vanishes below the corrected conduction-band edge. In the Boltzmann approximation, the auxiliary electron spectral density is therefore

$$n(E,z) \simeq f_n(E)LDOS_n^{\mathrm{DG}}(E,z) = \exp\left(\frac{F_n - E}{k_BT}\right)LDOS_n^{\mathrm{DG}}(E,z). \tag{4.36}$$

The multi-phonon electron capture coefficient derived in the NEGF formulation is

$$c_n(E,z) = \frac{2\pi}{\hbar}\sum_{l\geq 0} M_{dc}^0(z)\frac{(S-l)^2}{S}L(l)\delta\left(E - \varepsilon_d(z) - l\hbar\omega_0\right). \tag{4.37}$$

The corresponding electron capture integral becomes

$$\int_{-\infty}^{+\infty} N_t c_n(E,z)n(E,z)\,dE = N_t\int_{-\infty}^{+\infty} c_n(E,z)\exp\left(\frac{F_n - E}{k_BT}\right)LDOS_n^{\mathrm{DG}}(E,z)\,dE. \tag{4.38}$$

Substitution of Eqs. (4.35) and (4.37), followed by integration using the Dirac delta, gives

$$\begin{aligned}\int_{-\infty}^{+\infty} N_t c_n(E,z)n(E,z)\,dE = {} & N_t\frac{2\pi}{\hbar}M_{dc}^0(z)\frac{1}{2\pi^2}\left(\frac{2m_n^*}{\hbar^2}\right)^{3/2} \\ & \times\sum_{l\geq 0}\frac{(S-l)^2}{S}L(l)\exp\left(\frac{F_n - \varepsilon_d(z) - l\hbar\omega_0}{k_BT}\right) \\ & \times\sqrt{\varepsilon_d(z) + l\hbar\omega_0 - E_C(z) + q\Lambda_n(z)}.\end{aligned} \tag{4.39}$$

The total density obtained within the same local Boltzmann approximation is

$$n_Q(z) = N_C\exp\left(\frac{F_n - E_C(z) + q\Lambda_n(z)}{k_BT}\right). \tag{4.40}$$

Using Eq. (4.40) to eliminate the quasi-Fermi level, the capture integral can be written as

$$\int_{-\infty}^{+\infty} N_t c_n(E,z) n(E,z)\, dE = n_Q(z) N_t \frac{8}{h}\left(\frac{\pi}{k_B T}\right)^{3/2} M_{dc}^0(z) \times \sum_{l\geq 0} \frac{(S-l)^2}{S} L(l) \exp\left(\frac{E_C(z) - q\Lambda_n(z) - \varepsilon_d(z) - l\hbar\omega_0}{k_B T}\right) \times \sqrt{\varepsilon_d(z) + l\hbar\omega_0 - E_C(z) + q\Lambda_n(z)}. \tag{4.41}$$

Comparison with Eq. (4.33) gives the density-gradient electron lifetime:

$$\frac{1}{\tau_{n,Q}^{\mathrm{DG}}(z)} = N_t \frac{8}{h}\left(\frac{\pi}{k_B T}\right)^{3/2} M_{dc}^0(z) \times \sum_{l\geq 0} \frac{(S-l)^2}{S} L(l) \exp\left(\frac{E_C(z) - q\Lambda_n(z) - \varepsilon_d(z) - l\hbar\omega_0}{k_B T}\right) \times \sqrt{\varepsilon_d(z) + l\hbar\omega_0 - E_C(z) + q\Lambda_n(z)}. \tag{4.42}$$

The prefactor M_{dc}^0 is determined through the same zero-field bulk calibration introduced in the NEGF formulation. For completeness, imposing $\Lambda_n = 0$ and $\tau_{n,Q}^{\mathrm{DG}} = \tau_{n,0}$ gives

$$M_{dc}^0(z) = \left[N_t \frac{8}{h}\left(\frac{\pi}{k_B T}\right)^{3/2} \tau_{n,0} \sum_{\substack{l\geq 0 \\ \varepsilon_d(z)+l\hbar\omega_0 \geq E_C(z)}} \frac{(S-l)^2}{S} L(l) \times \exp\left(\frac{E_C(z) - \varepsilon_d(z) - l\hbar\omega_0}{k_B T}\right) \sqrt{\varepsilon_d(z) + l\hbar\omega_0 - E_C(z)} \right]^{-1}. \tag{4.43}$$

The same derivation can be applied to holes. Introducing the density-gradient corrected valence-band edge

$$E_V^{\mathrm{DG}}(z) = E_V(z) + q\Lambda_p(z), \tag{4.44}$$

the corresponding effective hole lifetime is

$$\frac{1}{\tau_{p,Q}^{\mathrm{DG}}(z)} = N_t \frac{8}{h}\left(\frac{\pi}{k_B T}\right)^{3/2} M_{dv}^{0}(z) \times \sum_{l\geq 0} \frac{(S+l)^2}{S} L(l) \exp\left(\frac{\varepsilon_d(z) - l\hbar\omega_0 - E_V(z) - q\Lambda_p(z)}{k_B T}\right) \times \sqrt{E_V(z) + q\Lambda_p(z) - \varepsilon_d(z) + l\hbar\omega_0}. \tag{4.45}$$

The zero-field bulk calibration similarly gives

$$M_{dv}^{0}(z) = \left[N_t \frac{8}{h}\left(\frac{\pi}{k_B T}\right)^{3/2} \tau_{p,0} \sum_{\substack{l\geq 0 \\ \varepsilon_d(z) - l\hbar\omega_0 \leq E_V(z)}} \frac{(S+l)^2}{S} L(l) \times \exp\left(\frac{\varepsilon_d(z) - l\hbar\omega_0 - E_V(z)}{k_B T}\right) \sqrt{E_V(z) - \varepsilon_d(z) + l\hbar\omega_0} \right]^{-1}. \tag{4.46}$$

The density-gradient recombination rate is finally obtained by inserting the quantum-corrected carrier densities and the effective lifetimes of Eqs. (4.42) and (4.45) into the SRH-like expression derived in the NEGF section:

$$U_{\mathrm{DG}} = \frac{n_Q p_Q - \gamma_n n_1 \gamma_p p_1}{\tau_{p,Q}^{\mathrm{DG}}\left(n_Q + \gamma_n n_1\right) + \tau_{n,Q}^{\mathrm{DG}}\left(p_Q + \gamma_p p_1\right)}. \tag{4.47}$$

Trap-assisted tunneling is therefore included through two related effects: the modification of the carrier densities produced by the density-gradient correction and the corresponding variation of the multi-phonon effective capture lifetimes. In the limit $\Lambda_n, \Lambda_p \to 0$, the bulk semiclassical lifetimes and the conventional SRH recombination rate are recovered. Unlike the NEGF and Poisson-Schrödinger formulations, the density-gradient model does not provide an independently calculated spectral recombination rate. The auxiliary spectral distributions introduced above are used only to average the multi-phonon capture coefficients and determine the integrated effective lifetimes.

4.3.2 Numerical results

The density-gradient multi-phonon model is now applied to the same visible InGaN/GaN LED considered in the previous sections. Differently from NEGF and Poisson-Schrödinger drift-

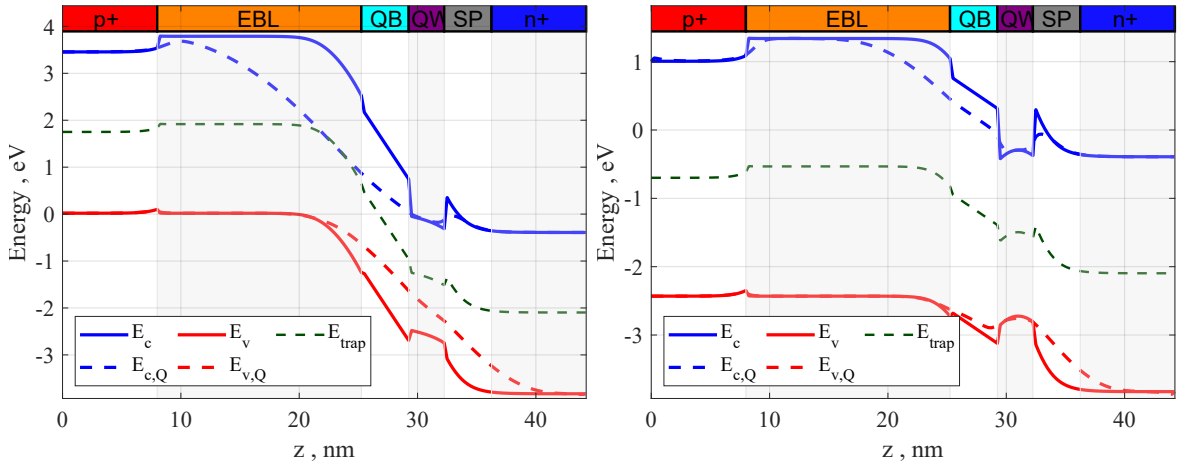


Fig. 4.11 Semiclassical and density-gradient quantum-corrected band diagrams of the visible InGaN/GaN LED at $V = 0.10$ V (left) and $V = 2.55$ V (right).

diffusion, the density-gradient method does not provide an explicit energy-resolved local density of states. Quantum effects are instead introduced through the density-gradient quantum potentials, which modify the effective conduction and valence band edges used in the drift-diffusion equations. The multi-phonon capture lifetimes derived in the previous subsection are therefore evaluated using these effective quantum band edges. Figure 4.11 shows the semiclassical and density-gradient quantum-corrected band diagrams at two representative voltages, $V = 0.10$ V and $V = 2.55$ V. These plots are useful because the effective quantum band edges indicate where the density-gradient correction introduces confinement and tunneling effects in the drift-diffusion description.

At low bias, the electric field across the junction is large and the density-gradient correction produces a visible extension of the effective band edges inside the classically forbidden region. This behavior represents the DGDD analogue of the tunneling correction discussed in the NEGF and Poisson-Schrödinger sections. The effective potential is particularly pronounced in the junction and active-region side of the device, where the strong electric field enhances the probability for carriers to access states inside the gap.

At $V = 2.55$ V, the forward bias reduces the electric field across the junction. As a consequence, tunneling inside the gap is reduced and the effective quantum potential becomes less pronounced. The correction remains visible mainly around the quantum barrier, quantum well, and spacer region, where confinement effects and local tunneling across the heterostructure barriers are still relevant.

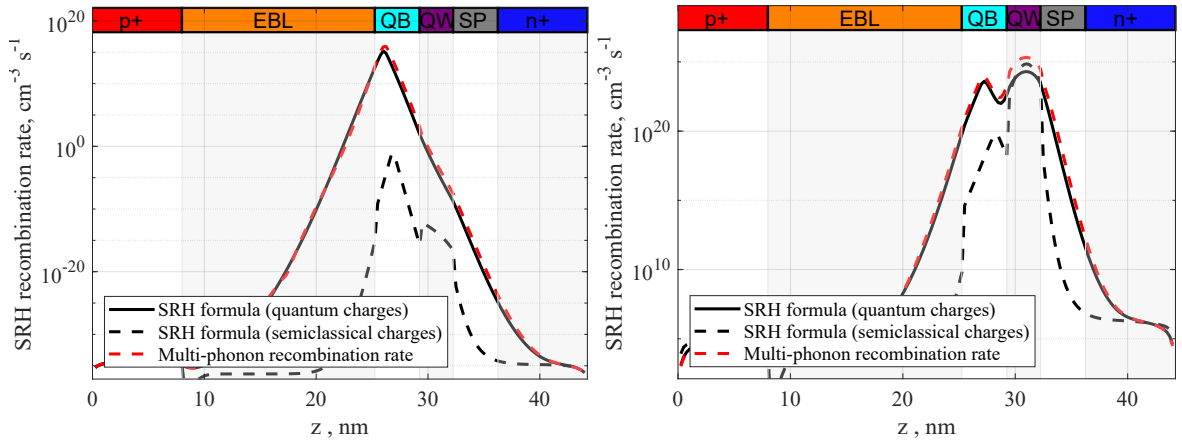


Fig. 4.12 Comparison between the SRH recombination rate evaluated with semiclassical carrier densities, the SRH recombination rate evaluated with density-gradient quantum-corrected carrier densities and constant bulk-like lifetimes, and the multi-phonon DGDD recombination rate at $V = 0.10$ V (left) and $V = 2.55$ V (right).

The impact of the quantum correction on the recombination rate is shown in Fig. 4.12. Three quantities are compared: the SRH recombination rate evaluated with semiclassical carrier densities, the SRH recombination rate evaluated with density-gradient quantum carrier densities and constant bulk-like lifetimes, and the multi-phonon recombination rate obtained from the effective density-gradient lifetimes.

At $V = 0.10$ V, the recombination is localized close to the junction side of the active region, where the strong electric field makes tunneling effects more relevant. However, differently from what was observed in the NEGF and Poisson-Schrödinger cases, the multi-phonon DGDD rate remains very close to the SRH rate evaluated with quantum-corrected quantities and bulk-like lifetimes. This behavior can be understood from the approximate quantum charge obtained from the density-gradient formulation. In this case, the additional tunneling-related charge introduced by the DG correction remains mostly compatible with the energy range involved in the trap-assisted capture process. Therefore, the multi-phonon formulation does not reveal a large amount of quantum charge excluded from SRH-like recombination, as occurred more clearly in the previous models.

At $V = 2.55$ V, the recombination peak moves toward the quantum barrier, quantum well, and spacer region, where the carrier densities become larger under forward bias. Also in this case, the multi-phonon DGDD rate follows closely the SRH rate evaluated with quantum-corrected carrier densities and constant bulk-like lifetimes. The main difference with respect to the semiclassical SRH rate is therefore mainly associated with the carrier-density correction introduced by DGDD, rather than with a strong modification of the capture process itself. The same behavior is reflected

in the integrated recombination current shown in Fig. 4.13. As in the previous sections, three calculations are compared: the semiclassical SRH rate obtained from a pure drift-diffusion simulation, the SRH rate evaluated with density-gradient quantum carrier densities and constant bulk-like lifetimes, and the multi-phonon DGDD recombination rate.

The current-voltage characteristics confirm the trend observed in the local recombination profiles. Both quantum-corrected calculations give a recombination current much larger than the semiclassical SRH result over most of the subthreshold and intermediate-bias range. This increase is mainly due to the density-gradient correction to the carrier densities, which introduces tunneling-related charge in regions that are classically forbidden in the pure drift-diffusion description. In contrast, the multi-phonon DGDD current remains very close to the SRH current evaluated with quantum charges and constant bulk-like lifetimes over most of the voltage range. This indicates that, in the density-gradient formulation, the multi-phonon lifetime correction does not introduce a strong additional modification of the integrated recombination current beyond the quantum-density correction itself.

The corresponding ideality factor shows the same qualitative transition already observed in the NEGF and Poisson-Schrödinger analyses. At low and intermediate bias, the ideality factor is much larger than two, consistently with a tunneling-assisted recombination regime localized near the EBL/QB side of the active region. At higher forward bias, the ideality factor decreases as carrier injection into the quantum well increases and recombination in the well becomes dominant. The multi-phonon DGDD curve exhibits a pronounced peak around the low-bias transition region. Since the ideality factor is obtained from the derivative of $\log J$, this peak is sensitive to small variations in the current and to the voltage discretization; therefore, its absolute value should not be overinterpreted. The robust information is the overall trend: DGDD reproduces a high-ideality TAT regime followed by a transition toward lower ideality when quantum-well recombination dominates. These results highlight the main limitation of the density-gradient approximation for the description of trap-assisted tunneling in this structure. The method is able to introduce quantum corrections within the drift-diffusion framework and to define effective multi-phonon lifetimes rooted in the same multi-phonon capture theory used in the previous sections. However, the quantum correction is obtained from an approximate density-gradient potential, rather than from the direct solution of a quantum transport or Schrödinger problem. As a consequence, the DGDD quantum charge can differ from that obtained with NEGF or Poisson-Schrödinger, both in magnitude and spatial distribution.

In the present case, the DGDD quantum charge remains mostly compatible with the trap-assisted recombination window. Therefore, the multi-phonon DGDD rate stays close to the SRH expression evaluated with quantum-corrected densities and constant bulk-like lifetimes. This

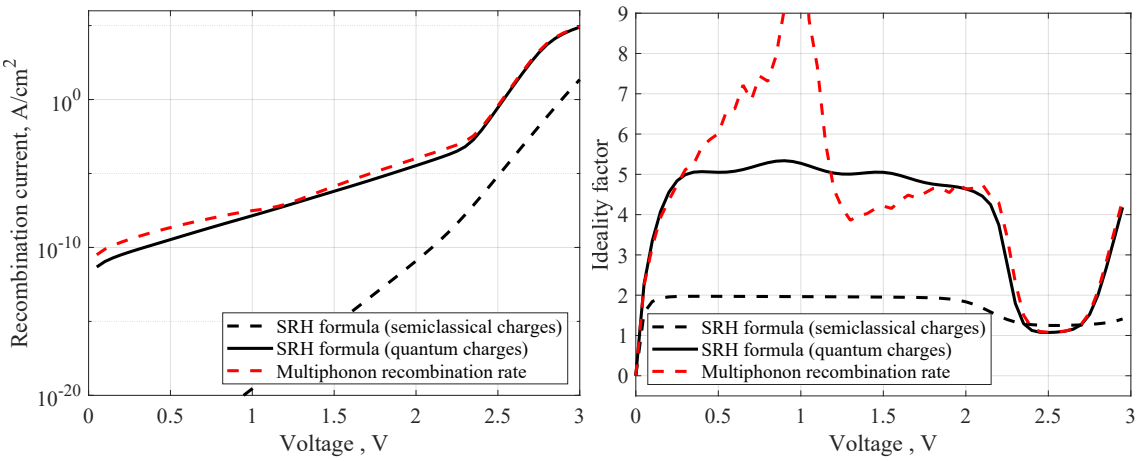


Fig. 4.13 Current-voltage characteristics (left) and ideality factor (right) obtained from the semiclassical SRH rate (black dashed lines), from the SRH formula evaluated with density-gradient quantum carrier densities and constant bulk-like lifetimes (black solid lines), and from the multi-phonon density-gradient recombination rate (red dashed lines).

should not be interpreted as a failure of the multi-phonon formulation. Rather, it indicates that the quantum charge obtained from the density-gradient approximation does not create as strong a separation between the charge made available by tunneling and the charge effectively participating in the trap-assisted multi-phonon capture process as in the more microscopic descriptions. In NEGF and Poisson-Schrödinger, the tunneling-related charge can extend more deeply into the gap, making the difference between the multi-phonon rate and the constant-lifetime SRH expression more visible.

Moreover, since DGDD does not provide an explicit energy-resolved density of states, this interpretation cannot be verified through spectral recombination maps as in the NEGF and Poisson-Schrödinger cases. It must instead be inferred from the effective band diagrams, spatial recombination profiles, and integrated electrical characteristics. This makes the method computationally attractive, but also more approximate when the energy distribution of the tunneling charge is important.

4.4 Comparison between the models

The previous sections analyzed the description of trap-assisted tunneling within three different frameworks: NEGF, Poisson-Schrödinger drift-diffusion (PSDD), and density-gradient drift-diffusion (DGDD). In this section, the results are compared directly in order to clarify which physical information is retained by the quantum-corrected drift-diffusion models and where

their approximations become relevant. NEGF is used as the most microscopic reference, since it describes the quantum density of states and the open-boundary transport problem explicitly, although at a much larger computational cost, which makes it unsuitable for large device simulation campaigns. The comparison is carried out at $V = 0.10$ V, representative of the low-bias TAT-dominated regime, and at $V = 2.55$ V, where injection into the active region becomes significant.

Figure 4.14 compares the semiclassical band diagram with the effective quantum band diagrams obtained from PSDD and DGDD. This comparison is useful because it directly shows how the two quantum-corrected drift-diffusion methods modify the band edges used in the transport calculation. At $V = 0.10$ V, both methods introduce an effective band-edge extension inside the gap, especially around the EBL/QB side of the active region, where tunneling-assisted capture is expected to be most relevant. However, the DGDD quantum correction is less pronounced in the EBL region than the PSDD one, indicating that DGDD predicts a weaker tunneling-related contribution in this region. In the quantum well, instead, the two methods give very similar effective quantum band edges, showing that confinement is described in a comparable way by PSDD and DGDD.

At $V = 2.55$ V, the electric field across the junction is reduced by the forward bias and the tunneling-related extension of the effective band edges becomes less pronounced. Nevertheless, the comparison between PSDD and DGDD remains consistent with the low-bias case: the two methods give similar confinement corrections in the quantum well, while their description of the tunneling correction near the EBL/QB region differs more significantly. This already suggests that the largest differences between PSDD and DGDD should appear in the tunneling-dominated part of the recombination process, rather than in the description of quantum confinement inside the well. The effect of these different quantum corrections on the carrier densities is shown in Fig. 4.15. At low bias, the most relevant differences appear around the EBL/QB side of the structure. In this region, DGDD gives a lower quantum charge than both PSDD and NEGF, consistently with the weaker quantum-potential correction observed in Fig. 4.14. PSDD, instead, follows the NEGF charge profile more closely in the tunneling region, indicating that the Poisson-Schrödinger correction provides a more accurate approximation of the quantum charge responsible for TAT.

At $V = 2.55$ V, the carrier densities show an additional feature. Both PSDD and DGDD predict a larger hole density inside the quantum well than NEGF. This difference is important because it directly affects the recombination current at higher forward bias. The discrepancy is mainly associated with the description of hole transport through the EBL/QB interface. In the quantum-corrected drift-diffusion calculations, after the quantum correction has been included

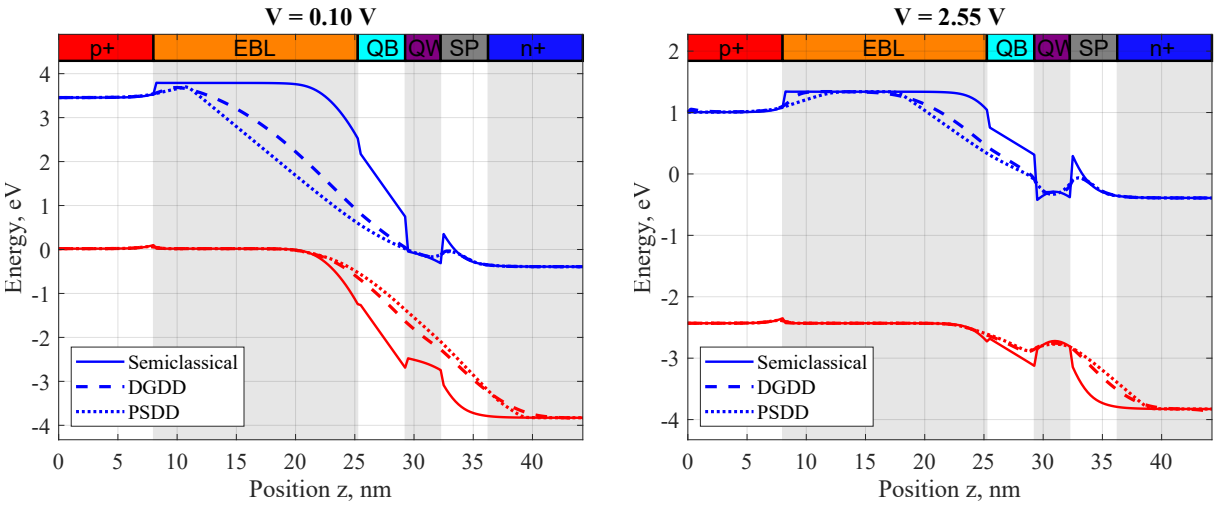


Fig. 4.14 Comparison between the semiclassical band diagram and the effective quantum band diagrams obtained from Poisson-Schrödinger drift-diffusion (dotted lines) and density-gradient drift-diffusion (dashed lines). Left: $V = 0.10$ V. Right: $V = 2.55$ V.

through the effective band edges, the transport is still governed by drift-diffusion equations and by the adopted mobility model. Since bulk mobilities are used here, the tunneling component of hole transport through the EBL/QB barrier is not described with the same field dependence as in the quantum transport calculation. This can lead to an overestimation of the hole injection into the quantum well in the QCDD approaches. The same trends are reflected in the product np, shown

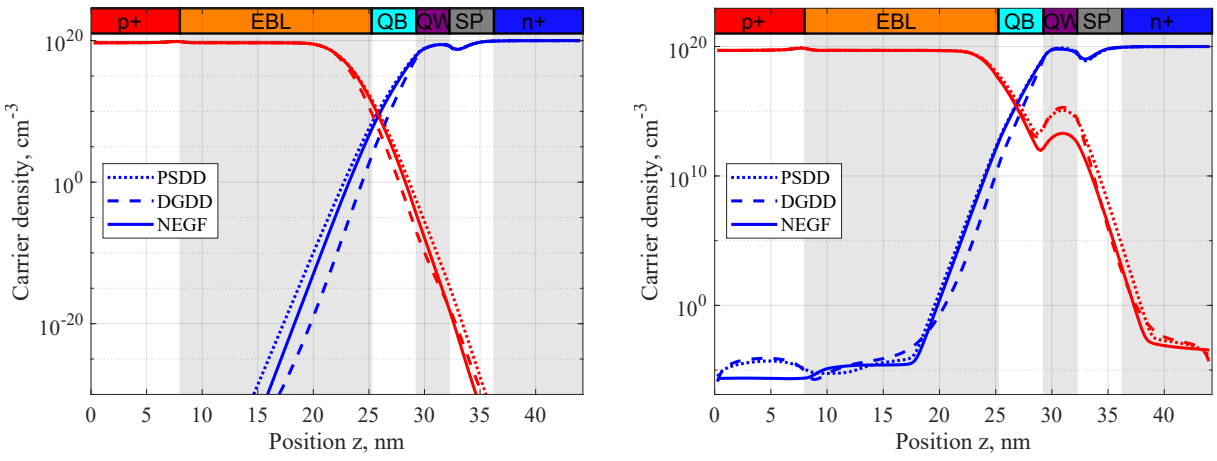


Fig. 4.15 Comparison of the electron density (blue lines) and hole density (red lines) obtained from NEGF (solid lines), Poisson-Schrödinger drift-diffusion (dotted lines), and density-gradient drift-diffusion (dashed lines). Left: $V = 0.10$ V. Right: $V = 2.55$ V.

in Fig. 4.16. This quantity is a useful indicator since recombinations depend strongly on it, so that we can observe the regions where SRH-like recombination can be enhanced. At $V = 0.10$ V, the np product is localized close to the EBL/QB side of the active region, consistently with a low-bias TAT regime. In this region, PSDD remains closer to NEGF, whereas DGDD gives a weaker carrier overlap. At $V = 2.55$ V, the np product increases strongly in the quantum well, where carrier injection becomes more efficient. The larger hole density predicted by the QCDD models therefore translates into a larger np product in the well region, which anticipates the differences observed in the recombination rates and in the high-bias current. The multi-phonon

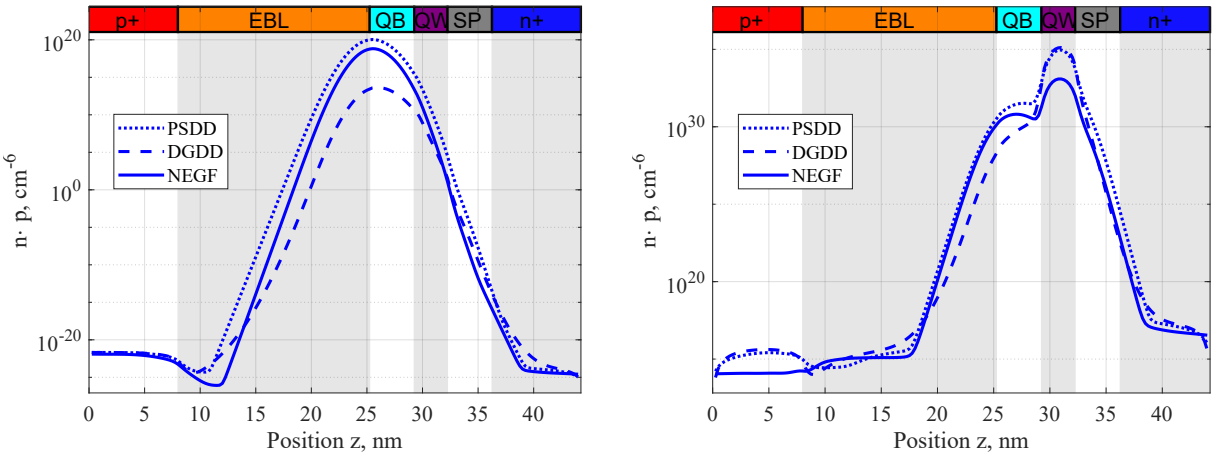


Fig. 4.16 Comparison of the product of electron and hole densities, np, obtained from NEGF (solid lines), Poisson-Schrödinger drift-diffusion (dotted lines), and density-gradient drift-diffusion (dashed lines). Left: $V = 0.10$ V. Right: $V = 2.55$ V.

recombination rates, described in the previous sections, are compared in Fig. 4.17. This figure is the most direct comparison of the TAT description provided by the three methods, because it includes not only the quantum carrier densities but also the multi-phonon capture probabilities. At $V = 0.10$ V, PSDD follows the NEGF recombination rate very well around the main TAT region. This confirms that the Poisson-Schrödinger spectral formulation is able to reproduce the main low-bias trap-assisted tunneling contribution obtained from NEGF. DGDD, instead, gives a different spatial profile and a slightly larger recombination rate in part of the junction region. This is consistent with the interpretation discussed in Sec. 4.3.2: in DGDD, the approximate quantum charge is less extended into the gap, so the multi-phonon capture process samples almost the same charge distribution that enters the SRH formula with constant bulk-like lifetimes. As a result, the multi-phonon DGDD rate remains close to the rate obtained with bulk-like lifetimes. In NEGF and PSDD, instead, part of the tunneling-related spectral weight can extend beyond the energy range effectively coupled to the selected trap, making the multi-phonon rate

smaller than what would be obtained from the total quantum charge alone. At $V = 2.55$ V, recombination in the quantum well becomes dominant. In this regime, both PSDD and DGDD overestimate the recombination rate inside the well with respect to NEGF. This follows directly from the carrier-density comparison in Fig. 4.15, where both QCDD methods predict a larger hole population in the quantum well. Therefore, the recombination-rate comparison connects the high-bias charge-density discrepancy to the electrical characteristics: the overestimated hole injection in the QCDD simulations leads to an overestimated recombination contribution around the LED working point. The final comparison is made on the current-voltage characteristics

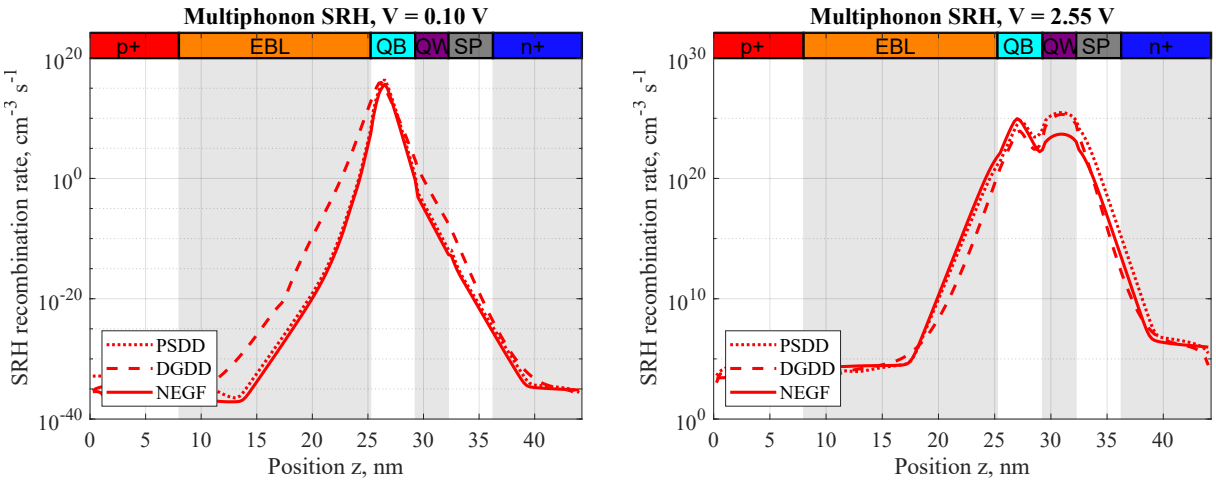


Fig. 4.17 Comparison of the multi-phonon recombination rates obtained with NEGF (solid lines), Poisson-Schrödinger drift-diffusion (dotted line), and density-gradient drift-diffusion (dashed lines). Left: $V = 0.10$ V. Right: $V = 2.55$ V.

and ideality factor, shown in Fig. 4.18. In this figure, all three models are evaluated using the multi-phonon recombination formulation. The experimental data from Ref. [45] are also shown for comparison. A series resistance of $2.5 \times 10^{-2}\ \Omega\,\mathrm{cm}^2$ is included in the simulated curves to account for resistive contributions not explicitly represented in the simulated device, including non-idealities in the contact regions and limitations associated with the one-dimensional simulation framework.

The NEGF calculation reproduces the experimental IV characteristic very closely over most of the investigated voltage range, supporting its use as the microscopic reference for the comparison. At very low bias, however, the simulated current remains lower than the experimental one. To assess whether this residual discrepancy could originate from the uncertainty in the parameters describing the defect, a sensitivity analysis was performed by varying the trap-energy position, the Huang–Rhys factor, and the Gaussian broadening of the trap-energy distribution around

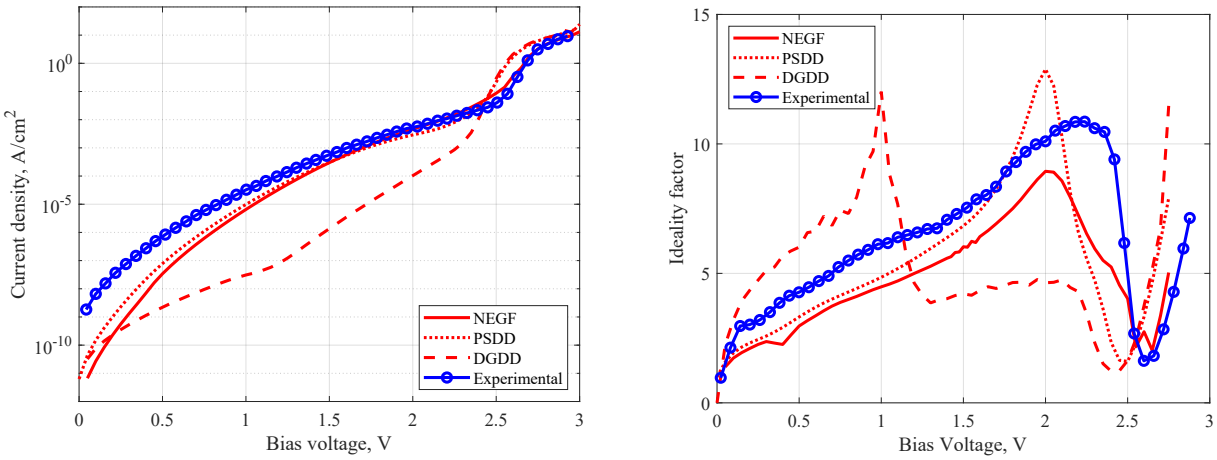


Fig. 4.18 Current-voltage characteristics (left) and ideality factor (right) obtained with NEGF (solid red line), Poisson-Schrödinger drift-diffusion (dotted red line), and density-gradient drift-diffusion (dashed red line) using the multi-phonon recombination formulation. The experimental data from Ref. [45] are also reported (blue line).

the reference values reported in Table 4.1. The resulting current-voltage characteristics and ideality factors are shown in Fig. 4.19. The trap-energy position has a significant influence on the subthreshold current. When the defect level is moved sufficiently far from the center of the bandgap, trap-assisted recombination is strongly suppressed, as carrier capture from one of the two bands becomes increasingly unfavorable. For trap energies closer to midgap, the dependence is less monotonic, since the resulting recombination rate depends on the spatial and energetic distributions of the carrier densities together with the corresponding effective capture lifetimes. Variations of the trap position consequently modify both the magnitude of the subthreshold current and the position of the maximum in the ideality factor.

The Huang–Rhys factor produces an even stronger variation of the trap-assisted contribution. Reducing S increases the low-bias current and can improve its agreement with experiment in this regime, but the same variation leads to an overestimation of the current at intermediate and higher forward bias and shifts the maximum of the ideality factor toward lower voltage. In contrast, variations of the trap-energy broadening produce only a comparatively small modification of the current-voltage characteristic over the investigated range, with a somewhat larger effect on the ideality factor. At sufficiently high forward bias, the curves obtained for the different parameter values tend to converge, consistently with the reduced importance of trap-assisted recombination once carrier injection becomes dominant. Overall, the considered parameter variations affect the quantitative agreement with experiment, but no individual variation provides a systematic improvement of both the current and ideality factor over the full voltage range. In particular, the

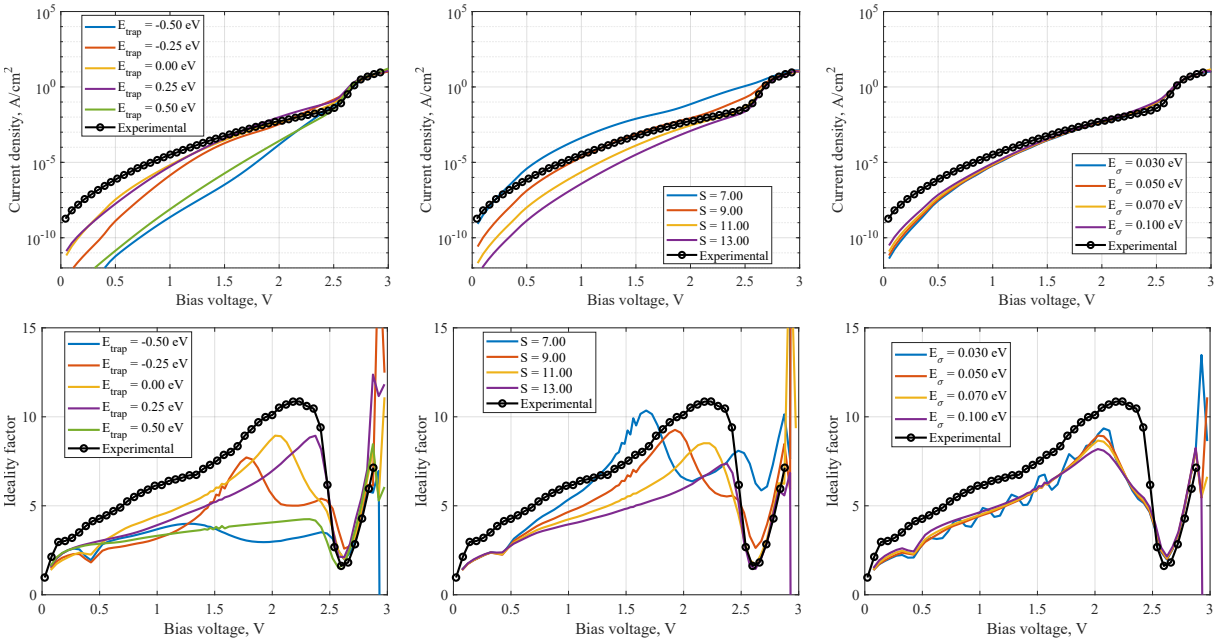


Fig. 4.19 Sensitivity of the calculated current-voltage characteristics (top row) and ideality factor (bottom row) to the parameters of the multi-phonon trap-assisted recombination model. From left to right, the trap-energy position E_{trap}, the Huang–Rhys factor S, and the Gaussian trap-energy broadening E_σ are varied. The trap-energy position is expressed relative to midgap, such that $E_{\text{trap}} = 0$ corresponds to a midgap defect. Experimental data are shown for comparison.

residual low-bias discrepancy cannot be attributed simply to the assumed midgap position of the trap level. Although some parameter variations can improve the agreement with the experimental current in specific voltage regions, the same variations do not provide a consistent improvement over the full voltage range. This suggests that the measured current may include additional low-bias tunneling-assisted channels not included in the present trap-assisted recombination model. One possible contribution is phonon-assisted tunneling recombination not mediated by the selected trap state. Such a process would be associated with different microscopic parameters and transition probabilities, and is therefore outside the scope of the present SRH-like multi-phonon TAT formulation. The ideality factor reflects the same behavior: the NEGF curve shows a strong increase in the subthreshold regime, similarly to the experimental trend, but its value remains slightly lower in the lowest-bias range where the current is underestimated.

The PSDD result follows NEGF closely in the subthreshold and intermediate-bias regimes. This agreement is consistent with the recombination-rate comparison of Fig. 4.17, where the PSDD multi-phonon rate reproduces the NEGF rate well in the TAT-dominated region. Therefore, the Poisson-Schrödinger approach retains the main quantum information needed to describe the low-bias TAT contribution, while remaining much less computationally demanding than

NEGF. At higher forward bias, however, PSDD tends to predict a larger current than NEGF and experiment. As shown in Fig. 4.15, this behavior is correlated with the larger hole density predicted inside the quantum well. This indicates that, once the quantum correction is mapped into the drift-diffusion framework, hole transport through the EBL/QB interface can become too efficient. Since the drift-diffusion part of the calculation uses bulk mobilities, the tunneling-assisted component of the transport through this barrier is not described with the same field dependence as in NEGF. A calibrated effective mobility, specifically accounting for tunneling through the EBL/QB barrier, could therefore improve the high-bias agreement without changing the underlying quantum-density correction.
The DGDD result captures the qualitative behavior of the IV and ideality factor, including the increase of the current in the subthreshold regime and the presence of an ideality factor much larger than two. However, it deviates more clearly from NEGF and PSDD. In the low-bias regime, the DGDD current is lower, consistently with the weaker quantum correction and the smaller charge overlap near the EBL/QB tunneling region discussed in Figs. 4.14-4.16. At higher voltage, DGDD is affected by the same drift-diffusion transport limitation discussed for PSDD, namely the overestimated hole population in the quantum well. The ideality factor curve should therefore be interpreted mainly in terms of its qualitative trend: DGDD captures the transition from a high-ideality TAT regime to a lower-ideality quantum-well recombination regime, but it is less accurate quantitatively than PSDD.

The comparison between the three methods is summarized in Table 4.2. NEGF provides the most complete description of the problem, because it treats tunneling, confinement, open boundaries, spectral densities, and scattering within the same quantum-transport framework. However, this comes at a very high computational cost, making full IV simulations difficult for routine device optimization. PSDD represents the best compromise among the quantum-corrected drift-diffusion methods considered here. It reproduces the NEGF carrier densities and multi-phonon recombination rates reasonably well in the TAT-dominated regime, while requiring only a fraction of the computational cost. Its main limitations are the phenomenological spectral broadening and the fact that transport remains semiclassical after the quantum correction is mapped into the drift-diffusion solver. DGDD is the cheapest approach and captures the qualitative effect of quantum corrections, including the emergence of high ideality factors associated with TAT. However, its approximated quantum charge differs more strongly from NEGF and PSDD in the EBL/QB tunneling region, which limits its quantitative accuracy for trap-assisted tunneling in this structure.

Table 4.2 Comparison of the transport models considered for quantum-transport effects, including representative wall-clock times per bias point.

Model	**Advantages**	**Limitations**	**Assessment**
NEGF	Full quantum-transport description; explicit LDOS, tunneling, open boundaries, and spectral quantities and transport.	Very high computational cost; impractical for large-scale or routine device simulations.	Most rigorous approach; best used as a reference for calibration and validation. Representative runtime: $\sim$ 12 h per bias point.
PSDD	Captures confinement and tunneling-related spectral charge; provides spectral recombination rates.	Requires phenomenological broadening; transport remains semiclassical after mapping to DD.	Best compromise between accuracy and computational cost for quantum-corrected device simulations. Representative runtime: $\sim$ 10 min per bias point.
DGDD	Low computational cost; directly compatible with DD; feasible for 2D/3D simulations.	Quantum charge comes from an approximate DG potential; less accurate in tunneling-dominated regions.	Fast qualitative approximation, useful for large simulations but less reliable quantitatively. Representative runtime: $\sim$ 1 min per bias point.

The runtimes reported in Table 4.2 are representative wall-clock times per bias point for the implementations used in this study and should not be interpreted as a controlled hardware benchmark. DGDD and PSDD were executed in MATLAB on a standard workstation, whereas the NEGF calculations used a parallelized Fortran implementation on a computational server. The comparison therefore reflects both the intrinsic computational requirements of the different formalisms and differences in implementation, hardware, and parallelization.

Overall, the comparison shows that trap-assisted tunneling in the visible InGaN/GaN LED cannot be described by a purely semiclassical drift-diffusion model with conventional SRH recombination. Quantum carrier densities are essential to reproduce the large subthreshold

current and the high ideality factors associated with tunneling-assisted recombination. Among the quantum-corrected drift-diffusion methods, PSDD provides the closest agreement with NEGF in the TAT-dominated regime, because it retains a spectral description of both the quantum charge and the multi-phonon recombination process, although transport remains semiclassical. DGDD captures the qualitative behavior at much lower computational cost, but its more approximate quantum correction leads to larger deviations in the tunneling region.

The results of this chapter therefore identify PSDD as the most suitable compromise between physical accuracy and computational feasibility among the reduced models considered here. The following chapter builds on this conclusion by moving from the microscopic analysis of TAT in a benchmark visible LED to the device-level simulation and optimization of AlGaN-based UV-C LEDs. In these structures, quantum effects are still central to the device operation. Carrier confinement, transport across heterostructure barriers, and trap-assisted tunneling strongly affect carrier injection and recombination, while tunnel junctions rely directly on interband tunneling. However, full NEGF simulations would be impractical for the larger structures and multiple design variations required in a device-optimization study. For this reason, the PSDD framework is used as a practical quantum-corrected tool to analyze and optimize the performance of UV-C LED structures.

Chapter 5

AlGaN UV-C LED optimization

Ultraviolet (UV) light sources gained renewed importance during the COVID-19 pandemic, when UV-C radiation was shown to effectively inactivate SARS-CoV-2 [47, 48]. More generally, UV-C sources are widely used for the sterilization of medical equipment and the disinfection of water supplies. The most established technology is the low-pressure mercury lamp [49, 50], which emits at 254 nm, a wavelength with strong germicidal effectiveness. Despite their widespread use, mercury lamps present several limitations, including long warm-up times, non-uniform radiation profiles, and safety concerns associated with mercury leakage caused by mechanical damage or device degradation.

These limitations have motivated the development of alternative UV sources. Among them, AlGaN-based light-emitting diodes (LEDs) represent a promising solid-state technology. Recent studies have shown that AlGaN UV-C LEDs can provide effective sterilization [51, 52], and commercial disinfection systems based on this technology have already been demonstrated. Depending on the target application, particular interest has been devoted to emission near 275 nm and near 230 nm. Radiation around 275 nm lies within an atmospheric transmission window, combines efficient propagation in air with strong germicidal activity, and is therefore attractive for disinfection. However, it is harmful to skin and eyes and requires appropriate shielding. In contrast, far-UV-C radiation around 230 nm is absorbed within the outer non-living layers of the skin, making it potentially safer for use in occupied spaces and enabling continuous disinfection. Its main drawback is the strong absorption by ozone and air, which limits the propagation distance.

Despite these advantages, the performance of AlGaN-based UV-C LEDs is still far below that of InGaN/GaN blue and green LEDs. While blue and green emitters can reach external quantum efficiencies close to 80% [53], AlGaN-based devices typically reach values around 10% for 275 nm LEDs at best, and below 1% for 230 nm LEDs [54, 55]. Several factors contribute to

this limitation. High dislocation densities in AlGaN enhance nonradiative recombination, the lack of ideal barrier materials increases leakage losses, and the large activation energy of p-type dopants strongly limits hole injection [56, 57].
Magnesium remains the most effective p-type dopant for AlGaN, but its high activation energy makes efficient p-type conductivity difficult to achieve at room temperature [58]. This problem becomes more severe as the aluminum content increases and complicates the formation of low-resistance ohmic contacts. For this reason, GaN is commonly used as a p-contact layer. However, GaN strongly absorbs UV-C radiation, reducing the light extraction efficiency (LEE). Moreover, GaN-based p-contacts typically rely on Ni/Au electrodes, whose reflectivity in the UV spectral range is poor, further limiting the extracted optical power.

In this chapter, the optimization of AlGaN UV-C LED structures is investigated by considering epitaxial designs aimed at improving both carrier injection and optical extraction. The goal is to increase the internal quantum efficiency (IQE) by improving carrier confinement and hole injection, and to increase the external quantum efficiency (EQE) and wall-plug efficiency (WPE) by reducing absorption losses in the p-side contact region.
The internal quantum efficiency is defined as the fraction of the injected current that contributes to radiative recombination in the active region. The external quantum efficiency further accounts for the fraction of generated photons that is extracted from the device, and can equivalently be expressed as the ratio between the emitted photon flux and the injected carrier flux. Finally, the wall-plug efficiency measures the conversion efficiency from electrical input power to emitted optical power. These quantities can be written as

$$\mathrm{IQE} = \frac{J_{\mathrm{rad}}}{J_{\mathrm{tot}}}, \qquad \mathrm{EQE} = \mathrm{IQE} \cdot \mathrm{LEE} = \frac{qP_{\mathrm{opt}}}{J_{\mathrm{tot}} h\nu}, \qquad \mathrm{WPE} = \frac{P_{\mathrm{opt}}}{J_{\mathrm{tot}} V} = \mathrm{EQE}\frac{h\nu}{qV}. \tag{5.1}$$

The reference structure considered in this chapter is a 275 nm AlGaN LED. It consists of an n^{++}-AlGaN contact layer with aluminum molar fraction $x = 0.6$, Si-doped at 10^{20} cm^{-3}, followed by the multi-quantum-well active region. The active region contains three 2 nm $Al_{0.4}GaN$ quantum wells embedded in 10 nm $Al_{0.6}GaN$ barriers, followed by an additional 10 nm barrier. A 30 nm $Al_{0.95}GaN$ electron blocking layer (EBL), p-doped at 5×10^{18} cm^{-3}, is then introduced to reduce electron leakage at high bias. Finally, the device is completed by a conventional p-GaN contact layer, doped at 10^{20} cm^{-3}, which improves hole injection and contact formation but reduces LEE because of its strong UV absorption. This structure is representative of designs reported in the literature [59, 60]. The corresponding band diagram is shown in Fig. 5.1, where the semiclassical band edges are compared with the quantum-corrected potentials describing confinement and intraband tunneling effects. The optimization is divided into two main parts. First, the electron

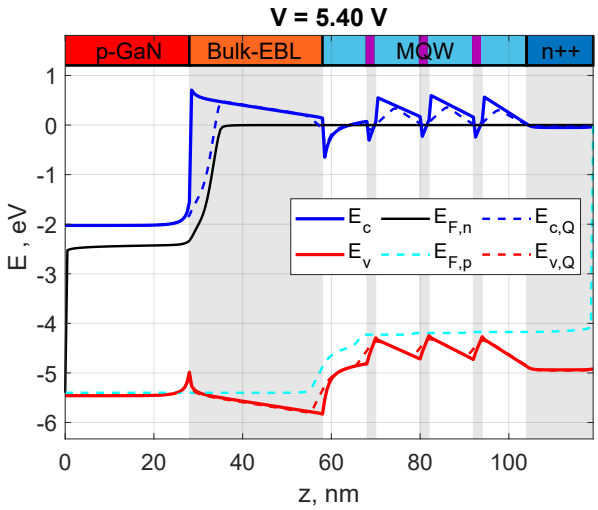


Fig. 5.1 Band diagram of the reference LED structure at 5.40 V, with semiclassical potentials (solid blue and red lines) and quantum potentials (dashed blue and red lines).

blocking region is modified in order to improve hole injection, carrier confinement, and radiative recombination. This is done by comparing a conventional bulk EBL with two alternative designs: a polarization-doped graded EBL and a superlattice EBL. Second, the p-side contact region is optimized to reduce UV absorption while preserving efficient hole injection. In this work, only intrinsic epitaxial and structural improvements are considered; packaging, external optics, and extraction techniques external to the epitaxial structure are left outside the scope of the analysis. The simulations are performed using the Poisson-Schrödinger drift-diffusion

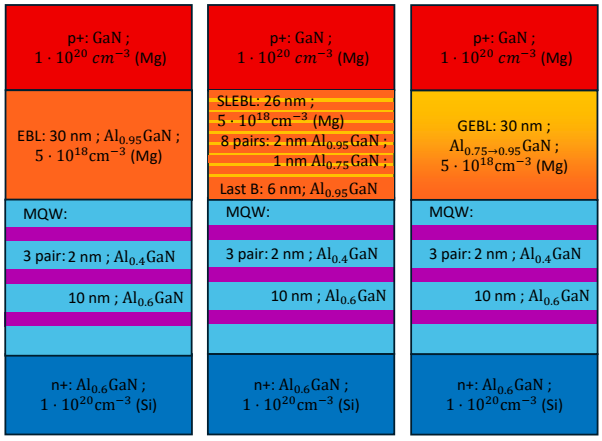


Fig. 5.2 Schematic layer structures of the reference LED with a bulk electron blocking layer (left), a superlattice EBL (center), and a graded EBL (right).

framework described in Sec. 3.2. This choice is motivated by the results of the previous chapters,

Table 5.1 Simulation parameters for the recombination terms, consistent with the literature [61, 59].

Name	Value	Unit
τ_{SRH}	5×10^{-9}	s
$B_{\mathrm{radiative}}$	2×10^{-11}	cm^3/s
C_{Auger}	2.7×10^{-31}	cm^6/s

where this approach was shown to provide a practical compromise between quantum-transport accuracy and computational efficiency. In the present context, it allows quantum confinement and tunneling effects to be included while keeping the computational cost compatible with device-level optimization of full UV-C LED structures. The recombination parameters are summarized in Table 5.1. The SRH lifetime is set to $\tau_{\mathrm{SRH}} = 5$ ns, assuming midgap traps. The radiative and Auger coefficients are set to $B = 2 \times 10^{-11}$ cm^3/s and $C = 2.7 \times 10^{-31}$ cm^6/s, respectively [61, 59].

5.1 Electron Blocking Layer (EBL) optimization

A major limitation of AlGaN UV-C LEDs is the poor hole injection caused by the high activation energy of Mg acceptors in Al-rich AlGaN. This limits the maximum internal quantum efficiency, because large forward biases are required to inject a sufficient hole density into the active region. At the same time, increasing the voltage enhances electron leakage over or through the EBL, which can become the dominant current-loss mechanism.

The role of the EBL is therefore twofold. It must provide a sufficiently large barrier for electrons, so that they remain confined in the quantum-well region, while introducing the smallest possible obstacle for holes injected from the p-side. Ideally, the EBL should be transparent for holes. In practice, however, the valence-band offset at the EBL/MQW interface is fixed by the material composition and cannot be tuned independently from the electron barrier. For this reason, the hole barrier must be reduced mainly through the electrostatic design of the layer. A high p-type charge density in the EBL is desirable because it bends the bands in a way that reduces the barrier experienced by holes. Since Mg ionization is inefficient in Al-rich AlGaN, alternative ways to increase the effective p-type charge in the EBL are required.

Several strategies have been proposed to mitigate this problem. In this section, two approaches based on EBL engineering are investigated: a polarization-doped graded EBL and a superlattice EBL. Both designs exploit polarization charges, which are characteristic of III-nitride heterostruc-

tures and arise at interfaces or in graded layers with different alloy compositions. The three device structures considered in this part of the analysis are shown schematically in Fig. 5.2.

5.1.1 Polarization-doped electron blocking layer (PD-EBL)

In III-nitride heterostructures, polarization charges arise from the discontinuity of the total polarization across interfaces. This total polarization includes both the spontaneous contribution, associated with the non-centrosymmetric wurtzite crystal structure, and the piezoelectric contribution, induced by strain due to lattice mismatch. The resulting bound sheet charges can strongly modify the local electrostatic potential and carrier distribution. In quantum wells, for example, polarization fields can separate electron and hole wavefunctions, reducing their overlap and lowering the radiative recombination efficiency.
At the same time, polarization charges can also be deliberately exploited to improve carrier injection. Simon et al. [62] showed that, by growing a compositionally graded III-nitride layer, the two-dimensional sheet charge normally localized at an abrupt heterointerface can be transformed into a three-dimensional distribution of polarization charge. This mechanism, commonly referred to as polarization doping or distributed polarization doping, provides a way to control carrier densities without relying only on substitutional dopants. In Ref. [63], for example, this concept was used to realize an $Al_{0.0\leftarrow 0.1}GaN/Al_{0.1\leftarrow 0.0}GaN$ junction without intentional Mg or Si doping, while still obtaining the electrostatic behavior of a conventional *p-n* junction. The corresponding band diagram and charge density are shown in Figs. 5.3 and 5.4, where the built-in potential typical of a *p-n* junction can be observed. To estimate the effective charge density generated by a compositionally graded AlGaN layer, a simple analytical model can be obtained by applying Vegard's law to the polarization vectors, following the Bernardini-Fiorentini model [46]. The spontaneous polarization P_{sp} and piezoelectric polarization P_{pz} can be written as functions of the aluminum molar fraction x_{mol}:

$$P_{\mathrm{sp}}(x_{\mathrm{mol}}) = (1 - x_{\mathrm{mol}})P_{\mathrm{sp,GaN}} + x_{\mathrm{mol}}P_{\mathrm{sp,AlN}} + b_{\mathrm{sp}}x_{\mathrm{mol}}(1 - x_{\mathrm{mol}}), \tag{5.2}$$

$$P_{\mathrm{pz}}(x_{\mathrm{mol}}) = P_{\mathrm{pz,AlN}}x_{\mathrm{mol}} + b_{\mathrm{pz}}x_{\mathrm{mol}}(1 - x_{\mathrm{mol}}). \tag{5.3}$$

The polarization parameters used in the model are reported in Table 5.2. By definition, the polarization charge density is obtained from the spatial derivative of the polarization vector. Assuming a linear variation of the molar fraction,

$$x_{\mathrm{mol}}(z) = \frac{x_{\mathrm{mol,end}} - x_{\mathrm{mol,ini}}}{L}(z - L) + x_{\mathrm{mol,end}},$$

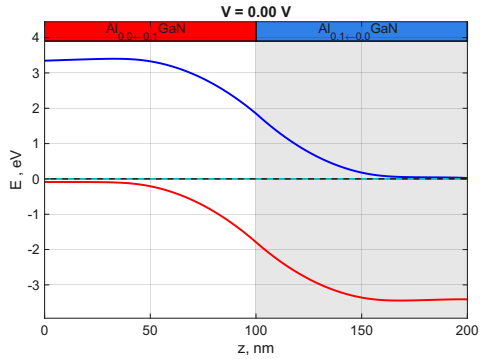


Fig. 5.3 Band diagram of the reference polarization-doped junction at equilibrium, with semiclassical potentials (solid lines) and quantum potentials (dashed lines).

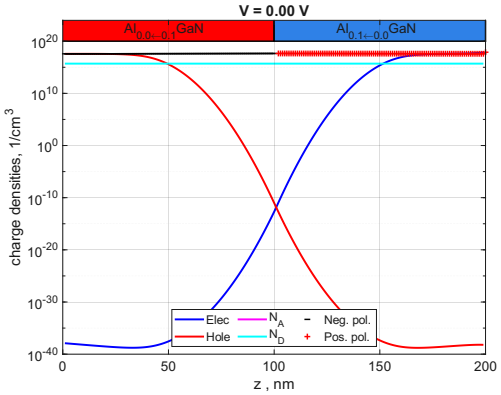


Fig. 5.4 Charge density of the reference polarization-doped junction at equilibrium.

Table 5.2 Polarization parameters for GaN and AlN.

Parameter	Value	Unit
$P_{\mathrm{sp,GaN}}$	-0.034	$\mathrm{C/m^2}$
$P_{\mathrm{sp,AlN}}$	-0.090	$\mathrm{C/m^2}$
b_{sp}	0.019	$\mathrm{C/m^2}$
$P_{\mathrm{pz,AlN}}$	-0.0525	$\mathrm{C/m^2}$
b_{pz}	0.0282	$\mathrm{C/m^2}$

the spontaneous (Q_{sp}) and piezoelectric polarization (Q_{pz}) charge densities can be written as

$$\begin{aligned} Q_{\mathrm{sp}}(z) &= \frac{\partial x_{\mathrm{mol}}(z)}{\partial z}\left[P_{\mathrm{sp,AlN}} - P_{\mathrm{sp,GaN}} + b_{\mathrm{sp}}\left(1 - 2x_{\mathrm{mol}}(z)\right)\right] \\ &= \frac{x_{\mathrm{mol,end}} - x_{\mathrm{mol,ini}}}{L}\left[P_{\mathrm{sp,AlN}} - P_{\mathrm{sp,GaN}} + b_{\mathrm{sp}}\left(1 - 2\left(\frac{x_{\mathrm{mol,end}} - x_{\mathrm{mol,ini}}}{L}(z - L) + x_{\mathrm{mol,end}}\right)\right)\right], \end{aligned} \tag{5.4}$$

$$\begin{aligned} Q_{\mathrm{pz}}(z) &= \frac{\partial x_{\mathrm{mol}}(z)}{\partial z}\left[P_{\mathrm{pz,AlN}} + b_{\mathrm{pz}}\left(1 - 2x_{\mathrm{mol}}(z)\right)\right] \\ &= \frac{x_{\mathrm{mol,end}} - x_{\mathrm{mol,ini}}}{L}\left[P_{\mathrm{pz,AlN}} + b_{\mathrm{pz}}\left(1 - 2\left(\frac{x_{\mathrm{mol,end}} - x_{\mathrm{mol,ini}}}{L}(z - L) + x_{\mathrm{mol,end}}\right)\right)\right]. \end{aligned} \tag{5.5}$$

These expressions allow the effective doping concentration generated by polarization charges in a graded AlGaN layer to be estimated directly from the initial and final molar fractions and from the grading length. The results are shown in Figs. 5.5 and 5.6. As expected, the effective polarization

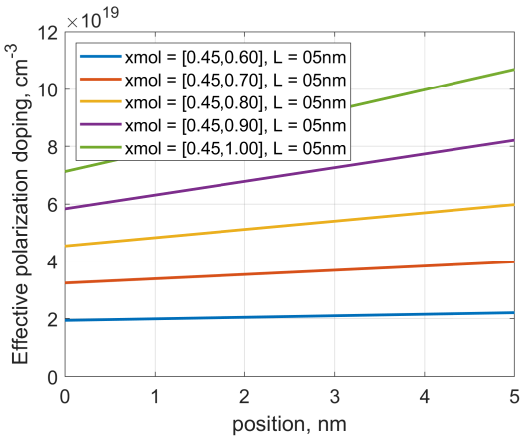


Fig. 5.5 Polarization doping in a graded layer as a function of the final molar fraction.

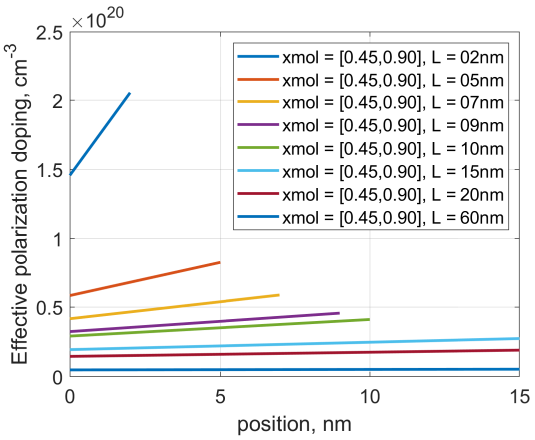


Fig. 5.6 Polarization doping in a graded layer as a function of the layer length.

doping increases when the grading length is reduced and when the difference between the initial and final molar fractions is increased. Achieving effective charge densities of the order of 10^{20} cm^{-3} would require very short and steep composition gradients, which are difficult to realize experimentally. Conversely, for applications where effective doping levels around 10^{18} cm^{-3} are sufficient, polarization doping represents a viable alternative to conventional Mg doping. The values obtained from this simple model are consistent with Hall-effect measurements reported in Ref. [60]. In that work, a 60 nm-thick graded layer with the Al molar fraction varying from 35% to 90% exhibited a hole concentration of approximately 6.9×10^{18} cm^{-3}, while the present model predicts an average value of 6.88×10^{18} cm^{-3}. Distributed polarization doping has recently been proposed for the EBL of AlGaN UV LEDs [60, 59]. In this configuration, the polarization charge increases the hole density in the EBL, mimicking the effect of a larger effective p-type doping even when the density of ionized Mg acceptors is very low or absent. This improves hole injection into the active region and can increase the radiative recombination rate and the internal quantum efficiency.

A graded EBL is therefore introduced by gradually varying the aluminum molar fraction from 75% to 100% along the EBL. The high final aluminum content is chosen to preserve UV-C transparency and electron confinement. The band diagram of the resulting structure is shown in Fig. 5.7. The graded barrier behaves similarly to the bulk barrier from the point of view of electron blocking, even when quantum corrections are included. The quantum band edges remain close to the semiclassical ones through most of the EBL, with tunneling contributions limited mainly to the upper part of the barrier. This indicates that electron leakage should not become a dominant issue. The same conclusion is supported by the charge-density profile in Fig. 5.8, where the minority electron density on the p-side remains comparable to that obtained with

the bulk EBL. The main advantage of the graded EBL is visible in the hole density. As shown

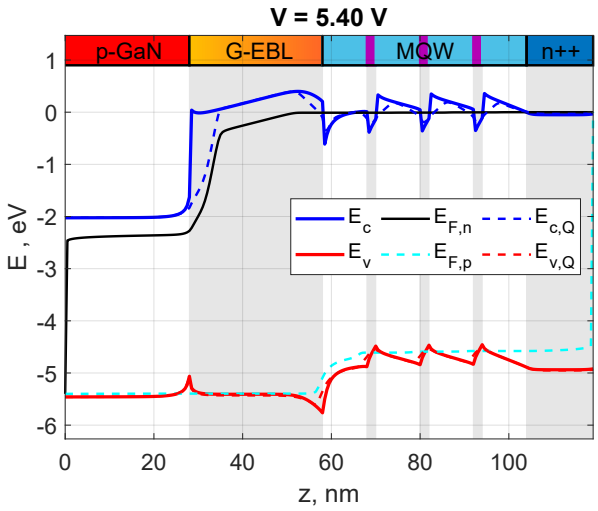


Fig. 5.7 Band diagram of the LED with graded EBL at 5.40 V, including the semiclassical and quantum-corrected band edges.

in Fig. 5.8, the hole concentration in the EBL reaches values close to 10^{19} cm^{-3}, exceeding the nominal Mg doping level of 5×10^{18} cm^{-3}. This occurs even though the concentration of ionized acceptors remains much lower, around 10^{10} cm^{-3}. Therefore, it would actually be possible to completely remove Mg doping in this layer since its charge density is much lower than the polarization one. The enhanced effective p-type charge in the EBL reduces the barrier that holes must overcome to reach the active region. The remaining barrier is mainly localized at the EBL/MQW interface, where positive polarization charge creates a local potential barrier. However, because the EBL is more effectively doped, this barrier becomes thinner and can be crossed more efficiently through tunneling. As a result, the hole concentration in the active region increases strongly, from approximately 10^{16} cm^{-3} in the reference structure to nearly 10^{20} cm^{-3} at the same applied voltage. This demonstrates that the graded EBL can substantially improve hole injection into the MQW region while preserving electron blocking capability.

5.1.2 Superlattice electron blocking layer (SL-EBL)

In recent years, several works have shown that replacing a bulk EBL with a superlattice (SL) structure can improve the efficiency of GaN-based blue LEDs [64–66]. Similar concepts have also been extended to deep-UV LEDs, with promising results [67]. In blue LEDs, the improved hole transport has often been associated with the enhancement of Mg ionization induced by the

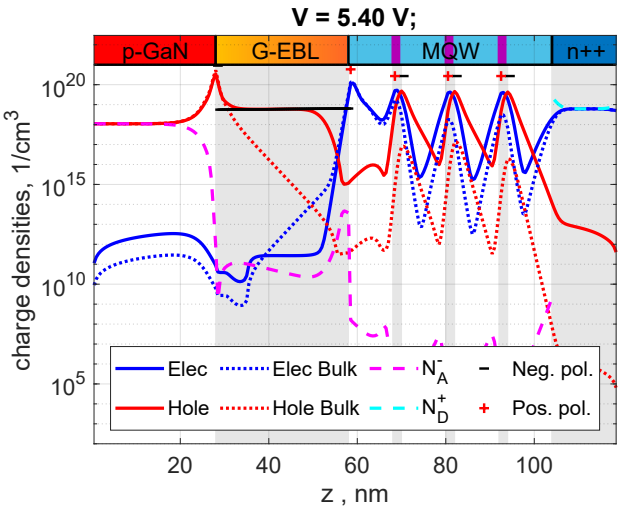


Fig. 5.8 Charge and ionized dopant densities of the LED with graded EBL at 5.40 V. The polarization charge density is also reported, showing its effect on the carrier distribution.

electrostatic effect of interface polarization charges. According to the incomplete-ionization model, the fraction of ionized acceptors is

$$\frac{N_A^-}{N_A} = \frac{1}{1 + g_A \exp\left(\frac{E_A - E_{F,p}}{k_B T}\right)}. \tag{5.6}$$

Therefore, acceptors become efficiently ionized when the acceptor level E_A lies below the hole quasi-Fermi level $E_{F,p}$, so that the acceptor states are occupied by electrons and become negatively charged. In bulk GaN, the Mg acceptor level lies about 200 meV above the valence-band edge. As a result, only a small fraction of the nominal Mg doping is ionized at room temperature. For example, a nominal Mg concentration of 10^{20} cm^{-3} can correspond to an ionized acceptor concentration of only about 10^{17} cm^{-3}.

In a superlattice, the polarization fields generated at the interfaces can locally shift the acceptor level below the Fermi level, increasing the ionized fraction of the dopants. This effect is shown in Fig. 5.9, which reports the band diagram of an 8 nm $Al_{0.20}Ga_{0.80}N$/8 nm GaN superlattice p-doped with Mg at a nominal concentration of 5×10^{19} cm^{-3}, similar to the structure analyzed in Ref. [64]. As shown in Fig. 5.10, the ionized acceptor density increases significantly at the GaN/$Al_{0.20}Ga_{0.80}N$ interfaces, reaching nearly full ionization. This is in contrast with the much lower ionized doping expected in the corresponding bulk material. At the opposite interfaces, namely $Al_{0.20}Ga_{0.80}N$/GaN, the acceptor level is pushed farther from the Fermi level by the negative polarization charge, producing a local reduction of the ionization efficiency.

Nevertheless, the negative interface charge itself contributes to the effective p-type charge, similarly to polarization doping, and can therefore increase the hole concentration and mobility, as observed experimentally in similar superlattice structures [64]. For UV-C LEDs, however,

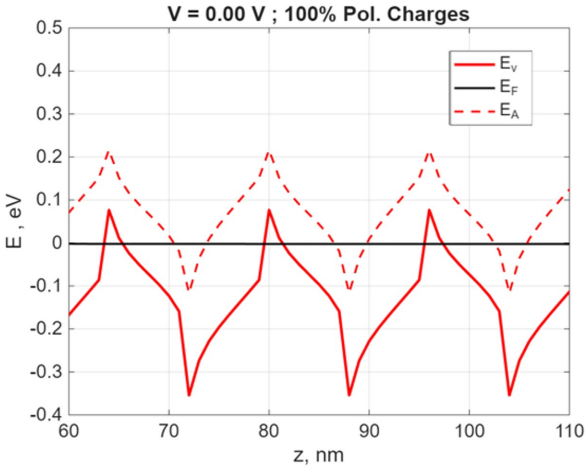


Fig. 5.9 Band diagram of a p-doped 8 nm $Al_{0.20}Ga_{0.80}N$/8 nm GaN superlattice at thermodynamic equilibrium, with Mg acceptor level E_A (dashed red line) at 200 meV above the valence-band edge.

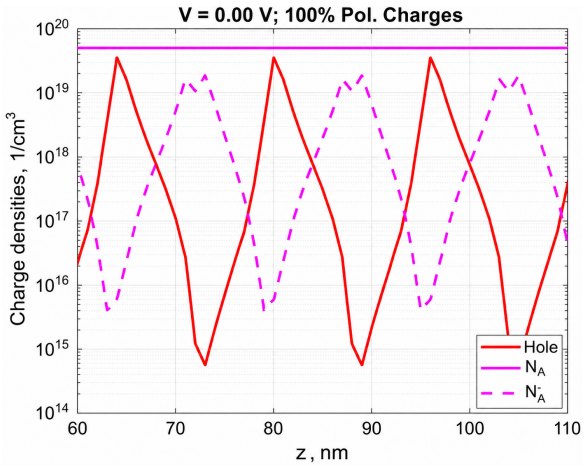


Fig. 5.10 Nominal doping (solid magenta line), ionized doping (dashed magenta line), and hole density (solid red line) of a p-doped 8 nm $Al_{0.20}Ga_{0.80}N$/8 nm GaN superlattice at thermodynamic equilibrium.

higher aluminum molar fractions are required in order to ensure transparency. In this case, the Mg acceptor level moves much farther from the valence band, increasing from about 200 meV to several hundred meV. As a consequence, even if the superlattice polarization fields improve the ionized acceptor density, the resulting concentration remains too low to provide efficient substitutional p-type doping. For example, for a nominal Mg concentration of 5×10^{19} cm^{-3}, the polarization fields can bring the Mg level closer to the Fermi level, but the acceptor level remains at least 150 meV above it. The maximum ionized acceptor density is therefore only about 10^{15} cm^{-3}, which is not sufficient for effective p-type doping.

Nevertheless, the hole concentration can still be strongly enhanced by the polarization charges at the superlattice interfaces, independently of the contribution from substitutional Mg ionization. Figs. 5.11 and 5.12 show a 2 nm $Al_{0.75}Ga_{0.25}N$/2 nm $Al_{0.95}Ga_{0.05}N$ superlattice with a negligible nominal Mg concentration of 10^{10} cm^{-3}. Despite the absence of effective substitutional p-type doping, a significant hole density is generated by the negative polarization charges at the superlattice interfaces. This mechanism will be referred to in the following as superlattice polarization doping. Since both positive and negative polarization charges are present in a superlattice, both electron and hole densities can in principle be enhanced. In practice, the electrostatics of the structure favors the enhancement of the majority carrier injected into the

superlattice. Therefore, to enhance hole transport, an efficient hole-injecting contact, such as a p-GaN layer or a tunnel junction, is still required. Superlattice structures therefore provide

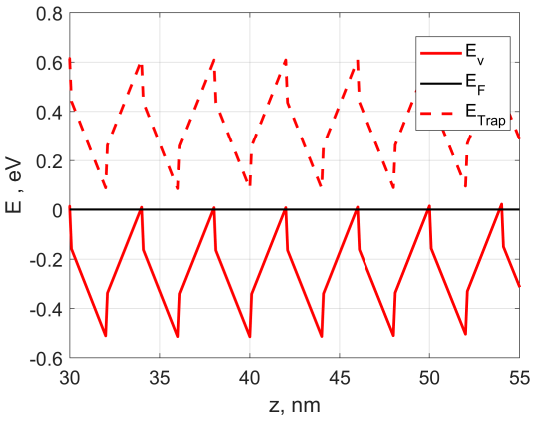


Fig. 5.11 Band diagram of a 2 nm $Al_{0.75}Ga_{0.25}N$/2 nm $Al_{0.95}Ga_{0.05}N$ superlattice at thermodynamic equilibrium with a negligible nominal Mg concentration of 10^{10} cm^{-3}, with Mg acceptor level E_{Trap} (dashed red line).

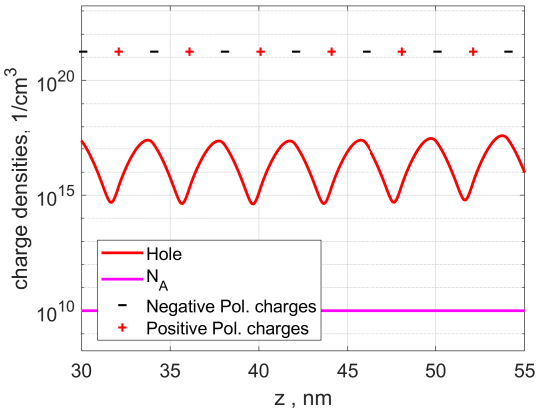


Fig. 5.12 Nominal doping (solid magenta line) and hole density (solid red line) with polarization charges (+,-) of a 2 nm $Al_{0.75}Ga_{0.25}N$/2 nm $Al_{0.95}Ga_{0.05}N$ superlattice at thermodynamic equilibrium.

two possible mechanisms to enhance the hole density in III-nitride devices. In low-Al-content materials, they can increase the Mg ionization efficiency through the interface electrostatics. In high-Al-content materials, where Mg ionization remains inefficient, they can still increase the hole density through superlattice polarization doping. This makes them attractive for UV-C LEDs, where high transparency requires Al-rich layers and conventional substitutional doping is strongly limited.

We now apply this concept to the EBL of the UV-C LED structure introduced above. Since a superlattice is intrinsically a quantum structure, a purely semiclassical drift-diffusion model cannot correctly describe transport through it. The Poisson-Schrödinger drift-diffusion model described in Sec. 3.2 is therefore used to include tunneling through the barriers, confinement in the wells, and miniband-assisted transport through the superlattice.

As a first example, a 2 nm $Al_{0.65}Ga_{0.35}N$/2 nm $Al_{0.85}Ga_{0.15}N$ superlattice is used as EBL, similarly to the design proposed in Ref. [65]. The corresponding band diagram and charge densities are shown in Figs. 5.13 and 5.14. The quantum-corrected band diagram provides direct information on transport through the superlattice. When the electron quantum potential lies below the semiclassical conduction-band edge, tunneling through the barriers becomes relevant. Conversely, when the quantum potential lies above the conduction-band edge, confinement raises the energy of the states inside the wells. In Fig. 5.13, electrons can tunnel through the thin

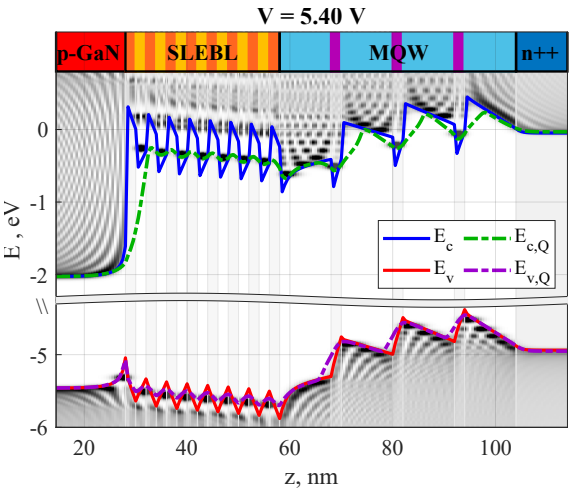


Fig. 5.13 Band diagram with PSDD LDOS at zero transverse wave vector and quantum potentials of a UV-C LED structure with a 2 nm $Al_{0.65}Ga_{0.35}N$/2 nm $Al_{0.85}Ga_{0.15}N$ superlattice at 5.40 V, with semiclassical band edges (solid lines) and quantum band edges (dashed lines).

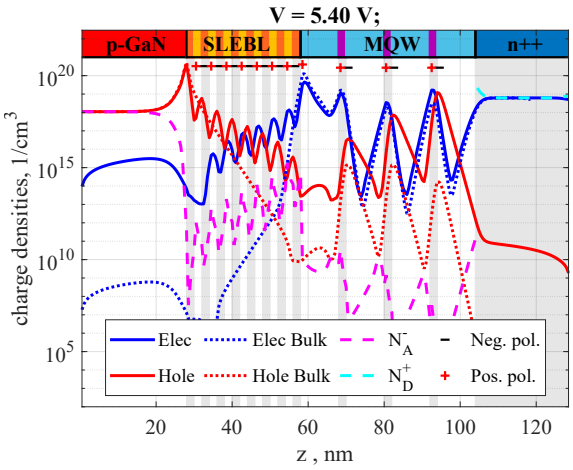


Fig. 5.14 Charge density of a UV-C LED structure with a 2 nm $Al_{0.65}Ga_{0.35}N$/2 nm $Al_{0.85}Ga_{0.15}N$ superlattice at 5.40 V. The ionized acceptor and donor densities are also shown. For comparison, the charge densities of the equivalent LED with bulk EBL are reported as dotted lines.

superlattice barriers and occupy the states formed inside the wells. The nearly flat quantum potential indicates the formation of a miniband, which facilitates electron transport through the EBL. Therefore, with respect to a semiclassical picture, quantum transport reduces the effectiveness of the superlattice as an electron blocking layer.

This behavior is confirmed by the charge densities in Fig. 5.14. The minority electron density on the p-side is significantly larger than in the bulk-EBL reference, indicating strong electron leakage through the superlattice. At the same time, the hole density in the quantum wells is substantially increased, showing that hole injection is improved. This improvement originates from the polarization charges at the superlattice interfaces, which enhance the hole density in the EBL even though the ionized acceptor concentration remains low, around 10^{12} cm^{-3}. As shown in Fig. 5.15, however, the resulting IQE remains only around 1%, because most of the injected current is lost through leakage rather than contributing to radiative recombination. This first structure shows that a superlattice EBL can improve hole injection, but also that an excessive leakage current can strongly reduce the IQE. A careful design of the superlattice is therefore required. The optimization starts by adding a thicker final barrier near the active region. This barrier must be thick enough to suppress electron leakage, while remaining thin enough to allow efficient hole injection into the MQW region.

The geometry of the superlattice wells is then varied to understand its impact on device

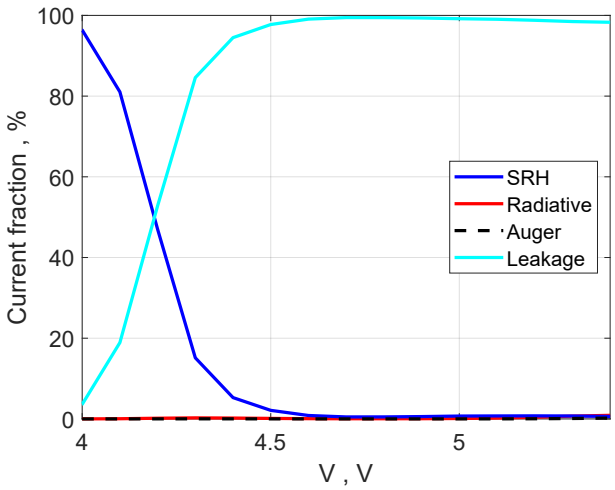


Fig. 5.15 Current fractions due to the different recombination mechanisms and leakage in the initial SL-EBL structure. The radiative fraction corresponds to the internal quantum efficiency.

performance. To isolate this effect, the superlattice period is kept fixed while the well and barrier thicknesses are changed. When the superlattice wells are made thicker, their quantum states become less confined and the corresponding miniband moves to lower energy. This facilitates electron leakage through the superlattice and reduces the efficiency. Conversely, when the barriers are too thick, hole injection is reduced. Because of their larger effective mass, holes tunnel less efficiently through thick barriers, and transport becomes more limited.

These effects are shown in Fig. 5.16, where two SL-EBL structures with the same 3 nm period are compared. The first structure uses 1 nm wells and 2 nm barriers, while the second uses 2 nm wells and 1 nm barriers. In the structure with thinner wells, the electron density on the p-side is lower, indicating reduced leakage. In the structure with thicker wells, instead, the thinner barriers allow more efficient hole injection and a larger hole density in the active region. The final device performance therefore results from the balance between electron leakage suppression and hole injection enhancement. A similar trade-off is obtained by varying the aluminum molar fraction in the superlattice wells. Increasing the Al content makes the wells shallower, raising the electron miniband energy and reducing electron tunneling through the EBL. This can improve electron confinement and reduce leakage. However, shallower wells also reduce the band discontinuity at the well/barrier interfaces, thereby decreasing the polarization charge and weakening the superlattice polarization-doping effect. As a result, hole injection into the active region becomes less effective.

This behavior is illustrated in Fig. 5.17, which compares two SL-EBL structures with different Al molar fractions in the wells. In the structure with shallower wells, corresponding to higher Al

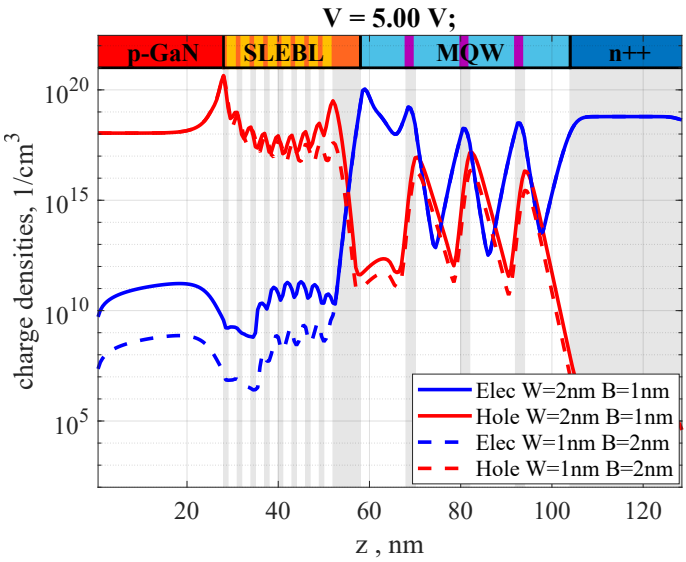


Fig. 5.16 Charge density profiles for two superlattice EBL geometries. Both structures have wells with an Al molar fraction of 75% and barriers with 95%. The first superlattice uses 1 nm wells and 2 nm barriers (dashed lines), while the second uses 2 nm wells and 1 nm barriers (solid lines).

content, electron leakage toward the p-side is reduced, but the hole density in the active region is also lower. Conversely, deeper wells enhance hole injection but lead to larger electron leakage. The design of the superlattice wells must therefore account for the interplay between polarization doping, hole injection, and electron confinement. To identify the optimal region, a simulation map is computed by varying the well Al molar fraction between 65% and 90% and the well thickness between 1 nm and 2.5 nm, while keeping the superlattice period fixed at 3 nm. The resulting peak IQE is shown in Fig. 5.18. As discussed above, wells that are too thin suppress leakage but reduce hole injection, whereas wells that are too thick enhance hole injection at the cost of increased electron leakage. Similarly, a well Al molar fraction that is too low produces excessive leakage, while an Al molar fraction that is too high reduces superlattice polarization doping and limits hole injection.

The voltage at which the IQE peak occurs provides additional information on the quality of hole injection. If the quantum wells are filled more easily, radiative recombination becomes dominant at lower bias and the efficiency peak is reached at a lower voltage. The peak voltages are shown in Fig. 5.19. A lower peak voltage is obtained for thinner barriers, because holes can more easily reach the active region. Similarly, reducing the Al molar fraction in the wells enhances the polarization-doping effect and improves hole injection. These results show that SL-EBLs offer a high degree of tunability, allowing the operating voltage and efficiency maximum to be adjusted through the superlattice geometry and composition. For the final comparison with the graded

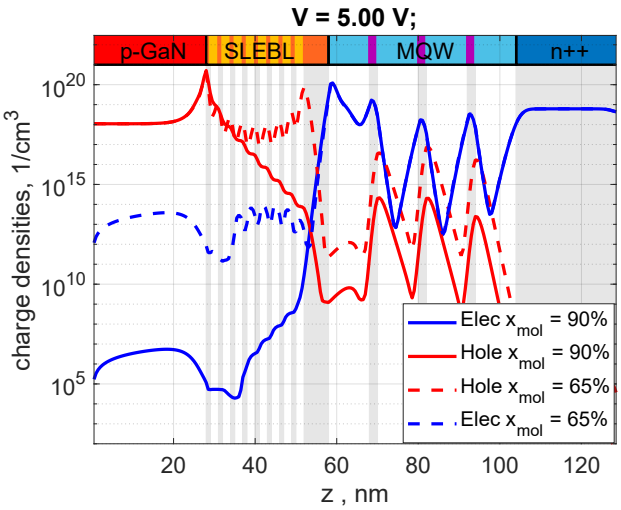


Fig. 5.17 Charge density profiles for two superlattice EBL geometries. Both structures use 1 nm wells and 2 nm barriers. One superlattice has an Al molar fraction of 65% in the wells (dashed lines), while the other has 90% (solid lines).

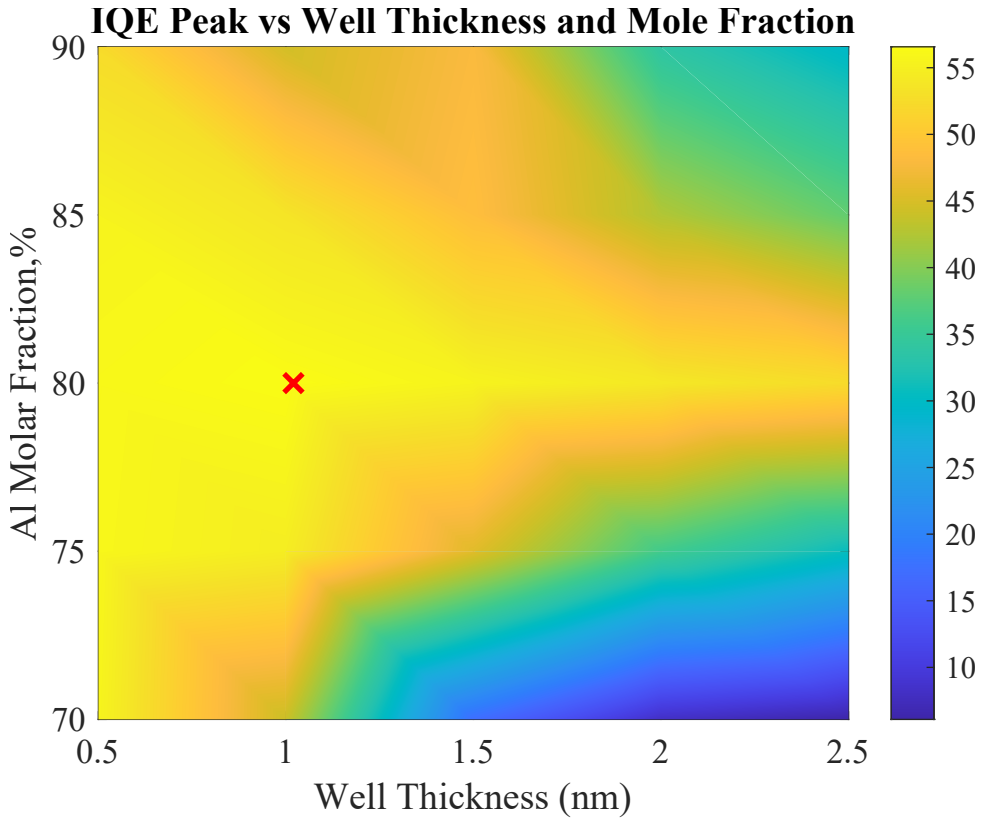


Fig. 5.18 IQE peak as a function of well thickness and Al molar fraction in the AlGaN superlattice EBL structure.

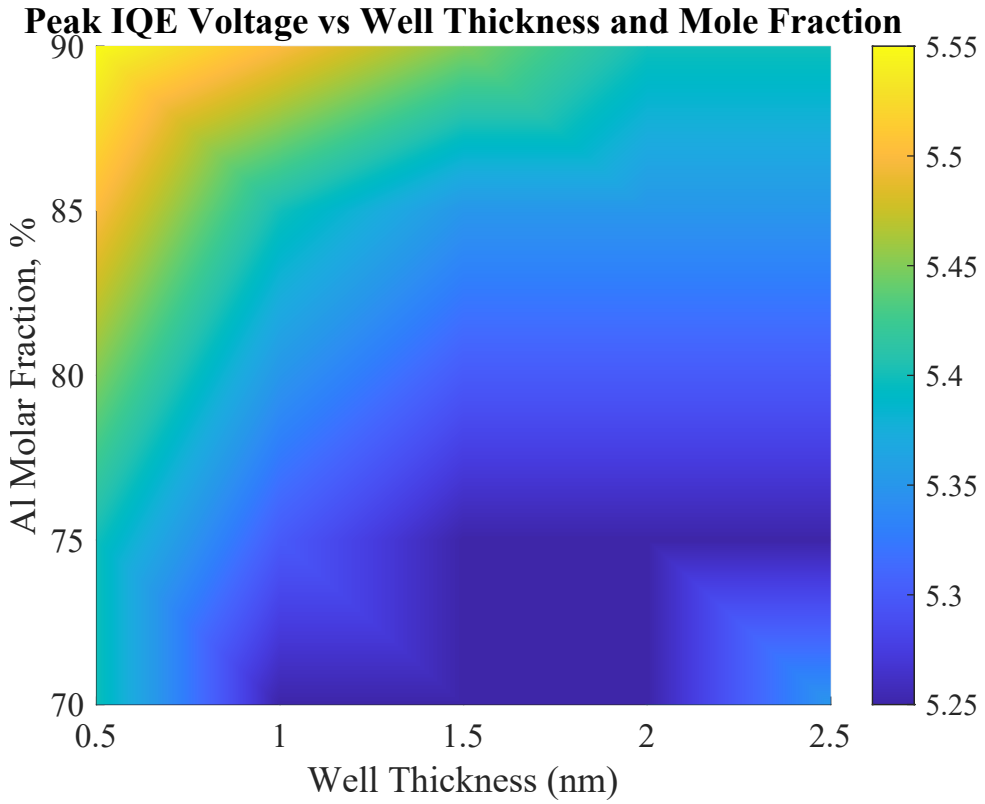


Fig. 5.19 IQE peak voltage as a function of well thickness and Al molar fraction in the AlGaN superlattice EBL structure.

EBL, the superlattice configuration giving the highest efficiency is selected. This optimized structure contains eight periods, with 1 nm wells and 2 nm barriers. The thin wells raise the miniband energy and reduce electron leakage, while the 2 nm barriers still allow a significant improvement of hole injection. A final 6 nm barrier is added near the active region to further suppress electron leakage. The band diagram and charge densities of this optimized SL-LED are shown in Fig. 5.20.

5.1.3 IQE comparison

The improvements in internal quantum efficiency obtained with the graded and superlattice EBLs are now compared with the conventional bulk EBL. All three devices show a turn-on voltage around 5 V, consistently with similar UV-C LED structures reported in the literature [60, 68, 59]. Figure 5.21 shows the fractions of the total current associated with radiative recombination, SRH recombination, Auger recombination, and leakage. As stated before, the internal quantum

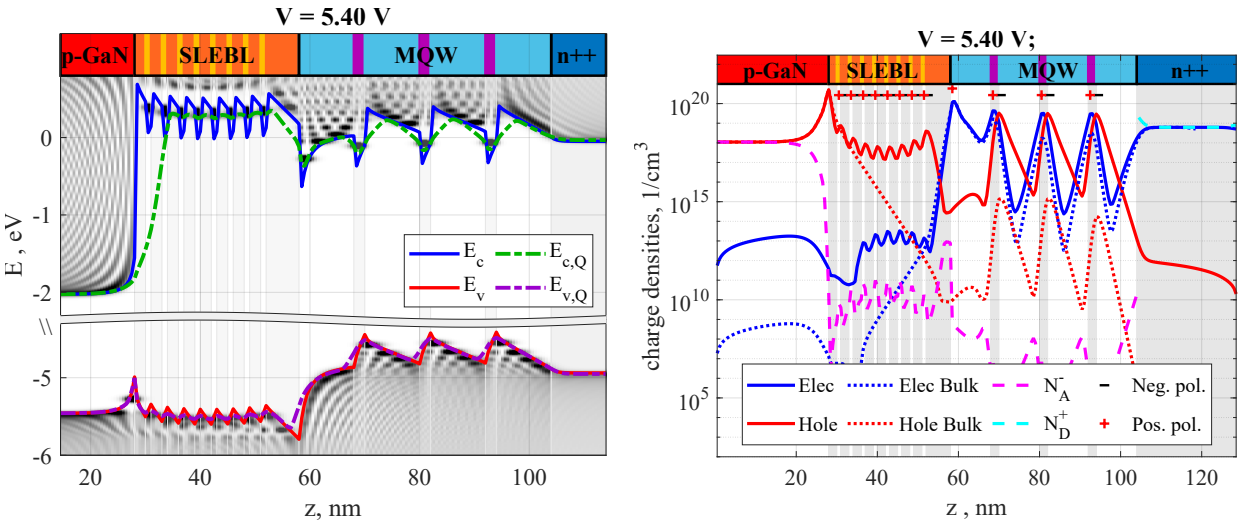


Fig. 5.20 (a) Band diagram with PSDD LDOS at zero transverse wavevector and quantum potentials and (b) charge densities of the optimized SL-LED structure at 5.40 V.

efficiency is defined as

$$IQE = \frac{J_{\text{rad}}}{J_{\text{TOT}}}. \tag{5.7}$$

Therefore, the radiative current fraction in Fig. 5.21 directly represents the IQE. At low injected current, the total current is mainly due to trap-assisted tunneling, a quantum process that enhances nonradiative recombination and is commonly associated with ideality factors much larger than two in III-nitride devices [69, 45]. In this regime, carriers tunnel into virtual states inside the gap and then recombine through trap states. As discussed in Chapter 4, this mechanism can be approximated within the SRH framework by using quantum carrier densities and effective multi-phonon lifetimes in the recombination rate. At higher current densities, the built-in potential is reduced and the quantum wells become sufficiently populated for radiative recombination to dominate. This corresponds to the IQE maximum. At still higher current densities, the radiative fraction decreases because Auger recombination becomes more important. Leakage current can also contribute to the efficiency reduction when carriers overcome or tunnel through the EBL instead of recombining in the active region.

With the bulk EBL, the maximum IQE is about 40%. This relatively low value is caused by poor hole injection: the quantum wells are efficiently filled only at higher voltages, where leakage becomes more significant. Both the graded and superlattice EBLs improve the IQE substantially, reaching peak values of 57% and 56%, respectively. The slightly lower efficiency of the superlattice EBL is attributed to the larger leakage current through the barrier. For the graded

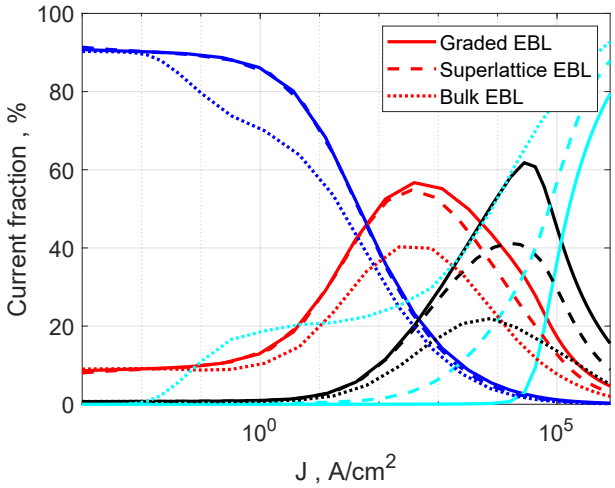


Fig. 5.21 Current fractions due to the different recombination mechanisms (SRH in blue, Radiative in red, Auger in black) and leakage (in cyan) with respect to the total current for LEDs with bulk EBL (dotted lines), graded EBL (solid lines), and superlattice EBL (dashed lines). The radiative fraction corresponds to the internal quantum efficiency.

EBL, the efficiency is mainly limited by SRH and Auger recombination, while leakage is strongly suppressed. This indicates that, once the EBL is properly engineered, further improvements in IQE would require better material quality, longer SRH lifetimes, or reduced Auger recombination, rather than additional modifications of the electron blocking layer.

5.2 p-contact layer optimization

Another major limitation of AlGaN UV-C LEDs is associated with the p-side contact region. Since AlGaN and GaN are wide-bandgap semiconductors, their p-type layers have a large work function. As a result, an extremely high metal work function would be required to form an ideal ohmic contact. For example, a work function close to 7 eV would be needed for an ideal p-GaN/metal contact. Since most metals have work functions between 4 and 6 eV, a significant Schottky barrier is formed at the metal/semiconductor interface. This problem becomes even more severe when the aluminum molar fraction is increased, because the band gap increases and the barrier for hole injection becomes larger.
A possible way to reduce the effect of this barrier is to make it thin enough for holes to tunnel through it. This requires a high hole density in the semiconductor layer adjacent to the metal contact. However, as discussed above, efficient p-type doping in III-nitride materials is difficult,

especially in Al-rich AlGaN. For this reason, conventional UV-C LEDs often use a p-GaN contact layer. In p-GaN, the Mg acceptor activation energy is lower than in Al-rich AlGaN, so a larger fraction of the nominal acceptor density is ionized and the contact resistance is reduced.
The use of p-GaN, however, introduces an important optical penalty. GaN strongly absorbs UV-C light, with an absorption coefficient of the order of 5×10^4 cm^{-1} at 275 nm, strongly reducing the light extraction efficiency. In addition, p-GaN contacts typically rely on high-work-function metals such as Ni and Au, whose reflectivity in the UV-C range is only about 20-30%. Since many UV-C LEDs are fabricated in flip-chip geometry, light is mainly extracted through the substrate. A highly reflective top metal could in principle redirect light emitted toward the p-side back into the device, increasing extraction. In practice, the combination of UV absorption in p-GaN and the low UV reflectivity of Ni/Au causes a significant fraction of the upward-emitted light to be lost.
In this section, two approaches are investigated to reduce p-side optical losses while preserving efficient hole injection. The starting point is the optimized LED with graded EBL described in Sec. 5.1.1, which still uses a conventional p-GaN contact layer. The first approach replaces p-GaN with a polarization-doped graded AlGaN contact, designed to remain transparent in the UV-C range while providing sufficient effective p-type charge. The second approach replaces the p-side contact with a tunnel junction, enabling the use of an n-type top contact. The three device stacks considered in this comparison are shown in Fig. 5.22.

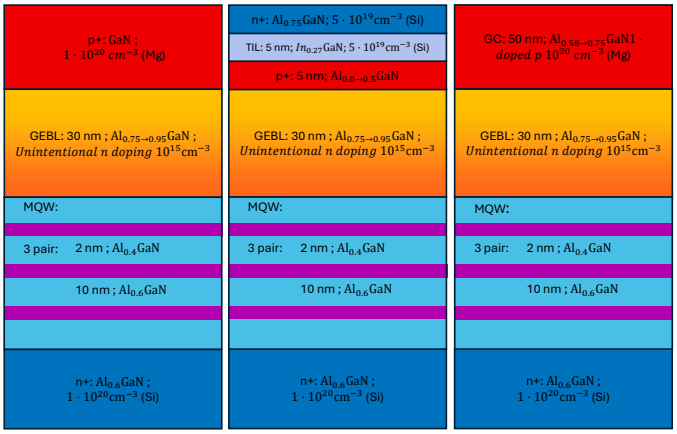


Fig. 5.22 Schematic structures of the simulated devices. Left: LED with conventional p-GaN contact. Center: LED with tunnel-junction contact on the p-side. Right: LED with polarization-doped graded contact.

5.2.1 Polarization-doped graded contact (GC)

The first proposed solution aims at replacing the absorbing p-GaN contact with a transparent AlGaN contact layer. A high aluminum molar fraction is needed to avoid UV-C absorption, but conventional Mg doping becomes inefficient in this composition range and would lead to large voltage losses. To overcome this limitation, polarization doping is used to generate an effective p-type charge density in the contact layer, following the mechanism described in Sec. 5.1.1.
To ensure transparency, the initial aluminum molar fraction of the graded contact is set to $x_{\mathrm{mol,ini}} = 58\%$. The final molar fraction is set to $x_{\mathrm{mol,end}} = 75\%$, in order to maintain compatibility with the graded EBL and avoid an abrupt discontinuity at the EBL/contact interface. The remaining design parameter is the length of the graded layer. In the reference device with p-GaN, the nominal Mg doping is $N_A = 10^{20}$ cm^{-3}, but the ionized acceptor density is only about 10^{18} cm^{-3}. Therefore, a comparable effective p-type charge density is required in the graded contact to preserve the electrical performance of the device.
Using the polarization-doping model introduced in Sec. 5.1.1, a grading from 58% to 75% Al provides an effective charge density of about 10^{18} cm^{-3} for layer lengths up to 145 nm. Shorter layers therefore provide larger effective doping and better hole injection. In this work, a 50 nm-long graded contact is selected. This length is short enough to provide a sufficient polarization-induced charge density, while remaining long enough to avoid an excessively abrupt composition change near the contact.
The resulting band diagram is shown in Fig. 5.23, while the corresponding charge density is reported in Fig. 5.24. The effective polarization doping in the graded contact reaches approximately 2×10^{18} cm^{-3}, even though the ionized Mg acceptor concentration remains very low, around 10^{13} cm^{-3}. The absence of an abrupt discontinuity between the contact layer and the EBL also improves the hole injection path toward the active region. In the reference device, a potential well forms at the p-GaN/EBL interface, as visible around 50 nm in Fig. 5.7. This well is removed in the graded-contact structure, and the higher effective p-type charge further facilitates hole transport. As a result, the efficiency peak is reached at 5.28 V, compared with 5.32 V for the device with p-GaN contact. The polarization-doped graded contact therefore provides a transparent p-side layer with sufficiently high effective doping to allow tunneling through the Schottky barrier and maintain adequate hole injection. Nevertheless, reducing the Schottky barrier still requires a high-work-function metal. Since such metals generally have low reflectivity in the UV-C range, the optical extraction remains partly limited by the metal contact.

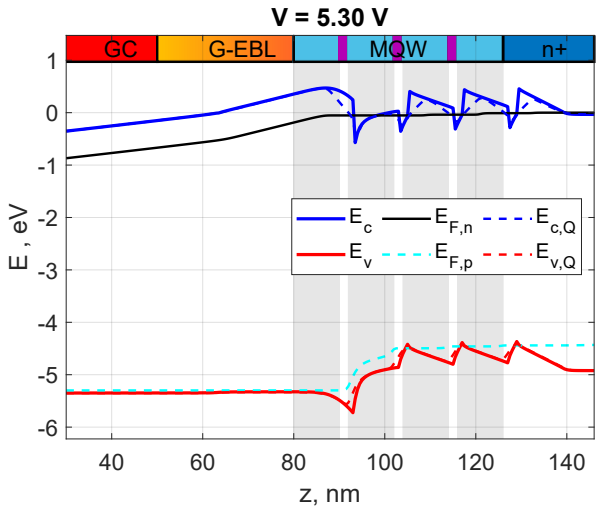


Fig. 5.23 Band diagram of the UV-C AlGaN LED at 5.3 V, where the p-GaN contact layer is replaced by a polarization-doped graded contact with Al molar fraction linearly increasing from 58% to 75%.

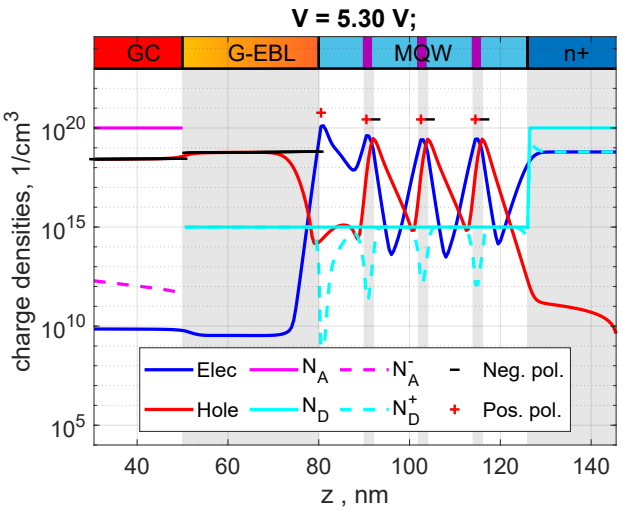


Fig. 5.24 Charge densities of the UV-C AlGaN LED at 5.3 V, where the p-GaN contact layer is replaced by a polarization-doped graded contact with Al molar fraction linearly increasing from 58% to 75%.

5.2.2 Tunnel-junction contact: optimization and application

The second proposed solution consists of replacing the p-side contact layer with a tunnel junction. This approach has already been demonstrated for visible and UV-B LEDs, where it can improve the external quantum efficiency by enabling the use of an n-type contact, which can be doped more efficiently and contacted with lower resistance [70–72]. However, an effective implementation in UV-C LEDs has not yet been reported, because the higher aluminum content required for UV-C

transparency makes conventional tunnel junctions intrinsically difficult to realize.
The main limitation of this approach is the trade-off between the optical benefit of removing the p-GaN contact and the electrical voltage drop introduced by the tunnel junction. If the tunnel junction requires a large additional voltage to provide the current density needed to drive the LED, the improvement in light extraction can be compensated, or even overcome, by a reduction in wall-plug efficiency. In the present device, it can be estimated that if more than about 3 V are required across the tunnel junction, the wall-plug efficiency of the TJ-LED becomes lower than that of the reference LED with p-GaN contact, despite the higher light extraction efficiency. Therefore, the tunnel junction must be optimized to deliver a large current density at the lowest possible voltage.
Tunnel junctions are intrinsically quantum devices, and cannot be accurately described by conventional drift-diffusion models. Moreover, because the relevant process is interband tunneling between valence-band and conduction-band states, a model in which the two bands are treated independently is not sufficient. For this reason, the tunnel junctions analyzed in this section are simulated using a 4×4 $\mathbf{k} \cdot \mathbf{p}$ NEGF model, as described in Chapter 2, which naturally includes interband coupling in the Hamiltonian and describes tunneling within a quantum-transport framework.
A conventional tunnel junction is first considered as a reference. It consists of a highly doped p-n homojunction. In wide-bandgap materials, however, the doping levels required to obtain an efficient tunnel junction can be extremely high. Figure 5.25 shows the IV characteristics of $Al_{0.58}Ga_{0.42}N$ p-n tunnel junctions with different nominal doping concentrations. The aluminum molar fraction is chosen to ensure transparency at 275 nm. The results show that a nominal doping of at least 10^{20} cm^{-3} is required to obtain even a small tunneling current. Taking 100 A/cm^2 as a reference current density for efficient LED operation, a tunnel junction doped at 10^{20} cm^{-3} would require almost 10 V, which is far too large for practical use. A doping concentration of 10^{21} cm^{-3} would reduce the required voltage to about 0.35 V, but such a high effective doping level is likely not technologically achievable.
The low current of these homojunctions has two main origins. The first and dominant one is the wide band gap of AlGaN, which increases the energy separation between conduction and valence bands and requires a large electric field to enable interband tunneling. The second is the low ionization efficiency of the dopants. In the depletion region under reverse bias, the strong band bending makes dopant ionization more effective. In the neutral regions, however, the ionized fraction can be several orders of magnitude lower than the nominal concentration. This reduces the effective carrier density and the built-in potential, so that a larger applied voltage is needed to align the valence and conduction bands for tunneling.

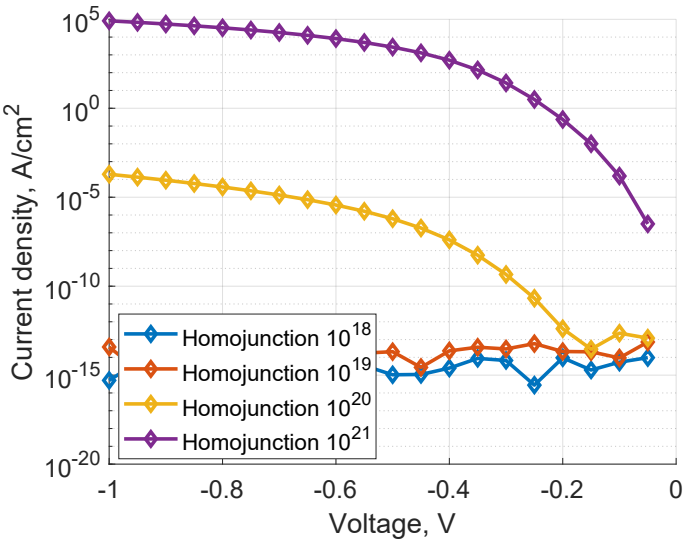


Fig. 5.25 IV characteristics of $Al_{0.58}Ga_{0.42}N$ p-n tunnel junctions with doping concentrations of 10^{18} cm^{-3} (blue), 10^{19} cm^{-3} (orange), 10^{20} cm^{-3} (yellow), and 10^{21} cm^{-3} (purple).

Since conventional UV-C-transparent AlGaN tunnel junctions do not provide sufficient current density, polarization doping is next considered as a possible way to improve their performance. At lower Al content, compositionally graded layers have been shown to enhance tunnel-junction characteristics [73]. Here, the same principle is applied to the higher Al molar fractions required for UV-C transparency. Since n-type doping is comparatively efficient, the modification is applied to the p-side of the junction. The first tested structure replaces the p-side with an undoped AlGaN layer graded from 58% to 78% Al. The corresponding IV characteristics are shown in Fig. 5.26. No significant current is obtained. Distributed polarization doping can reach effective charge densities of at most about 10^{19} cm^{-3}, which is not sufficient to strongly reduce the depletion width at the junction. As a result, the interband tunneling path remains too long, around 26 nm, to support efficient tunneling. This is confirmed by the band diagram and charge densities shown in Fig. 5.27. Polarization doping alone is therefore not sufficient to obtain an efficient tunnel junction. However, it can still help compensate the low ionization efficiency in the flat-band regions. This is shown in Fig. 5.28, where the p-side is replaced by an AlGaN layer graded from 58% to 78% Al and additionally Mg-doped at a nominal concentration of 10^{20} cm^{-3}. Compared with the conventional tunnel junction, polarization doping increases the current at low voltage. Near the junction, where the structure is under reverse bias, substitutional doping is effectively ionized and the local effective doping approaches 10^{20} cm^{-3}, which is sufficient to create a short tunneling path. The additional polarization doping mainly improves the flat-band regions, increasing the built-in potential and narrowing the interband tunneling barrier at low bias. In

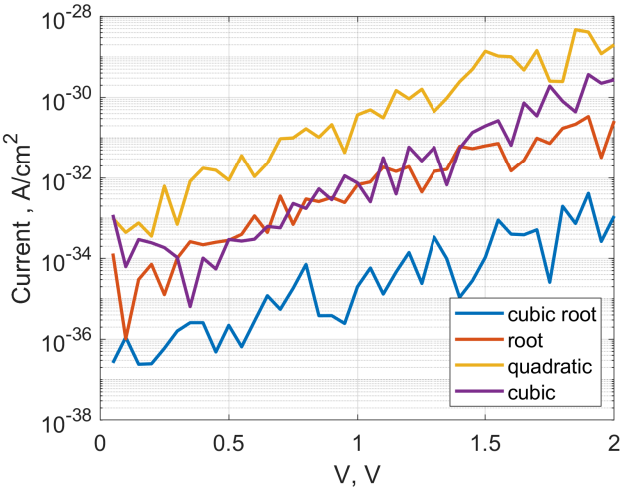


Fig. 5.26 IV characteristics of a graded tunnel junction where the p-side is replaced by an undoped AlGaN layer graded from 58% to 78% Al. Different grading profiles are explored.

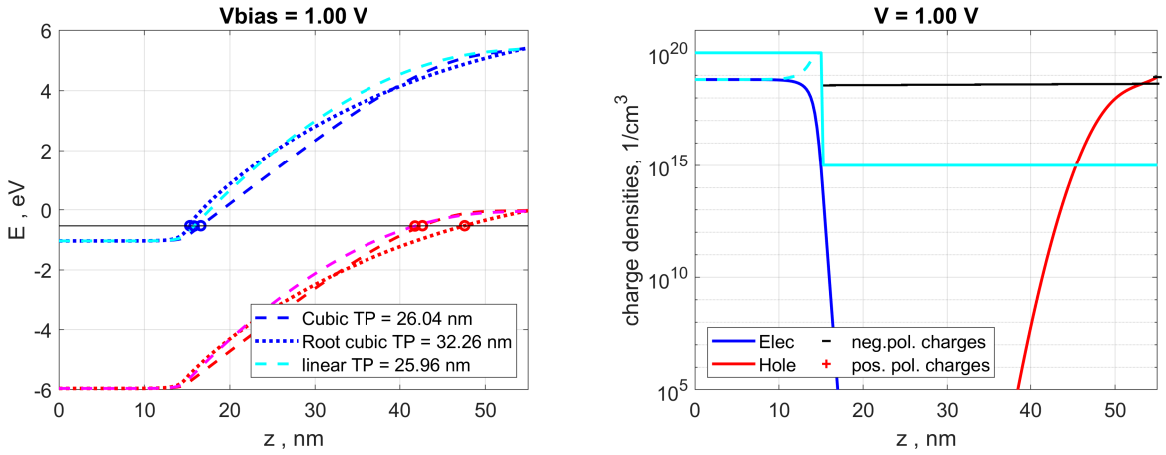


Fig. 5.27 (a) Band diagram and (b) charge densities of a graded tunnel junction where the p-side is replaced by an undoped AlGaN layer graded from 58% to 78% Al. Different grading profiles are explored.

this regime, a significant part of the tunneling current originates from energies close to the flat-band regions, so their electronic properties strongly affect the total current. At higher bias, the dominant tunneling energy shifts away from the flat-band edges and the junction electric field becomes the main factor controlling tunneling. Consequently, the benefit of polarization doping becomes less pronounced.

Although the current is improved, it remains too low for practical LED operation. Supplying the current density required by the LED would still require a voltage of the order of 10 V, which would eliminate the efficiency advantage expected from the improved LEE. The previous results

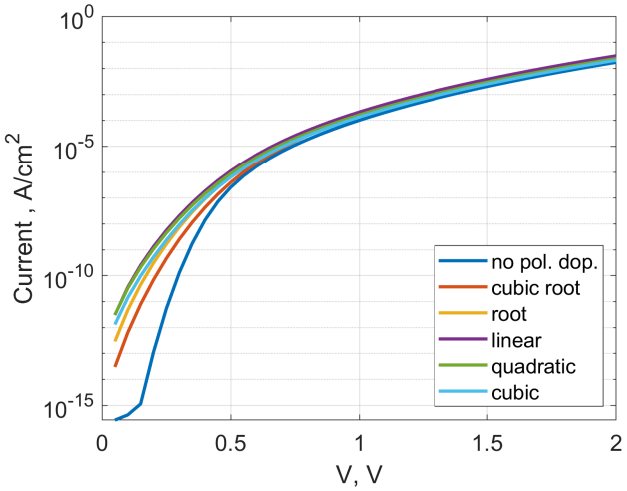


Fig. 5.28 IV characteristics of a graded tunnel junction where the p-side is replaced by a 10^{20} cm^{-3} Mg-doped AlGaN layer graded from 58% to 78% Al. Different grading profiles are explored.

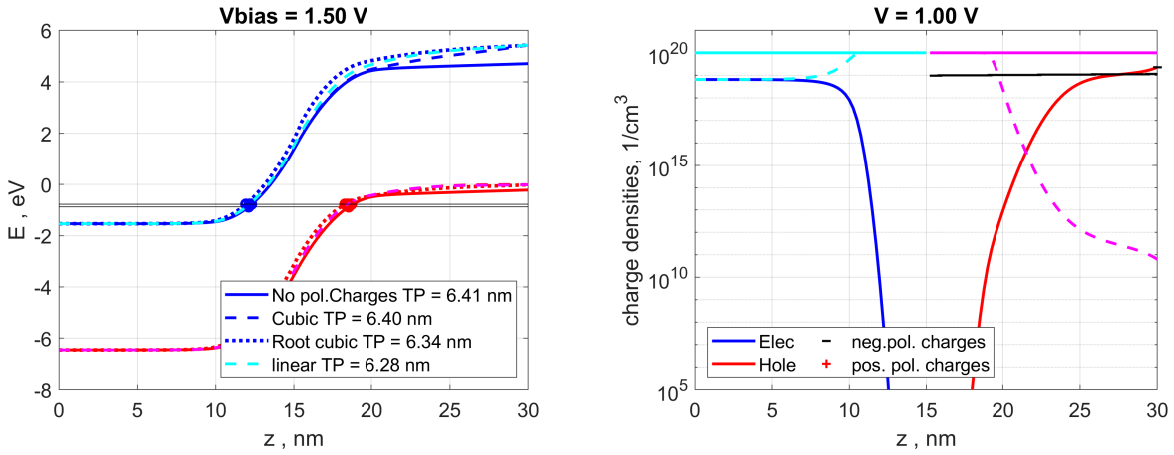


Fig. 5.29 (a) Band diagram and (b) charge densities of a graded tunnel junction where the p-side is replaced by a 10^{20} cm^{-3} Mg-doped AlGaN layer graded from 58% to 78% Al. Different grading profiles are explored.

show that an AlGaN homojunction tunnel junction would require extremely high doping, of the order of 10^{20}-10^{21} cm^{-3}, to provide sufficient current density. This limitation is caused mainly by the wide band gap of Al-rich AlGaN and by the limited ionization efficiency of p-type dopants. Polarization doping can improve the low-bias current, but it cannot generate sufficiently large charge densities unless very steep composition gradients are used, which may be difficult to fabricate.

A more effective strategy is to exploit polarization charges at abrupt heterointerfaces. At AlGaN/GaN or AlGaN/InGaN interfaces, the polarization discontinuity can generate very large

sheet charges, potentially corresponding to effective charge densities of the order of 10^{21} cm^{-3}. This can strongly enhance the electric field at the tunnel junction and increase the tunneling current at lower nominal doping concentrations. Heterostructure tunnel junctions have already been shown to improve tunneling performance [74, 75, 70]. In the following, the interlayer geometry is optimized in order to improve the current-voltage characteristics of a UV-C-transparent tunnel junction.

The starting point is a structure similar to that reported in Ref. [75]. The tunnel junction consists of an n-type $Al_{0.75}Ga_{0.25}N$ layer doped with Si at 5×10^{19} cm^{-3}, an 8 nm GaN interlayer n-doped at 10^{19} cm^{-3}, and a p-type $Al_{0.75}Ga_{0.25}N$ layer doped with Mg at 9×10^{19} cm^{-3}. The band diagram, spectral current density, and IV characteristic are shown in Fig. 5.30. Although the

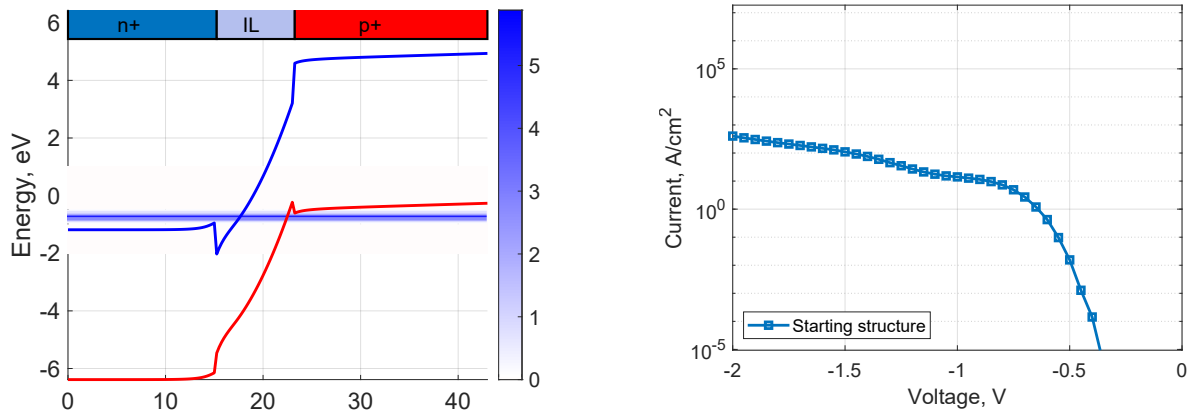


Fig. 5.30 (a) Band diagram and spectral current at −1.00 V of the reference tunnel junction with an 8 nm GaN interlayer and (b) corresponding IV characteristic.

heterostructure increases the current with respect to the AlGaN homojunction, the current density remains too low at low voltage. The current can be further improved by reducing the band gap of the interlayer and by increasing the polarization charge at the interfaces. Both effects can be obtained by replacing the GaN interlayer with an InGaN layer and varying the indium molar fraction. A parametric simulation campaign is therefore performed by varying both the interlayer thickness and the In molar fraction. The resulting tunneling current density at −1 V is shown in Fig. 5.31. The maximum current is obtained for an interlayer thickness of approximately 5 nm and an In molar fraction of about 27%. To understand the origin of this optimum, it is useful to analyze the tunneling path across the junction. In this structure, carrier transport can be interpreted as a sequence of three processes: first, intraband tunneling through the barrier formed at the n-layer/interlayer interface; second, interband tunneling inside the interlayer; and third, intraband tunneling through the barrier at the interlayer/p-layer interface.

The effect of the interlayer thickness is shown in Fig. 5.32. Increasing the interlayer thickness

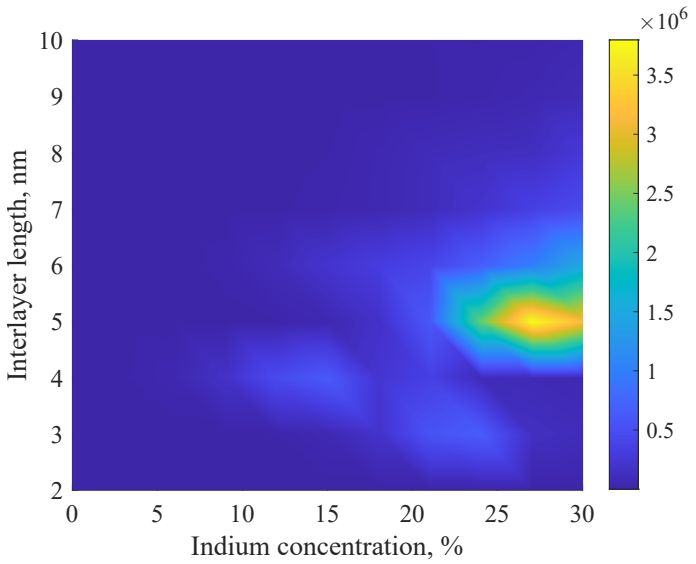


Fig. 5.31 Map of tunneling current density at −1 V as a function of interlayer thickness and indium molar fraction in the tunnel junction.

reduces the intraband barrier heights between the n-layer and the interlayer (E_1) and the barrier that forms between the interlayer and the p-layer E_2, because the electric field across the interlayer becomes smaller. This facilitates the intraband tunneling steps at the two interfaces. At the same time, the interband tunneling path increases almost linearly with the interlayer thickness, because the conduction and valence bands remain separated over a longer distance. The optimum thickness therefore results from a trade-off: a thicker interlayer reduces the intraband barriers, but increases the interband tunneling distance. In the present structure, this balance gives an optimum interlayer thickness close to 5 nm. A similar optimum is obtained when the In molar fraction is varied. Increasing the In content reduces the interlayer band gap, thereby shortening the interband tunneling path. However, the behavior of the intraband barriers is more complex because it results from two competing effects. The first is the band-offset difference between the interlayer and the AlGaN layers, which tends to increase the intraband barriers. The second is the polarization charge at the heterointerfaces, which modifies the local electric field and can reduce these barriers.

For the n-layer/interlayer barrier E_1, the band-offset contribution dominates, producing an overall increase of the barrier height as the In content increases. For the interlayer/p-layer barrier E_2, instead, the valence-band discontinuity is smaller and the polarization charge has a stronger influence. A change in the behavior of both barriers appears around an In molar fraction of 25%, where the polarization charge reaches its maximum according to Ref. [76]. The optimum In composition therefore results again from the balance between a shorter interband tunneling

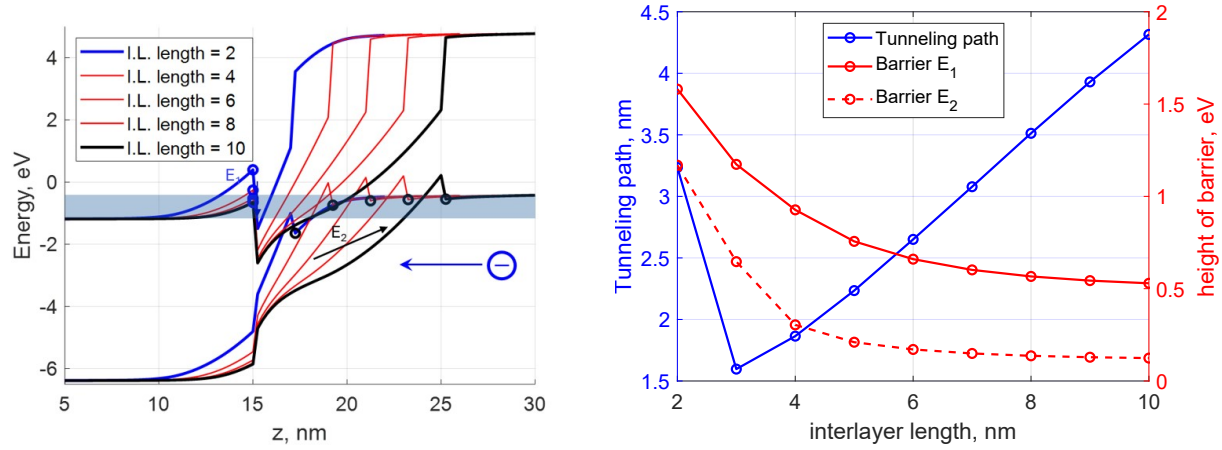


Fig. 5.32 (a) Band diagrams for different interlayer thicknesses. (b) Variation of the intraband barrier heights E_1, at the n-layer/interlayer interface, and E_2, at the interlayer/p-layer interface, together with the interband tunneling path.

path and the evolution of the intraband barriers. This behavior is shown in Fig. 5.33. The

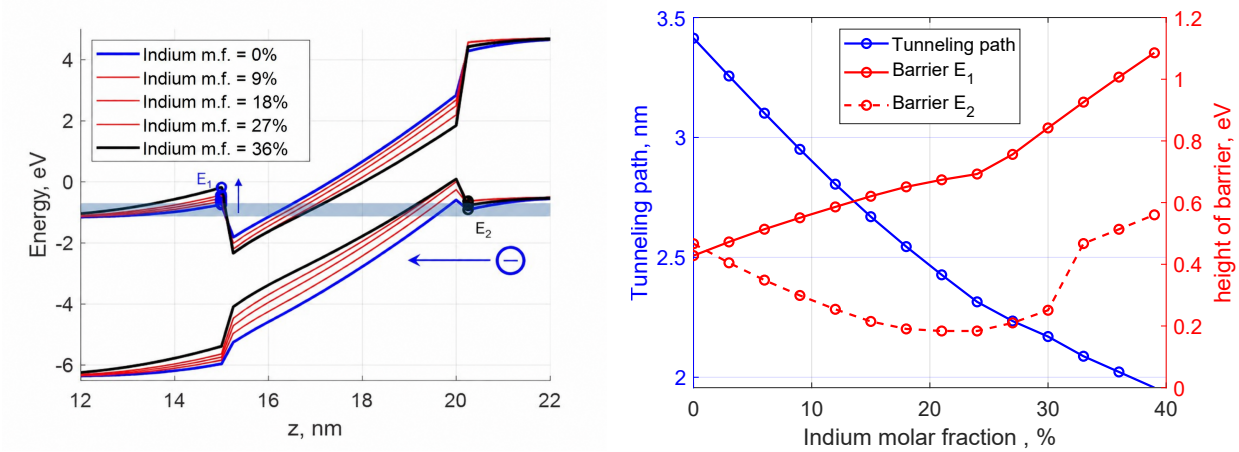


Fig. 5.33 (a) Band diagrams for different In molar fractions in the interlayer. (b) Variation of the intraband barrier heights E_1, at the n-layer/interlayer interface, and E_2, at the interlayer/p-layer interface, together with the interband tunneling path.

IV characteristic of the optimized tunnel junction is shown in Fig. 5.34, together with the corresponding band diagram and spectral current. The optimization of the interlayer geometry significantly increases the tunneling current density. Nevertheless, at low voltage, below approximately 0.3 V, the current remains lower than desired, so that a voltage of at least about 0.5 V is still required to supply the current density needed to drive the LED.

This residual limitation is mainly related to the p-side of the junction. The first contribution is the low ionization efficiency of Mg in Al-rich AlGaN. As discussed above, this reduces the effective

carrier density in the p-type layer and lowers the built-in electric field, so that a larger applied voltage is required to align the bands for direct interband tunneling. The second contribution is related to the valence-band structure at high Al molar fraction. In this composition range, the splitting and inversion of the light-hole and heavy-hole bands reduce the available valence-band density of states for interband tunneling. As a result, an energy interval of about 200 meV appears where only one valence-band branch effectively contributes to the tunneling process, reducing the overall current. A possible solution to improve the low-voltage current is to modify the p-side

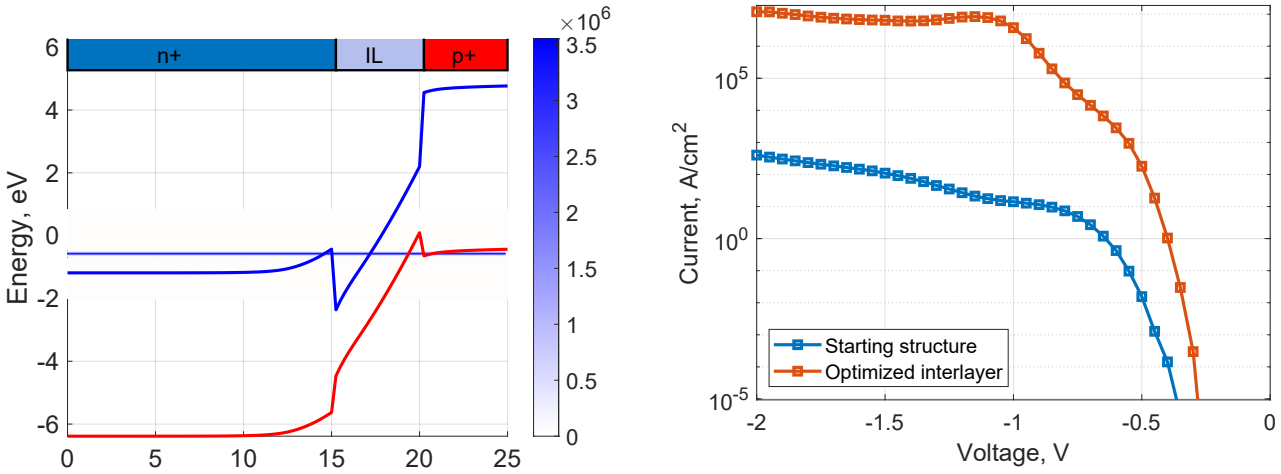


Fig. 5.34 (a) Band diagram and spectral current at −1.00 V of the tunnel junction with optimized 5 nm $In_{0.27}Ga_{0.73}N$ interlayer and (b) corresponding IV characteristic.

material. Replacing the Al-rich p-layer with GaN would reduce the valence-band splitting and improve Mg ionization, thereby increasing the tunneling current at low bias. However, this would also reintroduce strong UV-C absorption, defeating the purpose of replacing the p-GaN contact layer. A more balanced approach is therefore to combine the improved electrical properties of a lower-Al-content material with a limited optical penalty.

The solution proposed here is to introduce a steep 5 nm graded p-side layer, where the aluminum molar fraction changes from GaN to $Al_{0.50}Ga_{0.50}N$. This grading increases the effective p-type charge through polarization doping and improves the low-voltage tunneling current, while limiting the thickness of the UV-absorbing low-Al-content region. The band diagram, spectral current, and IV characteristic of this final optimized tunnel junction are shown in Fig. 5.35. With this structure, the tunnel junction adds only approximately 0.15 V to the LED operating voltage required to drive the device. A remaining question concerns the feasibility of growing this type of structure. InGaN is already difficult to grow on GaN in blue LEDs, because the lattice mismatch can lead to relaxation, crack formation, and other defects. For this reason, InGaN layers are usually limited to thicknesses of a few nanometers. Growth on AlGaN is even more

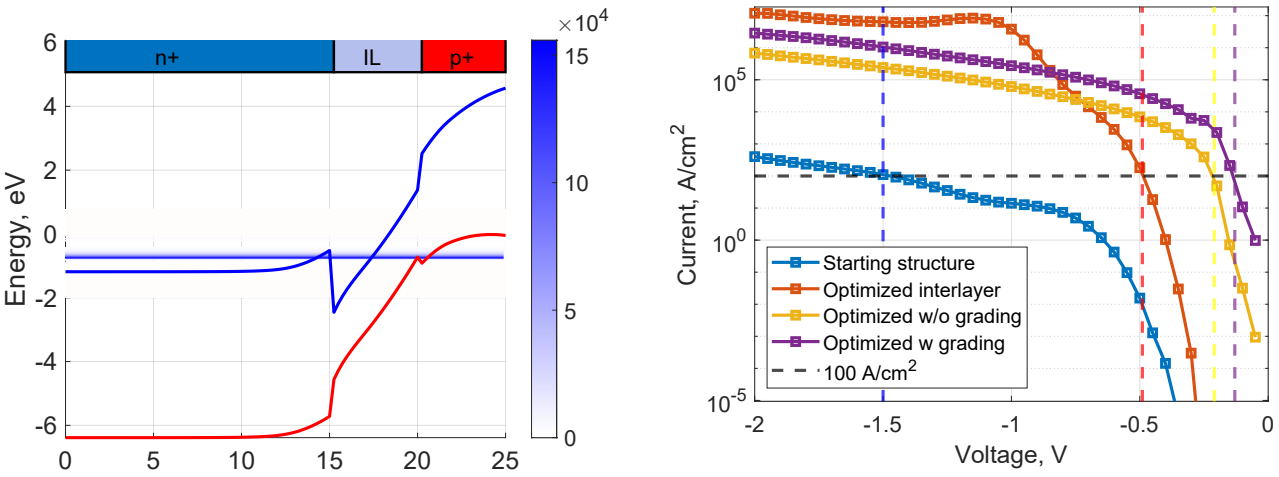


Fig. 5.35 (left) Band diagram and spectral current at −1.00 V of the tunnel junction with optimized 5 nm $In_{0.27}Ga_{0.73}N$ interlayer and optimized graded p-side layer and (right) corresponding IV characteristic.

challenging, especially by MOCVD, because of the different growth conditions required for InGaN and AlGaN. However, structures similar to the one proposed here have been successfully grown by plasma-assisted molecular beam epitaxy (PAMBE). For example, Zhang et al. [70] reported the growth of a 4 nm $In_{0.25}Ga_{0.75}N$ layer on $Al_{0.55}Ga_{0.45}N$.

Different models have been proposed to estimate the critical thickness before plastic relaxation occurs. One of the most widely used is the Matthews-Blakeslee model [77]. Applied to the present structure, this model gives a critical thickness of approximately 5.2 nm for $In_{0.27}Ga_{0.73}N$ grown on $Al_{0.75}Ga_{0.25}N$. This suggests that a similar structure could, in principle, be grown, although it would likely be challenging and expensive, requiring advanced epitaxial techniques such as PAMBE. Nevertheless, the optimization procedure proposed here remains useful as a guideline for identifying the best-performing structures among those that can realistically be fabricated.

The optimized tunnel junction must then be included in the full LED simulation. However, the larger size of the complete TJ-LED structure makes a full NEGF simulation of the entire device impractical. For this reason, the contribution of band-to-band tunneling (BTBT) in the tunnel junction is included in the Poisson-Schrödinger drift-diffusion framework as a generation term in the continuity equation, similarly to the approach used in Ref. [78]. The generation term, indicated as $G_{\mathrm{BTBT}}(z)$, is extracted from the IV characteristic of the tunnel junction obtained by NEGF:

$$G_{\mathrm{BTBT}}(z) = \frac{J_{\mathrm{TJ}}(V_{\mathrm{TJ}})}{qL_{\mathrm{TJ}}}. \tag{5.8}$$

Here, L_{TJ} is the length of the region where BTBT takes place, while V_{TJ} is the voltage drop across the tunnel junction, defined as the difference between the quasi-Fermi levels at the two ends of the TJ. In this way, the interband tunneling contribution obtained from the NEGF simulation is included in the device-level PSDD calculation without requiring a full NEGF treatment of the entire LED.

The optimized tunnel junction is then integrated into the UV-C LED structure. In this configuration, the absorbing p-GaN contact layer is removed and replaced by the tunnel junction followed by an n-type contact. The band diagram of the TJ-LED at 5.45 V is shown in Fig. 5.36(a). By examining the quasi-Fermi-level separation in the flat regions, the voltage drop across the tunnel junction and across the LED can be estimated. Thanks to the optimization described above, only 0.15 V is dropped across the TJ, while 5.30 V falls across the LED.

The effect of the tunnel junction on the device can also be seen in Fig. 5.36(b), which compares the IV characteristic of the TJ-LED with that of a conventional AlGaN LED using p-GaN as the contact layer. At low voltages, only a small current flows through the device, so the TJ requires very little voltage to drive the LED and the two IV curves remain very similar. As the voltage increases, the voltage drop across the TJ becomes more significant, leading to a small current reduction around 5.20 V with respect to the conventional structure. Since the TJ is highly optimized, this reduction remains limited, but it can still affect the final wall-plug efficiency.

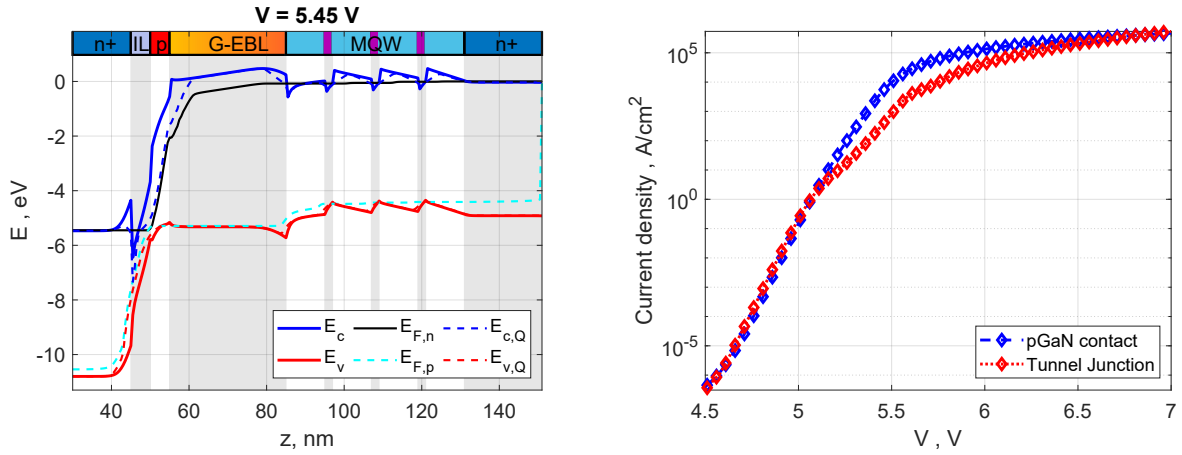


Fig. 5.36 (a) Band diagram of the optimized TJ-LED structure at 5.45 V. (b) IV characteristic of the TJ-LED (red) compared to a conventional structure with p-GaN instead of the TJ (blue).

5.2.3 Efficiency comparison

The three p-side contact solutions are now compared in terms of internal quantum efficiency (IQE), external quantum efficiency (EQE), and wall-plug efficiency (WPE), as defined in Eq. (5.1). The internal quantum efficiency of the LEDs with p-GaN contact, polarization-doped graded contact, and tunnel-junction contact is shown in Fig. 5.37. All three structures reach a similar maximum IQE of approximately 57%. This indicates that the contact-layer design does not significantly modify the recombination balance in the active region at a fixed current density, since all the peaks coincide for the same current density. However, the TJ-LED reaches its maximum IQE at a slightly higher voltage because of the additional voltage drop introduced by the tunnel junction. The light extraction efficiency is then evaluated as the fraction of photons generated in

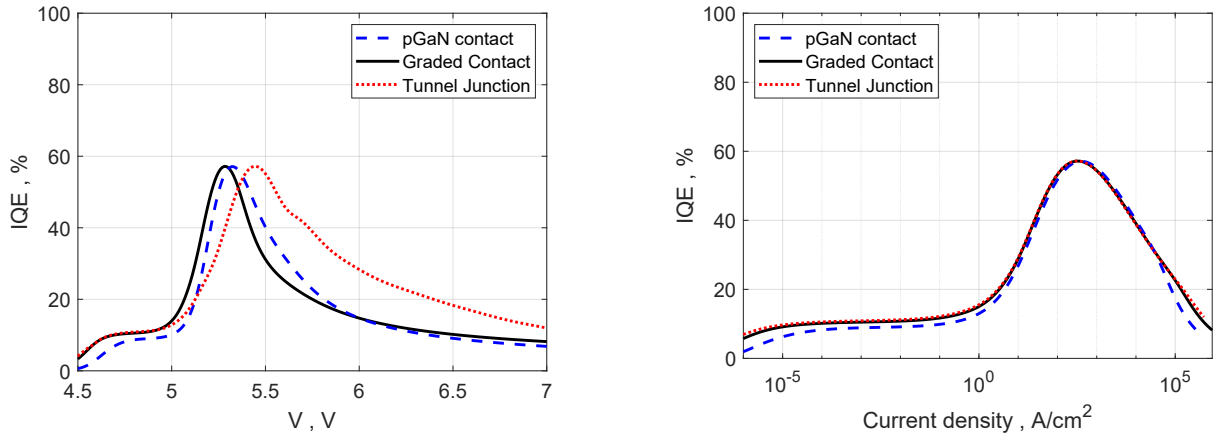


Fig. 5.37 IQE of the UV-C AlGaN LED vs voltage (left) and vs current density (right), with three different p-contact layers: conventional p-GaN contact (dashed blue line), polarization-doped graded contact (solid black line), and tunnel-junction contact (dotted red line).

the quantum wells that successfully escape the device. It is estimated using a ray-tracing model, where photons are emitted isotropically from the quantum wells and propagated as geometrical rays. At each material interface, refraction and reflection are computed from Snell's law and the Fresnel coefficients,

$$n_1 \sin\theta_1 = n_2 \sin\theta_2, \tag{5.9}$$

$$R = \left| \frac{n_1 \cos\theta_1 - n_2 \cos\theta_2}{n_1 \cos\theta_1 + n_2 \cos\theta_2} \right|^2, \tag{5.10}$$

where n_1 and n_2 are the complex refractive indices of the two materials at the interface, while θ_1 and θ_2 are the incident and transmitted angles, respectively. For simplicity, an AlN/sapphire substrate is considered, since this is a common platform for UV-C LED growth. More complex

substrate or interface structures, sometimes introduced to improve epitaxial quality or light extraction, are not included in this analysis. Absorption is accounted for through an exponential attenuation of the ray intensity,

$$I(x) = I_0 \exp(-\alpha x), \tag{5.11}$$

where α is the absorption coefficient of the material and x is the propagation distance inside the layer. The absorption coefficient is computed for the different structures using Fermi's golden rule, in order to account for band mixing in the tunnel-junction contact region. Figure 5.38 shows the angular distribution of the emitted optical power collected from the device. Because of the large refractive-index contrast between the semiconductor layers and air, the escape cone is narrow and a large fraction of the emitted light remains trapped inside the structure. Since the three devices share the same substrate side, the main differences between them are associated with the rays emitted toward the p-side. Therefore, optimizing the p-contact region can at most recover the fraction of light that would otherwise be absorbed or poorly reflected in the upper part of the device; it cannot remove the extraction losses associated with the escape cone and light trapping. In the conventional p-GaN contact, a large fraction of the light emitted toward the p-side is absorbed by the GaN layer before it can be reflected back toward the substrate. As a result, the LEE is limited to 10.59%, leading to an EQE of 6.01% when combined with the 57% IQE.

In the graded-contact structure, the collected optical power increases over the relevant emission angles. This improvement is due to the higher aluminum content of the graded layer, which makes it essentially transparent to UV-C radiation. Nevertheless, the LEE remains limited to 12.89%, corresponding to an EQE of 7.35%. The main remaining limitation is the metal contact: because high-work-function metals such as Au have low UV-C reflectivity, only a limited fraction of the upward-emitted light is reflected back toward the substrate and extracted.

The tunnel-junction contact gives the highest extracted optical power among the three structures. In this case, the p-GaN contact can be replaced by an n-type contact, allowing the use of a more reflective metal such as aluminum. The LEE increases to 18.61%, corresponding to an EQE of 10.60%. This represents a 75.50% improvement with respect to the conventional p-GaN-contact LED. Some optical power is still lost because the optimized tunnel junction is not fully transparent, mainly due to the presence of the InGaN interlayer, but the reduction of p-side absorption and the higher metal reflectivity more than compensate this loss. The external quantum efficiency and wall-plug efficiency are shown in Fig. 5.39 and summarized in Table 5.3. The wall-plug efficiency is defined as the ratio between the emitted optical power and the electrical power supplied to the device. Since the graded contact and tunnel-junction contact both increase the extracted optical power, they also improve the WPE. For the reference p-GaN contact, a peak

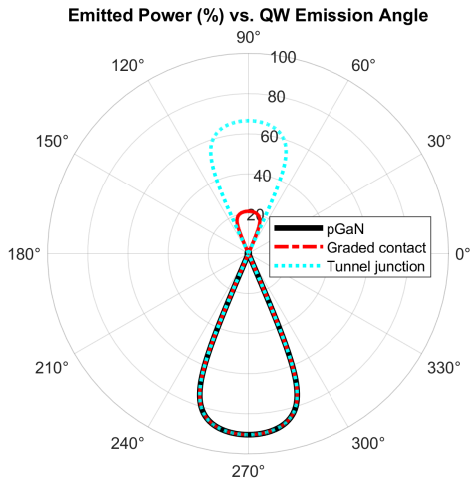


Fig. 5.38 Emitted power percentage as a function of the emission angle from the quantum well for three p-contact layers: p-GaN contact, polarization-doped graded contact, and tunnel-junction contact.

WPE of 5.08% is obtained. The polarization-doped graded contact increases the peak WPE to 6.25%, corresponding to a 23.00% improvement. This improvement is slightly larger than the EQE improvement, because the efficiency peak is reached at a slightly lower voltage as a consequence of the smoother band profile and improved effective doping. The tunnel-junction LED reaches a peak WPE of 8.78%, corresponding to a 72.62% improvement with respect to the p-GaN-contact device. This relative gain is lower than the EQE improvement because of the additional voltage drop introduced by the tunnel junction. However, thanks to the tunnel-junction optimization discussed above, the voltage penalty remains small and the WPE advantage is largely preserved. The dependence of the WPE improvement on the tunnel-junction voltage

Table 5.3 Internal quantum efficiency (IQE), light extraction efficiency (LEE), external quantum efficiency (EQE), and wall-plug efficiency (WPE) for the UV-C LEDs with p-GaN contact, polarization-doped graded contact (GC), and tunnel-junction contact (TJ).

	p-GaN	GC	TJ
IQE	57%	57%	57%
LEE	10.59%	12.89%	18.61%
EQE	6.01%	7.35%	10.60%
WPE	5.08%	6.25%	8.78%

drop can be estimated from

$$WPE = EQE \frac{\hbar\omega}{q\,(V_{\mathrm{LED}} + V_{\mathrm{TJ}})}, \tag{5.12}$$

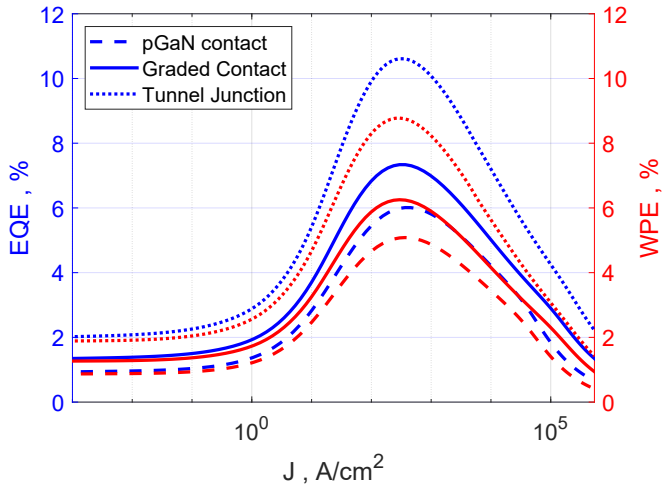


Fig. 5.39 EQE (blue lines, left axis) and WPE (red lines, right axis) of the UV-C AlGaN LED with three different p-contact layers: conventional p-GaN contact (dashed lines), polarization-doped graded contact (solid lines), and tunnel-junction contact (dotted lines).

where V_{LED} is the voltage drop across the LED part of the structure and V_{TJ} is the voltage drop across the tunnel junction. The relative WPE improvement with respect to the p-GaN-contact LED is shown in Fig. 5.40 as a function of V_{TJ}. As the tunnel-junction voltage increases, the WPE advantage progressively decreases. For voltage drops of the order of 4 V, the tunnel-junction contact no longer improves the WPE and may even become detrimental. This confirms that the optical benefit of the tunnel-junction contact can be fully exploited only if the junction is carefully optimized to minimize the voltage penalty. In conclusion, both proposed p-side contact solutions can improve the performance of UV-C LEDs with respect to the conventional p-GaN contact. The tunnel-junction contact provides the largest improvement in EQE and WPE, because it removes the absorbing p-GaN layer and enables the use of a highly reflective n-type metal contact. However, this solution requires careful tunnel-junction optimization to avoid excessive voltage losses, and it may increase fabrication complexity and cost because of the need for MBE growth. The polarization-doped graded contact provides a smaller improvement, but it is technologically simpler and can be grown by MOCVD. These results highlight the trade-off between maximum performance and manufacturability, and show how quantum-corrected device simulations can be used to identify realistic optimization paths for AlGaN UV-C LEDs.

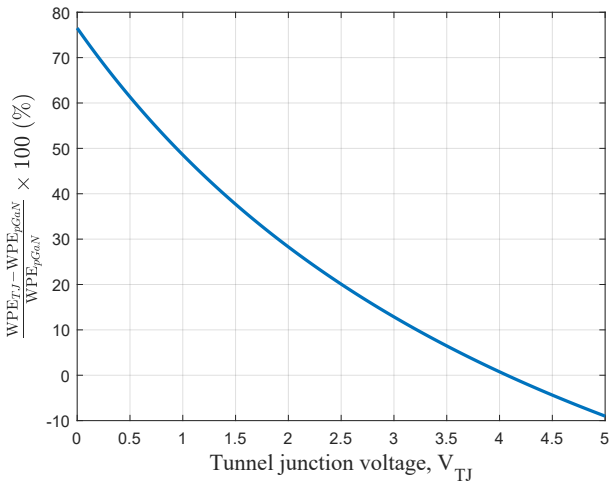


Fig. 5.40 Relative WPE improvement of the AlGaN UV-C LED when using a tunnel-junction contact instead of a p-GaN contact layer, plotted as a function of the tunnel-junction voltage.

Chapter 6

Conclusions and outlook

This work investigated the use of advanced quantum transport and quantum-corrected drift-diffusion models for the simulation of III-nitride optoelectronic devices. Several mechanisms governing the operation of modern nitride LEDs cannot be described accurately within a purely semiclassical framework. Quantum confinement, tunneling and interband tunneling in tunnel junctions all affect carrier injection, recombination, and optical efficiency. Therefore, predictive simulations of these devices require models able to include quantum effects while remaining computationally feasible for realistic structures.

NEGF is the most rigorous approach considered in this work, since it is a genuine quantum-kinetic method that provides direct access to spectral quantities such as the local density of states, spectral currents, open-boundary effects, and tunneling transport. However, its computational cost makes it difficult to apply to full LED structures or to optimization studies involving many design variations. This motivates the use of reduced quantum-corrected approaches, such as density-gradient and Poisson-Schrödinger drift-diffusion, which retain the numerical efficiency of drift-diffusion while introducing quantum effects through effective quantum potentials or quantum carrier densities.

The NEGF analysis of the subthreshold regime of an InGaN/GaN LED showed that the quantum density of states and the associated tunneling-related charge are essential to describe the enhancement of the recombination current due to trap-assisted tunneling, a mechanism commonly observed in III-nitride devices. The multi-phonon recombination model also showed that the total quantum charge is not always sufficient to determine the recombination rate, because only the spectral charge effectively coupled to the selected trap contributes to the capture process.

When the quantum-corrected drift-diffusion models were compared with the NEGF framework, the Poisson-Schrödinger approach reproduced the NEGF behavior with good accuracy in the

TAT-dominated regime. This is because it retains an explicit spectral representation of the quantum charge and allows the recombination rate to be inspected in both position and energy. The density-gradient method, instead, captured the qualitative effect of quantum corrections and reproduced the emergence of high ideality factors, but its approximate quantum charge led to larger deviations in the EBL/QB tunneling region. This comparison identified the Poisson-Schrödinger drift-diffusion model as the most suitable compromise between physical accuracy and computational cost among the reduced models considered in this work.

The application of the Poisson-Schrödinger drift-diffusion framework to AlGaN-based UV-C LEDs demonstrated that quantum-corrected simulations can serve not only as a tool for understanding transport mechanisms, but also as a practical framework for device design. Unlike visible III-nitride LEDs, UV-C devices operate in a regime where material limitations, polarization effects, carrier confinement, and optical losses become increasingly intertwined. As a result, the impact of a structural modification cannot be assessed by considering a single figure of merit in isolation. Instead, the overall device performance emerges from the interplay between carrier injection, leakage, recombination, optical absorption, and electrical efficiency, making a comprehensive simulation approach essential for identifying meaningful design improvements.

The optimization of the electron-blocking region highlighted this balance. Polarization-doped and superlattice EBLs can enhance hole injection by reshaping the electrostatic profile and by introducing quantum-corrected transport paths through thin layers. However, the same design choices that favor hole transport can also weaken electron confinement if the barrier becomes too transparent. The relevant optimization parameter is therefore not simply the hole density or the barrier height alone, but the compromise between hole injection, electron blocking, and recombination in the active region.

A similar trade-off emerged in the p-side contact design. Removing or reducing the absorbing p-GaN contact is beneficial for light extraction, but the electrical cost of the alternative contact scheme must be considered at the same time. The polarization-doped graded contact provides a technologically simpler route to improve transparency and carrier injection, while the tunnel-junction contact gives the largest optical benefit when its voltage drop is sufficiently low. In this sense, the comparison between the conventional p-GaN contact, the graded contact, and the tunnel-junction contact shows that the best UV-C LED design cannot be identified from optical or electrical considerations alone, but requires a combined analysis of carrier transport, recombination, light extraction, and voltage efficiency.

Overall, the work demonstrates that quantum-transport modeling is essential for understanding and optimizing III-nitride optoelectronic devices whose operation depends strongly on tunneling and confinement. NEGF remains the most complete and physically rigorous tool, and is therefore

well suited for microscopic validation and for the analysis of intrinsically quantum devices such as tunnel junctions. However, its computational cost limits its use in large-scale optimization. Quantum-corrected drift-diffusion, and in particular the Poisson-Schrödinger approach, provides a practical bridge between microscopic quantum transport and device-level design. It retains the quantum information needed to describe confinement, tunneling-assisted recombination, and carrier injection across heterostructure barriers, while remaining efficient enough for optimization studies.

The results also highlight that device optimization in III-nitrides requires considering electrical transport and optical extraction simultaneously. A structure that improves injection or internal quantum efficiency does not necessarily maximize the final wall-plug efficiency if it introduces additional voltage losses or optical absorption. For this reason, the most promising designs are those that balance carrier injection, leakage suppression, transparency, and voltage efficiency. In the UV-C LED case studied here, both polarization-doped graded contacts and tunnel-junction contacts improve the performance with respect to the conventional p-GaN contact, but they represent different trade-offs between efficiency gain, voltage penalty, and fabrication complexity.

Future work may extend these simulations toward more complete device descriptions. A first direction is the inclusion of multidimensional effects, such as lateral current spreading, current crowding, and realistic contact geometries, which cannot be captured in a one-dimensional framework. A second important direction is the detailed quantum study of metal/semiconductor contacts. Schottky barriers are particularly relevant in III-nitride devices because the large band offsets, high contact barriers, and relatively low effective doping in some layers can make carrier injection strongly dependent on tunneling and barrier-shape effects. A quantum-transport analysis of these contacts could therefore provide useful insight into contact resistance, injection efficiency, and the validity of simplified boundary conditions used in device simulations. Finally, further NEGF studies of tunnel junctions and tunneling-assisted recombination could be used to develop physically grounded reduced models for quantum-corrected drift-diffusion simulations. Such a multiscale strategy would preserve the microscopic accuracy of quantum transport where it is needed, while enabling efficient optimization of complete III-nitride optoelectronic devices.

References

[1] I. Akasaki. High efficiency blue led utilizing gan film with aln buffer layer grown by movpe. *Inst. Phys. Conf. Ser.*, 91:633–636, 1988.

[2] Shuji Nakamura, Takashi Mukai, and Masayuki Senoh. Candela-class high-brightness ingan/algan double-heterostructure blue-light-emitting diodes. *Applied Physics Letters*, 64(13):1687–1689, 03 1994.

[3] Isamu Akasaki. Nobel lecture: Fascinated journeys into blue light. *Rev. Mod. Phys.*, 87:1119–1131, Oct 2015.

[4] I. Vurgaftman and J. R. Meyer. Band parameters for nitrogen-containing semiconductors. *J. Appl. Phys.*, 94(6):3675–3696, September 2003.

[5] Nicholas G. Reed. The history of ultraviolet germicidal irradiation for air disinfection. *Public Health Reports*, 125(1):15–27, 2010.

[6] Vijay Kumar Sharma et al. Bright future of deep-ultraviolet photonics: Emerging uvc chip-scale light-source technology platform. *ACS Photonics*, 2022.

[7] Michael Kneissl, Tae-Yeon Seong, Jung Han, and Hiroshi Amano. The emergence and prospects of deep-ultraviolet light-emitting diode technologies. *Nature photonics*, 13(4):233–244, 2019.

[8] 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 Letters*, 23(9):535–537, 2002.

[9] C. L. Reynolds and A. Patel. Tunneling entity in different injection regimes of InGaN light emitting diodes. *J. Appl. Phys.*, 103(8):086102, 2008.

[10] Kengo Nagata, Taichi Matsubara, Yoshiki Saito, Keita Kataoka, Tetsuo Narita, Kayo Horibuchi, Maki Kushimoto, Shigekazu Tomai, Satoshi Katsumata, Yoshio Honda, Tetsuya Takeuchi, and Hiroshi Amano. A review on the progress of algan tunnel homojunction deep-ultraviolet light-emitting diodes. *Crystals*, 13(3), 2023.

[11] Sriram Krishnamoorthy, Fatih Akyol, and Siddharth Rajan. Ingan/gan tunnel junctions for hole injection in gan light emitting diodes. *Applied Physics Letters*, 105(14):141104, 10 2014.

[12] A. L. Fetter and J. D. Walecka. *Quantum Theory of Many-Particle Physics*. Dover Publications, New York, 3rd edition, 2003.

[13] Takeo Matsubara. A new approach to quantum-statistical mechanics. *Progress of Theoretical Physics*, 14(4):351–378, 10 1955.

[14] Leo P. Kadanoff and Gordon Baym. *Quantum Statistical Mechanics. Green's Function Methods in Equilibrium and Nonequilibrium Problems*. W. A. Benjamin, New York, 1962.

[15] L. V. Keldysh. Diagram technique for nonequilibrium processes. *Sov. Phys. JETP*, 20(4):1018–1026, April 1965.

[16] Paul C. Martin and Julian Schwinger. Theory of many-particle systems. i. *Phys. Rev.*, 115:1342–1373, Sep 1959.

[17] Gyungock Kim and Gerald B. Arnold. Theoretical study of tunneling phenomena in double-barrier quantum-well heterostructures. *Phys. Rev. B*, 38:3252–3262, Aug 1988.

[18] Gianluca Stefanucci and Robert van Leeuwen. *Nonequilibrium Many-Body Theory Of Quantum Systems. A Modern Introduction*. Cambridge University Press, Cambridge, U.K., 2013.

[19] Jesus Alberto Gonzalez Montoya. *Full-band NEGF modeling of optoelectronic devices*. PhD thesis, Politecnico di Torino, June 2023.

[20] Urs Aeberhard. *A Microscopic Theory of Quantum Well Photovoltaics*. PhD thesis, Eidgenössische Technische Hochschule Zürich, 2008.

[21] Supriyo Datta. *Electronic Transport in Mesoscopic Systems*. Cambridge University Press, Cambridge, U.K., 1995.

[22] Xiangyu Zhou, Francesco Bertazzi, Michele Goano, Giovanni Ghione, and Enrico Bellotti. Deriving $k \cdot p$ parameters from full-Brillouin-zone descriptions: A finite-element envelope function model for quantum-confined wurtzite nanostructures. *J. Appl. Phys.*, 116(3):033709, July 2014.

[23] R. G. Veprek. *Computational modeling of semiconductor nanostructures for optoelectronics*. PhD thesis, Eidgenössische Technische Hochschule Zürich, 2009.

[24] Sebastian Steiger. *Modelling Nano-LEDs*. PhD thesis, Eidgenössische Technische Hochschule Zürich, 2009.

[25] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic. Single and multiband modeling of quantum electron transport through layered semiconductor devices. *J. Appl. Phys.*, 81(12):7845–7869, February 1997.

[26] Andreas Schenk. An improved approach to the Shockley-Read-Hall recombination in inhomogeneous fields of space-charge regions. *J. Appl. Phys.*, 71(7):3339–3349, 1992.

[27] Urs Aeberhard. Nonequilibrium Green's function picture of nonradiative recombination of the Shockley-Read-Hall type. *Phys. Rev. B*, 99:125302, 2019.

[28] Jesus Alberto Gonzalez Montoya, Alberto Tibaldi, Carlo De Santi, Matteo Meneghini, Michele Goano, and Francesco Bertazzi. Nonequilibrium Green's function modeling of trap-assisted tunneling in $In_xGa_{1-x}N$/GaN light-emitting diodes. *Phys. Rev. Appl.*, 16(4):044023, October 2021.

[29] A. Svizhenko, M. P. Anantram, T. R. Govindan, B. Biegel, and R. Venugopal. Two-dimensional quantum mechanical modeling of nanotransistors. *Journal of Applied Physics*, 91(4):2343–2354, 02 2002.

[30] Zhibin Ren, R. Venugopal, S. Goasguen, S. Datta, and M.S. Lundstrom. nanomos 2.5: A two-dimensional simulator for quantum transport in double-gate mosfets. *IEEE Transactions on Electron Devices*, 50(9):1914–1925, 2003.

[31] Peter Greck, Stefan Birner, Bernhard Mittermeier, and Peter Vogl. Efficient method for the calculation of dissipative quantum transport in quantum cascade lasers. *Optics Express*, 23:6587–6600, 03 2015.

[32] Jesse A. Vaitkus and Jared H. Cole. Büttiker probes and the recursive Green's function: an efficient approach to include dissipation in general configurations. *Phys. Rev. B*, 97(8):085149, 2018.

[33] Alberto Tibaldi, Jesus Alberto Gonzalez Montoya, Marco Vallone, Michele Goano, Enrico Bellotti, and Francesco Bertazzi. Modeling infrared superlattice photodetectors: From nonequilibrium Green's functions to quantum-corrected drift diffusion. *Phys. Rev. Appl.*, 16(4):044024, October 2021.

[34] W. Van Roosbroeck. Theory of the flow of electrons and holes in germanium and other semiconductors. *The Bell System Technical Journal*, 29(4):560–607, 1950.

[35] P.A. Markowich, C.A. Ringhofer, and C. Schmeiser. *Semiconductor Equations*. Mathematics and Statistics. Springer Vienna, 2012.

[36] S. Selberherr. *Analysis and Simulation of Semiconductor Devices*. Springer-Verlag, Wien, 1984.

[37] D. L. Scharfetter and H. K. Gummel. Large-signal analysis of a silicon read diode transistor. *IEEE Trans. Electron Devices*, ED-16(1):64–77, January 1969.

[38] Marco Mandurrino, Giovanni Verzellesi, Michele Goano, Marco Vallone, Francesco Bertazzi, Giovanni Ghione, Matteo Meneghini, Gaudenzio Meneghesso, and Enrico Zanoni. Physics-based modeling and experimental implications of trap-assisted tunneling in InGaN/GaN light-emitting diodes. *Phys. Status Solidi A*, 212(5):947–953, 2015.

[39] W. Shockley and W. T. Read. Statistics of the recombinations of holes and electrons. *Phys. Rev.*, 87(5):835–842, 1952.

[40] M. G. Ancona. Density-gradient theory: a macroscopic approach to quantum confinement and tunneling in semiconductor devices. *J. Comp. Electron.*, 10:65–97, 2011.

[41] Carl L. Gardner. The quantum hydrodynamic model for semiconductor devices. *Journal on Applied Mathematics*, 54(2):409–427, 1994.

[42] Giovanni Verzellesi, Davide Saguatti, Matteo Meneghini, Francesco Bertazzi, Michele Goano, Gaudenzio Meneghesso, and Enrico Zanoni. Efficiency droop in InGaN/GaN blue light-emitting diodes: Physical mechanisms and remedies. *J. Appl. Phys.*, 114(7):071101, August 2013.

[43] 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*, 39(2):331–338, 1992.

[44] Nicola Roccato, Francesco Piva, Carlo De Santi, Riccardo Brescancin, Kalparupa Mukherjee, Matteo Buffolo, Camille Haller, Jean-François Carlin, Nicolas Grandjean, Marco Vallone, Alberto Tibaldi, Francesco Bertazzi, Michele Goano, Giovanni Verzellesi, Gaudenzio Meneghesso, Enrico Zanoni, and Matteo Meneghini. Modeling the electrical characteristics of ingan/gan led structures based on experimentally-measured defect characteristics. *Journal of Physics D: Applied Physics*, 54(42):425105, aug 2021.

[45] M. Auf der Maur, B. Galler, I. Pietzonka, M. Strassburg, H. Lugauer, and A. Di Carlo. Trap-assisted tunneling in InGaN/GaN single-quantum-well light-emitting diodes. *Appl. Phys. Lett.*, 105(13):133504, 2014.

[46] Fabio Bernardini and Vincenzo Fiorentini. Nonlinear macroscopic polarization in iii-v nitride alloys. *Phys. Rev. B*, 64:085207, Aug 2001.

[47] Hiroko Inagaki, Akatsuki Saito, Hironobu Sugiyama, Tamaki Okabayashi, and Shouichi Fujimoto. Rapid inactivation of sars-cov-2 with deep-uv led irradiation. *Emerging Microbes & Infections*, 9(1):1744–1747, 2020.

[48] Jose-Luis Sagripanti and C David Lytle. Estimated inactivation of coronaviruses by solar radiation with special reference to covid-19. *Photochemistry and photobiology*, 96(4):731–737, 2020.

[49] Shinji Kobayashi and Akinori Hatano. High-intensity low-pressure electrodeless mercury-argon lamp for uv disinfection of wastewater. *Journal of Water and Environment Technology*, 3(1):71–76, 2005.

[50] Ram Prakash, Afaque M Hossain, UN Pal, N Kumar, K Khairnar, and M Krishna Mohan. Dielectric barrier discharge based mercury-free plasma uv-lamp for efficient water disinfection. *Scientific reports*, 7(1):17426, 2017.

[51] Kari A Sholtes, Kincaid Lowe, Glenn W Walters, Mark D Sobsey, Karl G Linden, and Lisa M Casanova. Comparison of ultraviolet light-emitting diodes and low-pressure mercury-arc lamps for disinfection of water. *Environmental Technology*, 37(17):2183–2188, 2016.

[52] Raffaella Pizzichetti, Mario Martín-Gamboa, Cristina Pablos, Ken Reynolds, Simon Stanley, Javier Dufour, and Javier Marugán. Environmental life cycle assessment of uv-c leds vs. mercury lamps and oxidant selection for diclofenac degradation. *Sustainable Materials and Technologies*, 41:e01002, 2024.

[53] Yukio Narukawa, Masatsugu Ichikawa, Daisuke Sanga, Masahiko Sano, and Takashi Mukai. White light emitting diodes with super-high luminous efficacy. *Journal of Physics D: Applied Physics*, 43(35):354002, aug 2010.

[54] Noritoshi Maeda, Masafumi Jo, and Hideki Hirayama. Improving the efficiency of algan deep-uv leds by using highly reflective ni/al p-type electrodes. *Physica Status Solidi (a)*, 215(8):1700435, 2018.

[55] Yusuke Matsukura, Tetsuhiko Inazu, Cyril Pernot, Naoki Shibata, Maki Kushimoto, Manato Deki, Yoshio Honda, and Hiroshi Amano. Improving light output power of algan-based deep-ultraviolet light-emitting diodes by optimizing the optical thickness of p-layers. *Applied Physics Express*, 14(8):084004, jul 2021.

[56] Kazuhito Ban, Jun-ichi Yamamoto, Kenichiro Takeda, Kimiyasu Ide, Motoaki Iwaya, Tetsuya Takeuchi, Satoshi Kamiyama, Isamu Akasaki, and Hiroshi Amano. Internal quantum efficiency of whole-composition-range algan multiquantum wells. *Applied physics express*, 4(5):052101, 2011.

[57] Jiaming Wang, Fujun Xu, Lisheng Zhang, Jing Lang, Xuzhou Fang, Ziyao Zhang, Xueqi Guo, Chen Ji, Chengzhi Ji, Fuyun Tan, et al. Progress in efficient doping of al-rich algan. *Journal of Semiconductors*, 45(2):021501, 2024.

[58] KB Nam, ML Nakarmi, J Li, JY Lin, and HX Jiang. Mg acceptor level in aln probed by deep ultraviolet photoluminescence. *Applied Physics Letters*, 83(5):878–880, 2003.

[59] Gregor Hofmann, Anton Muhin, Norman Susilo, Friedhard Römer, Tim Wernicke, Michael Kneissl, and Bernd Witzigmann. Simulation of carrier injection efficiency in algan-based uv-light-emitting diodes. *IEEE Photonics Journal*, 16(4):1–7, 2024.

[60] Hayata Takahata, Tomoaki Kachi, Naoki Hamashima, Ryunosuke Oka, Hisanori Ishiguro, Tetsuya Takeuchi, Satoshi Kamiyama, Motoaki Iwaya, Yoshiki Saito, and Koji Okuno. Hole generation in polarization-doped alxga1–xn (x= 0.9–0.35)-graded layer with heavily mg-doped al0. 35ga0. 65n contact layer for 275 nm deep-ultraviolet light-emitting diode. *Physica Status Solidi (a)*, page 2400054, 2024.

[61] Ryota Ishii, Shiki Tanaka, Norman Susilo, Tim Wernicke, Michael Kneissl, Mitsuru Funato, and Yoichi Kawakami. Radiative and nonradiative recombination processes in algan quantum wells on epitaxially laterally overgrown aln/sapphire from 10 to 500 K. *physica status solidi (b)*, 261(11):2400018, 2024.

[62] John Simon, Vladimir Protasenko, Chuanxin Lian, Huili Xing, and Debdeep Jena. Polarization-induced hole doping in wide–band-gap uniaxial semiconductor heterostructures. *Science*, 327(5961):60–64, 2010.

[63] Shibin Li, Morgan Ware, Jiang Wu, Paul Minor, Zhiming Wang, Zhiming Wu, Yadong Jiang, and Gregory J. Salamo. Polarization induced pn-junction without dopant in graded algan coherently strained on gan. *Applied Physics Letters*, 101(12):122103, 09 2012.

[64] Peter Kozodoy, Yulia P. Smorchkova, Monica Hansen, Huili Xing, Steven P. DenBaars, Umesh K. Mishra, A. W. Saxler, R. Perrin, and W. C. Mitchel. Polarization-enhanced mg doping of algan/gan superlattices. *Applied Physics Letters*, 75(16):2444–2446, 10 1999.

[65] Chao Liu, Taiping Lu, Lejuan Wu, Hailong Wang, Yian Yin, Guowei Xiao, Yugang Zhou, and Shuti Li. Enhanced performance of blue light-emitting diodes with ingan/gan superlattice as hole gathering layer. *IEEE Photonics Technology Letters*, 24(14):1239–1241, 2012.

[66] Yun Yan Zhang and Yi An Yin. Performance enhancement of blue light-emitting diodes with a special designed AlGaN/GaN superlattice electron-blocking layer. *Applied Physics Letters*, 99(22):221103, 11 2011.

[67] Songqing Liu, Chunya Ye, Xuefen Cai, Shuping Li, Wei Lin, and Junyong Kang. Performance enhancement of algan deep-ultraviolet light-emitting diodes with varied superlattice barrier electron blocking layer. *Applied Physics A*, 122:1–6, 2016.

[68] Wenhong Sun, Vinod Adivarahan, Maxim Shatalov, Youngbae Lee, Shuai Wu, Jinwei Yang, Jianping Zhang, and M Asif Khan. Continuous wave milliwatt power algan light emitting diodes at 280 nm. *Japanese journal of applied physics*, 43(11A):L1419, 2004.

[69] Nicola Roccato, Francesco Piva, Carlo De Santi, Matteo Buffolo, Normal Susilo, Daniel Hauer Vidal, Anton Muhin, Luca Sulmoni, Tim Wernicke, Micheal Kneissl, Gaudenzio Meneghesso, Enrico Zanoni, and Matteo Meneghini. Modeling of the electrical characteristics and degradation mechanisms of uv-c leds. *IEEE Photonics Journal*, 16(1):1–6, 2024.

[70] Yuewei Zhang, Sriram Krishnamoorthy, Fatih Akyol, Andrew A Allerman, Michael W Moseley, Andrew M Armstrong, and Siddharth Rajan. Design and demonstration of ultra-wide bandgap algan tunnel junctions. *Applied Physics Letters*, 109(12), 2016.

[71] Xu Liu, Zhenxing Lv, Zhefu Liao, Yuechang Sun, Ziqi Zhang, Ke Sun, Qianxi Zhou, Bin Tang, Hansong Geng, Shengli Qi, et al. Highly efficient algan-based deep-ultraviolet light-emitting diodes: from bandgap engineering to device craft. *Microsystems & Nanoengineering*, 10(1):110, 2024.

[72] Shengjun Zhou, Zhefu Liao, Ke Sun, Ziqi Zhang, Yinzuo Qian, Pengfei Liu, Peng Du, Jingjing Jiang, Zhenxing Lv, and Shengli Qi. High-power algan-based ultrathin tunneling junction deep ultraviolet light-emitting diodes. *Laser & Photonics Reviews*, 18(1):2300464, 2024.

[73] Agnes Maneesha Dominic Merwin Xavier, Arnob Ghosh, Sheikh Ifatur Rahman, Andrew Allerman, Shamsul Arafin, and Siddharth Rajan. Design and demonstration of efficient transparent 30% al-content algan interband tunnel junctions. *Applied Physics Letters*, 122(8):081108, 02 2023.

[74] Sriram Krishnamoorthy, Digbijoy N Nath, Fatih Akyol, Pil Sung Park, Michele Esposto, and Siddharth Rajan. Polarization-engineered gan/ingan/gan tunnel diodes. *Applied Physics Letters*, 97(20), 2010.

[75] A Pandey, WJ Shin, J Gim, R Hovden, and Z Mi. High-efficiency algan/gan/algan tunnel junction ultraviolet light-emitting diodes. *Photonics Research*, 8(3):331–337, 2020.

[76] F. Bernardini, V. Fiorentini, and D. Vanderbilt. Spontaneous polarization and piezoelectric constants of III-V nitrides. *Phys. Rev. B*, 56(16):R10024–R10027, October 1997.

[77] J.W. Matthews and A.E. Blakeslee. Defects in epitaxial multilayers: I. misfit dislocations. *Journal of Crystal Growth*, 27:118–125, 1974.

[78] Alberto Tibaldi, Jesus Alberto Gonzalez Montoya, Matteo G. C. Alasio, Alberto Gullino, Anders Larsson, Pierluigi Debernardi, Michele Goano, Marco Vallone, Giovanni Ghione, Enrico Bellotti, and Francesco Bertazzi. Analysis of carrier transport in tunnel-junction vertical-cavity surface-emitting lasers by a coupled nonequilibrium Green's function–drift-diffusion approach. *Phys. Rev. Appl.*, 14(2):024037, August 2020.

[79] P. Rinke, M. Winkelnkemper, A. Qteish, D. Bimberg, J. Neugebauer, and M. Scheffler. Consistent set of band parameters for the group-III nitrides AlN, GaN, and InN. *Phys. Rev. B*, 77(7):075202, 2008.

Appendix A

Wurtzite eight-band k · p Hamiltonian

The multiband calculations performed for the III-nitride structures employ an eight-band $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian for (0001)-oriented wurtzite semiconductors. The model explicitly includes the lowest conduction band and the three upper valence bands, together with their spin components. The formulation follows Refs. [19, 28, 22].

The Hamiltonian is expressed in the zone-centre basis

$$
\begin{aligned}
|1\rangle &= |iS\uparrow\rangle, & |2\rangle &= |iS\downarrow\rangle, \\
|3\rangle &= -\frac{1}{\sqrt{2}}|(X+iY)\uparrow\rangle, & |4\rangle &= \frac{1}{\sqrt{2}}|(X-iY)\uparrow\rangle, \\
|5\rangle &= |Z\uparrow\rangle, & |6\rangle &= \frac{1}{\sqrt{2}}|(X-iY)\downarrow\rangle, \\
|7\rangle &= -\frac{1}{\sqrt{2}}|(X+iY)\downarrow\rangle, & |8\rangle &= |Z\downarrow\rangle.
\end{aligned} \tag{A.1}
$$

The first two basis functions are predominantly s-like conduction-band states, while the remaining six are predominantly p-like valence-band states. In this basis, the complete bulk Hamiltonian is

$$H_{\mathbf{k}\cdot\mathbf{p}} = \begin{pmatrix} E_c & 0 & -\frac{k_+P_2}{\sqrt{2}} & \frac{k_-P_2}{\sqrt{2}} & k_zP_1 & 0 & 0 & 0 \\ 0 & E_c & 0 & 0 & 0 & \frac{k_-P_2}{\sqrt{2}} & -\frac{k_+P_2}{\sqrt{2}} & k_zP_1 \\ -\frac{P_2k_-}{\sqrt{2}} & 0 & F & -\left(K^\dagger\right)^* & -\left(H_+^\dagger\right)^* & 0 & 0 & 0 \\ \frac{P_2k_+}{\sqrt{2}} & 0 & -K^* & G & H_-^\dagger & 0 & 0 & \sqrt{2}\Delta_3 \\ P_1k_z & 0 & -H_+^* & H_- & \Lambda & 0 & \sqrt{2}\Delta_3 & 0 \\ 0 & \frac{P_2k_+}{\sqrt{2}} & 0 & 0 & 0 & F^* & -K^* & H_-^\dagger \\ 0 & -\frac{P_2k_-}{\sqrt{2}} & 0 & 0 & \sqrt{2}\Delta_3 & -\left(K^\dagger\right)^* & G^* & -\left(H_+^\dagger\right)^* \\ 0 & P_1k_z & 0 & \sqrt{2}\Delta_3 & 0 & H_- & -H_+^* & \Lambda \end{pmatrix}. \tag{A.2}$$

The transverse wave-vector combinations are defined as

$$k_\pm = k_x \pm ik_y. \tag{A.3}$$

The conduction-band contribution is

$$E_c = E_g + \Delta_1 + \Delta_2 + \frac{\hbar^2}{2m_0}\left[k_xA_tk_x + k_yA_tk_y + k_zA_zk_z\right], \tag{A.4}$$

where m_0 is the free-electron mass and A_t and A_z determine the transverse and longitudinal conduction-band dispersions, respectively.

The valence-band contributions are written in terms of

$$\Lambda = \frac{\hbar^2}{2m_0}\left[k_zA_1k_z + k_xA_2k_x + k_yA_2k_y\right], \tag{A.5}$$

and

$$\Theta = \frac{\hbar^2}{2m_0}\left[k_zA_3k_z + k_xA_4k_x + k_yA_4k_y\right]. \tag{A.6}$$

The diagonal terms F and G are

$$F = \Delta_1 + \Delta_2 + \Lambda + \Theta + \frac{\hbar^2}{2m_0}\left[-ik_y\left(A_5^+ - A_5^-\right)k_x + ik_x\left(A_5^+ - A_5^-\right)k_y\right], \tag{A.7}$$

and

$$G = \Delta_1 - \Delta_2 + \Lambda + \Theta + \frac{\hbar^2}{2m_0}\left[-ik_x\left(A_5^+ - A_5^-\right)k_y + ik_y\left(A_5^+ - A_5^-\right)k_x\right]. \tag{A.8}$$

The remaining valence-band coupling terms are

$$K = \frac{\hbar^2}{2m_0}\left[k_x A_5 k_x - i\left(k_x A_5 k_y + k_y A_5 k_x\right) - k_y A_5 k_y\right], \tag{A.9}$$

and

$$H_\pm = \frac{\hbar^2}{2m_0}\left[k_z A_6^+\left(k_x - ik_y\right) + \left(k_x - ik_y\right)A_6^- k_z\right] \pm iA_7\left(k_x - ik_y\right). \tag{A.10}$$

The ordered coefficients satisfy

$$A_5 = A_5^+ + A_5^-, \qquad A_6 = A_6^+ + A_6^-. \tag{A.11}$$

Their separation identifies the left- and right-acting contributions when the material parameters vary along the heterostructure. In the numerical implementation, these contributions are assembled directly through the matrices H_{1L} and H_{1R}, and the ordered coefficients are therefore not treated as additional independent material parameters.

The crystal-field and spin–orbit splittings are related to the quantities appearing in Eq. (A.2) through

$$\Delta_1 = \Delta_{\mathrm{cr}}, \qquad \Delta_2 = \Delta_3 = \frac{\Delta_{\mathrm{SO}}}{3}. \tag{A.12}$$

The quantities P_1, P_2, A_t, and A_z are not independent input parameters. They are determined from the band gap, the crystal-field and spin–orbit splittings, and the longitudinal and transverse conduction-band effective masses, following the parametrization adopted in the numerical implementation.

For compactness, the constant

$$C_0 = \frac{\hbar^2}{2m_0} \tag{A.13}$$

is introduced. The normalized electron effective masses are denoted by

$$\mu_z = \frac{m_e^z}{m_0}, \qquad \mu_t = \frac{m_e^t}{m_0}. \tag{A.14}$$

The offset between the reference p-like valence states and the highest valence-band state is

$$\Delta_v = \max\left[\Delta_1 + \Delta_2,\ \frac{\Delta_1 - \Delta_2}{2} + \sqrt{\left(\frac{\Delta_1 - \Delta_2}{2}\right)^2 + 2\Delta_3^2}\right]. \tag{A.15}$$

The corresponding energy separation between the conduction state and this reference is

$$E_c^{(0)} = E_g + \Delta_v. \tag{A.16}$$

Defining

$$\mathcal{D} = \left(E_g + \Delta_1 + \Delta_2\right)\left(E_g + 2\Delta_2\right) - 2\Delta_3^2, \tag{A.17}$$

the longitudinal conduction–valence coupling is obtained as

$$P_1^2 = C_0\left(\frac{1}{\mu_z} - 1\right)\frac{\mathcal{D}}{E_g + 2\Delta_2}, \tag{A.18}$$

while the transverse coupling is

$$P_2^2 = C_0 \left(\frac{1}{\mu_t} - 1 \right) E_g \frac{\mathcal{D}}{\left(E_g + \Delta_1 + \Delta_2\right)\left(E_g + \Delta_2\right) - \Delta_3^2}. \tag{A.19}$$

The positive square roots are used,

$$P_1 = \sqrt{P_1^2}, \qquad P_2 = \sqrt{P_2^2}. \tag{A.20}$$

The residual conduction-band coefficients entering Eq. (A.4) are

$$A_t = \frac{1}{\mu_t} - \frac{P_2^2}{C_0 E_g}, \tag{A.21}$$

and

$$A_z = \frac{1}{\mu_z} - \frac{P_1^2}{C_0 E_c^{(0)}}. \tag{A.22}$$

The parameters $A_1, \ldots, A_6$ provided by the material database contain the effect of the conduction band through the original band-structure fit. Since the conduction band is included explicitly in the eight-band model, its contribution must be removed to avoid double counting. The parameters entering the Hamiltonian are therefore obtained through the Löwdin renormalization

$$\begin{aligned}
\widetilde{A}_1 &= A_1 + \frac{P_1^2}{C_0 E_c^{(0)}}, & \widetilde{A}_3 &= A_3 - \frac{P_1^2}{C_0 E_c^{(0)}}, \\
\widetilde{A}_4 &= A_4 + \frac{P_2^2}{2 C_0 E_g}, & \widetilde{A}_5 &= A_5 + \frac{P_2^2}{2 C_0 E_g}, \\
\widetilde{A}_6 &= A_6 + \frac{P_1 P_2}{C_0 \sqrt{2 E_g E_c^{(0)}}}.
\end{aligned} \tag{A.23}$$

The parameters A_2 and A_7 are left unchanged. In Eqs. (A.5)–(A.10), the symbols A_1, A_3, A_4, A_5, and A_6 are understood to represent these renormalized values.

The independent material parameters used in the calculations are reported in Table A.1. The values are taken from Ref. [79].

Table A.1 Independent band parameters used for the wurtzite eight-band $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian.

Parameter	Unit	GaN	AlN	InN
E_g	eV	3.44	6.00	0.69
m_e^z/m_0	–	0.186	0.322	0.065
m_e^t/m_0	–	0.209	0.329	0.068
A_1	–	−5.947	−3.991	−15.803
A_2	–	−0.528	−0.311	−0.497
A_3	–	5.414	3.671	15.251
A_4	–	−2.512	−1.147	−7.151
A_5	–	−2.510	−1.329	−7.060
A_6	–	−3.202	−1.952	−10.078
A_7	eV Å	0.046	0.026	0.175
Δ_{cr}	meV	10	−295	66
Δ_{SO}	meV	17	19	5

For ternary alloys, the parameters are obtained by interpolation between the corresponding binary compounds. For an $\mathrm{Al}_x\mathrm{In}_y\mathrm{Ga}_{1-x-y}\mathrm{N}$ alloy, a generic parameter Q is evaluated as

$$Q(x, y) = (1 - x - y)Q_{\mathrm{GaN}} + xQ_{\mathrm{AlN}} + yQ_{\mathrm{InN}}. \tag{A.24}$$

The alloy band gap additionally includes the adopted bowing contributions. Once the independent alloy parameters have been determined, the derived quantities P_1, P_2, A_t, A_z, and the renormalized A_i are recalculated locally using Eqs. (A.15)–(A.23).

Appendix B

Recursive Green's function method

The recursive Green's function method provides an efficient solution of the Dyson and Keldysh equations when the discretized Hamiltonian and self-energies have a block-tridiagonal structure [29]. In the following derivation, the equations are first written using the complete block partitions $1{:}q$ and $q+1{:}N$. The block-tridiagonal structure is introduced only afterward, when the resulting matrix equations are reduced to recursive relations involving neighbouring spatial blocks.

The retarded Green's function satisfies

$$A^R G^R = \mathbf{I}, \qquad A^R = (E + i\eta)\mathbf{I} - H - \Sigma^R, \tag{B.1}$$

while

$$A^A = (A^R)^\dagger \tag{B.2}$$

is the corresponding advanced inverse Green's-function matrix. The energy dependence of all quantities is omitted in the following for compactness. All matrix indices refer to spatial blocks rather than individual scalar matrix elements.

B.1 Recursive formulation of the retarded Green's function

B.1.1 Forward recursion for the left-connected Green's function

The left-connected Green's function $g^{rL,(q)}$ is defined as the Green's function of the truncated subsystem containing the first q spatial blocks:

$$A^R_{1:q,1:q} g^{rL,(q)}_{1:q,1:q} = \mathbf{I}_q. \tag{B.3}$$

The superscript (q) indicates the number of blocks included in the truncated subsystem and is not a matrix index.

To determine the left-connected Green's function of the subsystem containing the first $q+1$ blocks, the Dyson equation is partitioned as

$$\begin{pmatrix} A^R_{1:q,1:q} & A^R_{1:q,q+1} \\ A^R_{q+1,1:q} & A^R_{q+1,q+1} \end{pmatrix} \begin{pmatrix} g^{rL,(q+1)}_{1:q,1:q} & g^{rL,(q+1)}_{1:q,q+1} \\ g^{rL,(q+1)}_{q+1,1:q} & g^{rL,(q+1)}_{q+1,q+1} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_q & 0 \\ 0 & \mathbf{I} \end{pmatrix}. \tag{B.4}$$

The corresponding block equations are

$$A^R_{1:q,1:q} g^{rL,(q+1)}_{1:q,1:q} + A^R_{1:q,q+1} g^{rL,(q+1)}_{q+1,1:q} = \mathbf{I}_q, \tag{B.5a}$$
$$A^R_{1:q,1:q} g^{rL,(q+1)}_{1:q,q+1} + A^R_{1:q,q+1} g^{rL,(q+1)}_{q+1,q+1} = 0, \tag{B.5b}$$
$$A^R_{q+1,1:q} g^{rL,(q+1)}_{1:q,1:q} + A^R_{q+1,q+1} g^{rL,(q+1)}_{q+1,1:q} = 0, \tag{B.5c}$$
$$A^R_{q+1,1:q} g^{rL,(q+1)}_{1:q,q+1} + A^R_{q+1,q+1} g^{rL,(q+1)}_{q+1,q+1} = \mathbf{I}. \tag{B.5d}$$

From the second equation,

$$g^{rL,(q+1)}_{1:q,q+1} = -\left(A^R_{1:q,1:q}\right)^{-1} A^R_{1:q,q+1} g^{rL,(q+1)}_{q+1,q+1}. \tag{B.6}$$

Using the definition in Eq. (B.3),

$$g^{rL,(q+1)}_{1:q,q+1} = -g^{rL,(q)}_{1:q,1:q} A^R_{1:q,q+1} g^{rL,(q+1)}_{q+1,q+1}. \tag{B.7}$$

Substitution into the fourth equation of Eqs. (B.5) gives

$$g^{rL,(q+1)}_{q+1,q+1} = \Big[A^R_{q+1,q+1} - A^R_{q+1,1:q} g^{rL,(q)}_{1:q,1:q} A^R_{1:q,q+1}\Big]^{-1}. \tag{B.8}$$

At this point the block-tridiagonal structure can be exploited. The matrices $A^R_{q+1,1:q}$ and $A^R_{1:q,q+1}$ contain only the neighbouring blocks $A^R_{q+1,q}$ and $A^R_{q,q+1}$, respectively. Equation (B.8) therefore reduces to

$$\boxed{g^{rL,(q+1)}_{q+1,q+1} = \Big[A^R_{q+1,q+1} - A^R_{q+1,q} g^{rL,(q)}_{q,q} A^R_{q,q+1}\Big]^{-1}} \tag{B.9}$$

with initial condition

$$g^{rL,(1)}_{1,1} = \left(A^R_{1,1}\right)^{-1}. \tag{B.10}$$

Equation (B.9) can be applied recursively from the left boundary up to the final block. Since the left-connected subsystem then coincides with the complete device,

$$G^R_{N,N} = g^{rL,(N)}_{N,N}. \tag{B.11}$$

The remaining elements in the last column of the left-connected Green's function follow from Eq. (B.7). After exploiting block tridiagonality,

$$g^{rL,(q+1)}_{1:q,q+1} = -g^{rL,(q)}_{1:q,q} A^R_{q,q+1} g^{rL,(q+1)}_{q+1,q+1}. \tag{B.12}$$

This relation is later used to reconstruct the off-diagonal elements of the full Green's function.

B.1.2 Backward recursion for the diagonal blocks of G^R

Once the forward recursion has been completed, the remaining diagonal blocks of the full Green's function are obtained through a backward sweep. The full Dyson equation is partitioned into the subspaces $1{:}q$ and $q+1{:}N$:

$$\begin{pmatrix} A^R_{1:q,1:q} & A^R_{1:q,q+1:N} \\ A^R_{q+1:N,1:q} & A^R_{q+1:N,q+1:N} \end{pmatrix} \begin{pmatrix} G^R_{1:q,1:q} & G^R_{1:q,q+1:N} \\ G^R_{q+1:N,1:q} & G^R_{q+1:N,q+1:N} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_q & 0 \\ 0 & \mathbf{I}_{N-q} \end{pmatrix}. \tag{B.13}$$

The lower-left block equation gives

$$G^R_{q+1:N,1:q} = -\left(A^R_{q+1:N,q+1:N}\right)^{-1} A^R_{q+1:N,1:q} G^R_{1:q,1:q}. \tag{B.14}$$

Substitution into the upper-left block equation yields

$$\begin{aligned} G^R_{1:q,1:q} = \Big[& A^R_{1:q,1:q} \\ & - A^R_{1:q,q+1:N}\left(A^R_{q+1:N,q+1:N}\right)^{-1} A^R_{q+1:N,1:q}\Big]^{-1}. \end{aligned} \tag{B.15}$$

Using the Woodbury identity,

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1}+VA^{-1}U\right)^{-1} VA^{-1}, \tag{B.16}$$

Eq. (B.15) can be rewritten as

$$\begin{aligned} G^R_{1:q,1:q} = & \left(A^R_{1:q,1:q}\right)^{-1} \\ & + \left(A^R_{1:q,1:q}\right)^{-1} A^R_{1:q,q+1:N} G^R_{q+1:N,q+1:N} \\ & \times A^R_{q+1:N,1:q}\left(A^R_{1:q,1:q}\right)^{-1}. \end{aligned} \tag{B.17}$$

Since

$$\left(A^R_{1:q,1:q}\right)^{-1} = g^{rL,(q)}_{1:q,1:q}, \tag{B.18}$$

one obtains

$$\begin{aligned} G^R_{1:q,1:q} = & g^{rL,(q)}_{1:q,1:q} \\ & + g^{rL,(q)}_{1:q,1:q} A^R_{1:q,q+1:N} G^R_{q+1:N,q+1:N} \\ & \times A^R_{q+1:N,1:q} g^{rL,(q)}_{1:q,1:q}. \end{aligned} \tag{B.19}$$

Because A^R is block tridiagonal, the coupling between the two partitions occurs only through $A^R_{q,q+1}$ and $A^R_{q+1,q}$. Extracting the (q,q) block gives

$$\boxed{G^R_{q,q} = g^{rL,(q)}_{q,q} + g^{rL,(q)}_{q,q} A^R_{q,q+1} G^R_{q+1,q+1} A^R_{q+1,q} g^{rL,(q)}_{q,q}} \tag{B.20}$$

for $q = N-1, \ldots, 1$. Using

$$G^R_{q+1,q} = -G^R_{q+1,q+1} A^R_{q+1,q} g^{rL,(q)}_{q,q}, \tag{B.21}$$

the same result can also be written as

$$G^R_{q,q} = g^{rL,(q)}_{q,q} - g^{rL,(q)}_{q,q} A^R_{q,q+1} G^R_{q+1,q}. \tag{B.22}$$

B.1.3 Reconstruction of the off-diagonal blocks

The upper-right block equation obtained from Eq. (B.13) is

$$G^R_{1:q,q+1:N} = -g^{rL,(q)}_{1:q,1:q} A^R_{1:q,q+1:N} G^R_{q+1:N,q+1:N}. \tag{B.23}$$

Although this expression involves complete block submatrices, block tridiagonality implies that $A^R_{1:q,q+1:N}$ contains only one nonzero block, namely $A^R_{q,q+1}$. The resulting structure can be illustrated by considering explicit examples.

For $q = 2$ and $N = 4$,

$$\begin{aligned} &\begin{pmatrix} G^R_{1,3} & G^R_{1,4} \\ G^R_{2,3} & G^R_{2,4} \end{pmatrix} \\ &= -\begin{pmatrix} g^{rL,(2)}_{1,1} & g^{rL,(2)}_{1,2} \\ g^{rL,(2)}_{2,1} & g^{rL,(2)}_{2,2} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ A^R_{2,3} & 0 \end{pmatrix} \begin{pmatrix} G^R_{3,3} & G^R_{3,4} \\ G^R_{4,3} & G^R_{4,4} \end{pmatrix}. \end{aligned} \tag{B.24}$$

Performing the multiplication gives

$$\begin{pmatrix} G^R_{1,3} & G^R_{1,4} \\ G^R_{2,3} & G^R_{2,4} \end{pmatrix} = -\begin{pmatrix} g^{rL,(2)}_{1,2} A^R_{2,3} G^R_{3,3} & g^{rL,(2)}_{1,2} A^R_{2,3} G^R_{3,4} \\ g^{rL,(2)}_{2,2} A^R_{2,3} G^R_{3,3} & g^{rL,(2)}_{2,2} A^R_{2,3} G^R_{3,4} \end{pmatrix}. \tag{B.25}$$

For $q = 3$ and $N = 6$,

$$\begin{aligned}
&\begin{pmatrix} G^R_{1,4} & G^R_{1,5} & G^R_{1,6} \\ G^R_{2,4} & G^R_{2,5} & G^R_{2,6} \\ G^R_{3,4} & G^R_{3,5} & G^R_{3,6} \end{pmatrix} \\
&= -\begin{pmatrix} g^{rL,(3)}_{1,3} \\ g^{rL,(3)}_{2,3} \\ g^{rL,(3)}_{3,3} \end{pmatrix} A^R_{3,4} \begin{pmatrix} G^R_{4,4} & G^R_{4,5} & G^R_{4,6} \end{pmatrix}.
\end{aligned} \tag{B.26}$$

The two examples show that only the last column of the left-connected Green's function and the first row of the right subspace contribute. Equation (B.23) can therefore be reduced to

$$\boxed{G^R_{1:q,q+1:N} = -g^{rL,(q)}_{1:q,q} A^R_{q,q+1} G^R_{q+1,q+1:N}} \tag{B.27}$$

and, for an individual element,

$$G^R_{i,q+1} = -g^{rL,(q)}_{i,q} A^R_{q,q+1} G^R_{q+1,q+1}, \qquad i = 1, \ldots, q. \tag{B.28}$$

Equivalently, the element on the ith upper diagonal can be written as

$$G^R_{q+1-i,q+1} = -g^{rL,(q)}_{q+1-i,q} A^R_{q,q+1} G^R_{q+1,q+1}. \tag{B.29}$$

The corresponding lower-left block is

$$\boxed{G^R_{q+1:N,1:q} = -G^R_{q+1:N,q+1} A^R_{q+1,q} g^{rL,(q)}_{q,1:q}} \tag{B.30}$$

and therefore

$$G^R_{q+1,i} = -G^R_{q+1,q+1} A^R_{q+1,q} g^{rL,(q)}_{q,i}, \qquad i = 1, \ldots, q. \tag{B.31}$$

The required columns of $g^{rL,(q)}$ are generated during the forward sweep using Eq. (B.12). Equations (B.27) and (B.30) can then be applied during the backward sweep to reconstruct the off-diagonal blocks of the complete retarded Green's function. The resulting recursion was verified numerically by comparison with direct inversion of block-tridiagonal matrices and was found to reproduce the complete retarded Green's function.

B.2 Recursive formulation of the lesser Green's function

The lesser Green's function satisfies the Keldysh equation

$$G^{<} = G^{R}\Sigma^{<}G^{A}. \tag{B.32}$$

B.2.1 Forward recursion for the left-connected lesser Green's function

The left-connected lesser Green's function of the subsystem containing the first q blocks is defined as

$$g^{<L,(q)}_{1:q,1:q} = g^{rL,(q)}_{1:q,1:q}\Sigma^{<}_{1:q,1:q}g^{aL,(q)}_{1:q,1:q}, \tag{B.33}$$

where

$$g^{aL,(q)} = \left(g^{rL,(q)}\right)^{\dagger}. \tag{B.34}$$

For the subsystem containing $q+1$ blocks, the Keldysh equation is partitioned as

$$\begin{aligned}
&\begin{pmatrix} g^{<L,(q+1)}_{1:q,1:q} & g^{<L,(q+1)}_{1:q,q+1} \\ g^{<L,(q+1)}_{q+1,1:q} & g^{<L,(q+1)}_{q+1,q+1} \end{pmatrix} \\
&= \begin{pmatrix} g^{rL,(q+1)}_{1:q,1:q} & g^{rL,(q+1)}_{1:q,q+1} \\ g^{rL,(q+1)}_{q+1,1:q} & g^{rL,(q+1)}_{q+1,q+1} \end{pmatrix} \begin{pmatrix} \Sigma^{<}_{1:q,1:q} & \Sigma^{<}_{1:q,q+1} \\ \Sigma^{<}_{q+1,1:q} & \Sigma^{<}_{q+1,q+1} \end{pmatrix} \\
&\quad \times \begin{pmatrix} g^{aL,(q+1)}_{1:q,1:q} & g^{aL,(q+1)}_{1:q,q+1} \\ g^{aL,(q+1)}_{q+1,1:q} & g^{aL,(q+1)}_{q+1,q+1} \end{pmatrix}.
\end{aligned} \tag{B.35}$$

The lower-right block is

$$\begin{aligned}
g^{<L,(q+1)}_{q+1,q+1} = {} & g^{rL,(q+1)}_{q+1,q+1}\Sigma^{<}_{q+1,q+1}g^{aL,(q+1)}_{q+1,q+1} \\
& + g^{rL,(q+1)}_{q+1,q+1}\Sigma^{<}_{q+1,1:q}g^{aL,(q+1)}_{1:q,q+1} \\
& + g^{rL,(q+1)}_{q+1,1:q}\Sigma^{<}_{1:q,q+1}g^{aL,(q+1)}_{q+1,q+1} \\
& + g^{rL,(q+1)}_{q+1,1:q}\Sigma^{<}_{1:q,1:q}g^{aL,(q+1)}_{1:q,q+1}.
\end{aligned} \tag{B.36}$$

From the partitioned Dyson equation,

$$g^{rL,(q+1)}_{q+1,1:q} = -g^{rL,(q+1)}_{q+1,q+1} A^{R}_{q+1,1:q} g^{rL,(q)}_{1:q,1:q}, \tag{B.37a}$$

$$g^{aL,(q+1)}_{1:q,q+1} = -g^{aL,(q)}_{1:q,1:q} A^{A}_{1:q,q+1} g^{aL,(q+1)}_{q+1,q+1}. \tag{B.37b}$$

Substituting Eqs. (B.37) into Eq. (B.36) and factoring the external retarded and advanced Green's functions gives

$$\begin{aligned} g^{<L,(q+1)}_{q+1,q+1} = g^{rL,(q+1)}_{q+1,q+1} \Big[& \Sigma^{<}_{q+1,q+1} \\ & - \Sigma^{<}_{q+1,1:q} g^{aL,(q)}_{1:q,1:q} A^{A}_{1:q,q+1} \\ & - A^{R}_{q+1,1:q} g^{rL,(q)}_{1:q,1:q} \Sigma^{<}_{1:q,q+1} \\ & + A^{R}_{q+1,1:q} g^{<L,(q)}_{1:q,1:q} A^{A}_{1:q,q+1} \Big] g^{aL,(q+1)}_{q+1,q+1}. \end{aligned} \tag{B.38}$$

Using the block-tridiagonal structure of A^R and $\Sigma^<$, the recursion reduces to

$$\boxed{\begin{aligned} g^{<L,(q+1)}_{q+1,q+1} = g^{rL,(q+1)}_{q+1,q+1} \Big[& \Sigma^{<}_{q+1,q+1} \\ & - \Sigma^{<}_{q+1,q} g^{aL,(q)}_{q,q} A^{A}_{q,q+1} \\ & - A^{R}_{q+1,q} g^{rL,(q)}_{q,q} \Sigma^{<}_{q,q+1} \\ & + A^{R}_{q+1,q} g^{<L,(q)}_{q,q} A^{A}_{q,q+1} \Big] g^{aL,(q+1)}_{q+1,q+1} \end{aligned}} \tag{B.39}$$

The initial condition is

$$g^{<L,(1)}_{1,1} = g^{rL,(1)}_{1,1} \Sigma^{<}_{1,1} g^{aL,(1)}_{1,1}. \tag{B.40}$$

At the final block,

$$G^{<}_{N,N} = g^{<L,(N)}_{N,N}. \tag{B.41}$$

B.2.2 Backward recursion for the diagonal blocks of $G^<$

The full Keldysh equation is partitioned into the subspaces $1{:}q$ and $q+1{:}N$. Its upper-left block is

$$\begin{aligned} G^<_{1:q,1:q} = {} & G^R_{1:q,1:q}\Sigma^<_{1:q,1:q}G^A_{1:q,1:q} \\ & + G^R_{1:q,1:q}\Sigma^<_{1:q,q+1:N}G^A_{q+1:N,1:q} \\ & + G^R_{1:q,q+1:N}\Sigma^<_{q+1:N,1:q}G^A_{1:q,1:q} \\ & + G^R_{1:q,q+1:N}\Sigma^<_{q+1:N,q+1:N}G^A_{q+1:N,1:q}. \end{aligned} \tag{B.42}$$

The lower-right block is

$$\begin{aligned} G^<_{q+1:N,q+1:N} = {} & G^R_{q+1:N,1:q}\Sigma^<_{1:q,1:q}G^A_{1:q,q+1:N} \\ & + G^R_{q+1:N,1:q}\Sigma^<_{1:q,q+1:N}G^A_{q+1:N,q+1:N} \\ & + G^R_{q+1:N,q+1:N}\Sigma^<_{q+1:N,1:q}G^A_{1:q,q+1:N} \\ & + G^R_{q+1:N,q+1:N}\Sigma^<_{q+1:N,q+1:N}G^A_{q+1:N,q+1:N}. \end{aligned} \tag{B.43}$$

The retarded and advanced block relations derived previously are

$$G^R_{1:q,q+1:N} = -g^{rL,(q)}_{1:q,1:q}A^R_{1:q,q+1:N}G^R_{q+1:N,q+1:N}, \tag{B.44a}$$

$$G^R_{q+1:N,1:q} = -G^R_{q+1:N,q+1:N}A^R_{q+1:N,1:q}g^{rL,(q)}_{1:q,1:q}, \tag{B.44b}$$

$$G^A_{1:q,q+1:N} = -g^{aL,(q)}_{1:q,1:q}A^A_{1:q,q+1:N}G^A_{q+1:N,q+1:N}, \tag{B.44c}$$

$$G^A_{q+1:N,1:q} = -G^A_{q+1:N,q+1:N}A^A_{q+1:N,1:q}g^{aL,(q)}_{1:q,1:q}. \tag{B.44d}$$

To make the reduction explicit, the last term of Eq. (B.42) is treated by rearranging the lower-right Keldysh block in Eq. (B.43) as

$$\begin{aligned} & G^R_{q+1:N,q+1:N}\Sigma^<_{q+1:N,q+1:N}G^A_{q+1:N,q+1:N} \\ & = G^<_{q+1:N,q+1:N} \\ & \quad - G^R_{q+1:N,1:q}\Sigma^<_{1:q,1:q}G^A_{1:q,q+1:N} \\ & \quad - G^R_{q+1:N,1:q}\Sigma^<_{1:q,q+1:N}G^A_{q+1:N,q+1:N} \\ & \quad - G^R_{q+1:N,q+1:N}\Sigma^<_{q+1:N,1:q}G^A_{1:q,q+1:N}. \end{aligned} \tag{B.45}$$

After substituting this relation and Eqs. (B.44) into Eq. (B.42), the terms containing $\Sigma^{<}_{1:q,1:q}$ combine into $g^{<L,(q)}_{1:q,1:q}$ through Eq. (B.33), while the remaining repeated mixed terms cancel. The resulting expression is:

$$\begin{aligned} G^{<}_{1:q,1:q} = {} & g^{<L,(q)}_{1:q,1:q} \\ & + g^{rL,(q)}_{1:q,1:q} A^{R}_{1:q,q+1:N} G^{<}_{q+1:N,q+1:N} A^{A}_{q+1:N,1:q} g^{aL,(q)}_{1:q,1:q} \\ & - G^{R}_{1:q,q+1:N} A^{R}_{q+1:N,1:q} g^{<L,(q)}_{1:q,1:q} \\ & - g^{<L,(q)}_{1:q,1:q} A^{A}_{1:q,q+1:N} G^{A}_{q+1:N,1:q} \\ & + g^{rL,(q)}_{1:q,1:q} \Sigma^{<}_{1:q,q+1:N} G^{A}_{q+1:N,1:q} \\ & + G^{R}_{1:q,q+1:N} \Sigma^{<}_{q+1:N,1:q} g^{aL,(q)}_{1:q,1:q}. \end{aligned} \tag{B.46}$$

This is the general block expression prior to using block tridiagonality. Extracting the (q, q) block gives the backward recursion

$$\boxed{\begin{aligned} G^{<}_{q,q} = {} & g^{<L,(q)}_{q,q} + g^{rL,(q)}_{q,q} A^{R}_{q,q+1} G^{<}_{q+1,q+1} A^{A}_{q+1,q} g^{aL,(q)}_{q,q} \\ & - G^{R}_{q,q+1} A^{R}_{q+1,q} g^{<L,(q)}_{q,q} - g^{<L,(q)}_{q,q} A^{A}_{q,q+1} G^{A}_{q+1,q} \\ & + g^{rL,(q)}_{q,q} \Sigma^{<}_{q,q+1} G^{A}_{q+1,q} + G^{R}_{q,q+1} \Sigma^{<}_{q+1,q} g^{aL,(q)}_{q,q} \end{aligned}} \tag{B.47}$$

Starting from Eq. (B.41), Eq. (B.47) provides the full lesser Green's-function diagonal through a backward sweep.

B.2.3 Reconstruction of the off-diagonal blocks of $G^{<}$

The recursive procedure can also be used to reconstruct the off-diagonal blocks of the lesser Green's function. During the forward sweep, these blocks are generated when block $q + 1$ is added to the left-connected subsystem containing the first q blocks.

More specifically, the upper-right block of the partitioned Keldysh equation in Eq. (B.35) is combined with the retarded and advanced cross-block relations in Eqs. (B.37). The lower-right block derived in the previous subsection is then used to collect the terms associated with the

newly added block. This gives

$$\begin{aligned} g^{<L,(q+1)}_{1:q,q+1} = & - g^{<L,(q)}_{1:q,q} A^{A}_{q,q+1} g^{aL,(q+1)}_{q+1,q+1} \\ & - g^{rL,(q)}_{1:q,q} A^{R}_{q,q+1} g^{<L,(q+1)}_{q+1,q+1} \\ & + g^{rL,(q)}_{1:q,q} \Sigma^{<}_{q,q+1} g^{aL,(q+1)}_{q+1,q+1} . \end{aligned} \tag{B.48}$$

The corresponding lower-left block row is obtained analogously from the lower-left block of the partitioned Keldysh equation:

$$\begin{aligned} g^{<L,(q+1)}_{q+1,1:q} = & - g^{rL,(q+1)}_{q+1,q+1} A^{R}_{q+1,q} g^{<L,(q)}_{q,1:q} \\ & - g^{<L,(q+1)}_{q+1,q+1} A^{A}_{q+1,q} g^{aL,(q)}_{q,1:q} \\ & + g^{rL,(q+1)}_{q+1,q+1} \Sigma^{<}_{q+1,q} g^{aL,(q)}_{q,1:q} . \end{aligned} \tag{B.49}$$

Consequently, for any $i = 1, \ldots, q$, the newly generated upper off-diagonal blocks of the left-connected lesser Green's function satisfy

$$\boxed{\begin{aligned} g^{<L,(q+1)}_{i,q+1} = & - g^{<L,(q)}_{i,q} A^{A}_{q,q+1} g^{aL,(q+1)}_{q+1,q+1} \\ & - g^{rL,(q)}_{i,q} A^{R}_{q,q+1} g^{<L,(q+1)}_{q+1,q+1} \\ & + g^{rL,(q)}_{i,q} \Sigma^{<}_{q,q+1} g^{aL,(q+1)}_{q+1,q+1} \end{aligned}} \tag{B.50}$$

while the corresponding lower off-diagonal blocks are

$$\boxed{\begin{aligned} g^{<L,(q+1)}_{q+1,i} = & - g^{rL,(q+1)}_{q+1,q+1} A^{R}_{q+1,q} g^{<L,(q)}_{q,i} \\ & - g^{<L,(q+1)}_{q+1,q+1} A^{A}_{q+1,q} g^{aL,(q)}_{q,i} \\ & + g^{rL,(q+1)}_{q+1,q+1} \Sigma^{<}_{q+1,q} g^{aL,(q)}_{q,i} \end{aligned}} \tag{B.51}$$

for $i = 1, \ldots, q$. Repeating these relations during the forward sweep generates the required off-diagonal blocks of each progressively enlarged left-connected subsystem.

Once the forward sweep has been completed, the same block reduction is applied to the complete, rather than left-connected, Keldysh equation. During the backward sweep, the

upper-right block column of the full lesser Green's function satisfies

$$\begin{aligned} G^{<}_{1:q,q+1} = & - g^{<L,(q)}_{1:q,q} A^{A}_{q,q+1} G^{A}_{q+1,q+1} \\ & - g^{rL,(q)}_{1:q,q} A^{R}_{q,q+1} G^{<}_{q+1,q+1} \\ & + g^{rL,(q)}_{1:q,q} \Sigma^{<}_{q,q+1} G^{A}_{q+1,q+1}. \end{aligned} \tag{B.52}$$

The corresponding lower-left block row is

$$\begin{aligned} G^{<}_{q+1,1:q} = & - G^{R}_{q+1,q+1} A^{R}_{q+1,q} g^{<L,(q)}_{q,1:q} \\ & - G^{<}_{q+1,q+1} A^{A}_{q+1,q} g^{aL,(q)}_{q,1:q} \\ & + G^{R}_{q+1,q+1} \Sigma^{<}_{q+1,q} g^{aL,(q)}_{q,1:q}. \end{aligned} \tag{B.53}$$

Therefore, all blocks in column $q+1$ above the main diagonal are reconstructed from

$$\boxed{\begin{aligned} G^{<}_{i,q+1} = & - g^{<L,(q)}_{i,q} A^{A}_{q,q+1} G^{A}_{q+1,q+1} \\ & - g^{rL,(q)}_{i,q} A^{R}_{q,q+1} G^{<}_{q+1,q+1} \\ & + g^{rL,(q)}_{i,q} \Sigma^{<}_{q,q+1} G^{A}_{q+1,q+1} \end{aligned}} \tag{B.54}$$

for $i = 1, \ldots, q$. Similarly, all blocks in row $q+1$ below the main diagonal are reconstructed from

$$\boxed{\begin{aligned} G^{<}_{q+1,i} = & - G^{R}_{q+1,q+1} A^{R}_{q+1,q} g^{<L,(q)}_{q,i} \\ & - G^{<}_{q+1,q+1} A^{A}_{q+1,q} g^{aL,(q)}_{q,i} \\ & + G^{R}_{q+1,q+1} \Sigma^{<}_{q+1,q} g^{aL,(q)}_{q,i} \end{aligned}} \tag{B.55}$$

for $i = 1, \ldots, q$.

For $i = q$, Eqs. (B.54) and (B.55) reduce to the first upper and lower off-diagonal blocks, respectively. At each step of the backward sweep, the relations reconstruct the complete block column and block row associated with block $q+1$. Repeating the procedure for $q = N-1, \ldots, 1$ therefore recovers all the off-diagonal blocks of the lesser Green's function.

The lower blocks can equivalently be obtained from the anti-Hermitian property

$$G^{<}_{j,i} = - \left(G^{<}_{i,j} \right)^{\dagger}. \tag{B.56}$$

Appendix C

Recovery of the SRH rate in the semiclassical bulk limit

The multi-phonon capture and emission integrals derived in Section 4.1.1 retain the algebraic structure of the conventional Shockley-Read-Hall formulation. This appendix shows explicitly that the conventional SRH rate is recovered when the spectral carrier distributions are restricted to the bulk conduction and valence bands and the nondegenerate approximation is adopted. All quantities are local in space, but the spatial dependence is omitted to simplify the notation.
In a homogeneous semiconductor, the semiclassical conduction-band electron density is

$$n_{B,\mathrm{cl}} = \int_{E_C}^{+\infty} n_c(E)\, dE = N_C \exp\!\left(\frac{F_n - E_C}{k_B T}\right), \tag{C.1}$$

where F_n is the electron quasi-Fermi level. In the Boltzmann limit, the occupied conduction-band spectral density can be written as

$$\frac{n_c(E)}{n_{B,\mathrm{cl}}} = \frac{\mathcal{D}_c(E)}{N_C} \exp\!\left(\frac{E_C - E}{k_B T}\right), \tag{C.2}$$

where $\mathcal{D}_c(E)$ is the conduction-band density of states. The average electron capture coefficient is therefore

$$\langle c_n \rangle = \frac{1}{n_{B,\mathrm{cl}}} \int_{E_C}^{+\infty} c_n(E) n_c(E)\, dE. \tag{C.3}$$

Defining the bulk electron capture lifetime as

$$\frac{1}{\tau_{n,\mathrm{SRH}}} = N_t \langle c_n \rangle, \tag{C.4}$$

the electron capture contribution in Eq. (4.14) becomes

$$I_n^c = \int_{E_C}^{+\infty} N_t c_n(E) n_c(E)\, dE = \frac{n_{B,\mathrm{cl}}}{\tau_{n,\mathrm{SRH}}}. \tag{C.5}$$

The corresponding electron emission contribution is

$$I_n^e = \int_{E_C}^{+\infty} N_t c_n(E) \exp\left(\frac{\varepsilon_d - E}{k_B T}\right) \overline{n}_c(E)\, dE. \tag{C.6}$$

In the nondegenerate limit, the conduction-band states are almost empty and $\overline{n}_c(E) \simeq \mathcal{D}_c(E)$. Introducing the usual SRH auxiliary density

$$n_1 = N_C \exp\left(\frac{\varepsilon_d - E_C}{k_B T}\right), \tag{C.7}$$

one obtains

$$\exp\left(\frac{\varepsilon_d - E}{k_B T}\right) \mathcal{D}_c(E) = \frac{n_1}{n_{B,\mathrm{cl}}} n_c(E). \tag{C.8}$$

Using Eqs. (C.3) and (C.4), Eq. (C.6) then reduces to

$$I_n^e = n_1 N_t \langle c_n \rangle = \frac{n_1}{\tau_{n,\mathrm{SRH}}}. \tag{C.9}$$

The same procedure can be applied to holes. The semiclassical valence-band hole density is

$$p_{B,\mathrm{cl}} = \int_{-\infty}^{E_V} p_v(E)\, dE = N_V \exp\left(\frac{E_V - F_p}{k_B T}\right), \tag{C.10}$$

where F_p is the hole quasi-Fermi level. Its normalized spectral distribution is

$$\frac{p_v(E)}{p_{B,\mathrm{cl}}} = \frac{\mathcal{D}_v(E)}{N_V} \exp\left(\frac{E - E_V}{k_B T}\right), \tag{C.11}$$

where $\mathcal{D}_v(E)$ is the valence-band density of states. The average hole capture coefficient and the corresponding bulk lifetime are defined as

$$\langle c_p \rangle = \frac{1}{p_{B,\mathrm{cl}}} \int_{-\infty}^{E_V} c_p(E) p_v(E)\, dE, \qquad \frac{1}{\tau_{p,\mathrm{SRH}}} = N_t \langle c_p \rangle. \tag{C.12}$$

The hole capture contribution is consequently

$$I_p^c = \int_{-\infty}^{E_V} N_t c_p(E) p_v(E)\, dE = \frac{p_{B,\mathrm{cl}}}{\tau_{p,\mathrm{SRH}}}. \tag{C.13}$$

The hole emission contribution is

$$I_p^e = \int_{-\infty}^{E_V} N_t c_p(E) \exp\left(\frac{E-\varepsilon_d}{k_B T}\right) \overline{p}_v(E)\, dE. \tag{C.14}$$

In the nondegenerate limit, the valence-band states are almost completely occupied and $\overline{p}_v(E) \simeq \mathcal{D}_v(E)$. Defining

$$p_1 = N_V \exp\left(\frac{E_V-\varepsilon_d}{k_B T}\right), \tag{C.15}$$

it follows that

$$\exp\left(\frac{E-\varepsilon_d}{k_B T}\right) \mathcal{D}_v(E) = \frac{p_1}{p_{B,\mathrm{cl}}} p_v(E), \tag{C.16}$$

and therefore

$$I_p^e = p_1 N_t \left\langle c_p \right\rangle = \frac{p_1}{\tau_{p,\mathrm{SRH}}}. \tag{C.17}$$

The four semiclassical capture and emission contributions can thus be written as

$$I_n^c = \frac{n_{B,\mathrm{cl}}}{\tau_{n,\mathrm{SRH}}}, \qquad I_n^e = \frac{n_1}{\tau_{n,\mathrm{SRH}}}, \tag{C.18a}$$

$$I_p^c = \frac{p_{B,\mathrm{cl}}}{\tau_{p,\mathrm{SRH}}}, \qquad I_p^e = \frac{p_1}{\tau_{p,\mathrm{SRH}}}. \tag{C.18b}$$

Substituting Eqs. (C.18) into the general stationary-defect rate in Eq. (4.10) gives

$$U_{\mathrm{SRH}} = \frac{n_{B,\mathrm{cl}} p_{B,\mathrm{cl}} - n_1 p_1}{\tau_{p,\mathrm{SRH}} \left(n_{B,\mathrm{cl}} + n_1\right) + \tau_{n,\mathrm{SRH}} \left(p_{B,\mathrm{cl}} + p_1\right)}. \tag{C.19}$$

Since $n_1 p_1 = n_i^2$, the conventional Shockley-Read-Hall recombination rate is recovered,

$$U_{\mathrm{SRH}} = \frac{n_{B,\mathrm{cl}} p_{B,\mathrm{cl}} - n_i^2}{\tau_{p,\mathrm{SRH}} \left(n_{B,\mathrm{cl}} + n_1\right) + \tau_{n,\mathrm{SRH}} \left(p_{B,\mathrm{cl}} + p_1\right)}. \tag{C.20}$$

The conventional SRH expression is therefore the homogeneous nondegenerate bulk limit of the multi-phonon capture and emission formulation. In Chapter 4, the same capture coefficients are instead averaged over the complete quantum spectral distributions, including virtual tail states

extending into the nominal bandgap, which gives the effective quantum lifetimes used to describe trap-assisted tunneling.