MULTI-SCALE QUANTUM TRANSPORT MODELING OF LIGHT EMITTING DIODES

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Junzhe Geng

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

June 2016

Purdue University

West Lafayette, Indiana

To My Family.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my major advisor, professor Klimeck for his continuous trust and support. I deeply appreciate the opportunity to learn and grow under his guidance over the years.

I would like to thank professor Tillmann Kubis for his guidance during my PhD, teaching me his knowledge, and helping me through various challenges and difficulties.

Special thanks goes to Erik Nelson, for his guidance and support during my PhD project, as well as during my internship at Lumileds.

I would like to thank professor Oana Malis and Michael Manfra for serving as my committee members, and being available for intellectual discussions when I needed.

I would like to thank all my colleagues, especially Prasad Sarangapani, Kuang-Chuang Wang, Yuanchen Chu, James Charles, Bozidar Novakovic, Daniel Mejia, Jim Fonseca,Yaohua Tan, Yu He, Michael Povolotskyi, Jun Huang, Kai Miao, Hesameddin Ilatikhameneh, Tarek Ameen and Saima Sharmin for their various help and support.

I would like extend special thanks to several other (formal) colleagues, besides ones mentioned above, who have also been my personal friends and given me invaluable support both in and outside of workplace: Matthias Tan, Zhengping Jiang, Pengyu Long, Yuling Hsueh, Fan Chen and Xufeng Wang.

Last but not least, none of these would have happened if not for the support of my family. From the deepest of my heart, I would like to thank my parents Yexin Geng and Dengjuan Zhou, my fiancee Natalie Yu-Tung Hou for their selfless, unconditional love and support throughout my life.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

ABSTRACT

Ph.D., Purdue University, June 2016. Multi-Scale Quantum Transport Modeling of Light Emitting Diodes. Major Professor: Gerhard Klimeck and Tillmann Kubis.

GaN/InGaN multi-quantum-well (MQW) structure is the centerpiece of most mid-to-high power light-emitting diodes (LED). The operation of MQW LEDs is determined by the carrier flow through complex, extended quantum states, the optical recombination between these states and the optical fields in the device. Most existing LED modeling tools are based on semiclassical physics with limited capabilities of modeling quantum phenomenon. Non-equilibrium Green Function Formalism (NEGF) is the state-of-the-art approach for quantum transport, however when it is applied in its textbook form it is numerically too demanding to handle realistically extended devices. This work introduces a new approach to LED modeling based on a multi-scaled NEGF approach that subdivides the critical device domains and separates the quantum transport from the recombination treatments. Several key modeling challenges and their solutions are addressed in this work. The model is applied to a commercial blue LED, and comparison between modeling and experimental results shows good promise.

1. INTRODUCTION

1.1 LED and Solid-State-Lighting

Solid-state-lighting is widely regarded as the future of light generation, due to its unparalleled electricity-to-light conversion efficiency compared to traditional lighting. Lighting today accounts for 15-22% of world's electricity consumption [1]. It is estimated with solid-states lighting, that number could be reduced to 4%, saving more than $150B/year in terms of energy and infrastructure cost , as well as reducing carbon emission by 200M tons/year [2]. As the cornerstone of solid-state-lighting, light-emitting diode (LED) has to significantly out-perform all forms of traditional lights in terms of efficiency, output, and cost of ownership, in order for solid-state-lighting to overthrow traditional lighting as the preferred approach. Under a few decades of rapid progress (Fig.1.1), today's LED can reach peak efficiency of 85% peak internal quantum efficiency (IQE) and light extraction efficiency [3]. However, there are several main hurdles that still needs to be overcome:

1. **Cost**: Costing is the main reason why LED has not replace conventional lighting in a wide scale. Figure 1.1 shows the evolution of LED in terms of cost ($ per lumen) and light output per package (lumen per lamp). Every decade, the cost per lumen (unit of useful light emitted) falls by a factor of 10, the amount of light generated per LED package increases by a factor of 20, this is known as Haitz's law [2]. Figure 1.2 shows the amount of energy saving LED could produce over various forms of traditional lighting. In order for LED to significantly penetrate the general lighting market, LED cost needs to be significantly lower than $5 \times 10^{-3}\$/lm$, which is produced by fluorescent light; and efficiency needs to significantly surpass 100 lm/W.

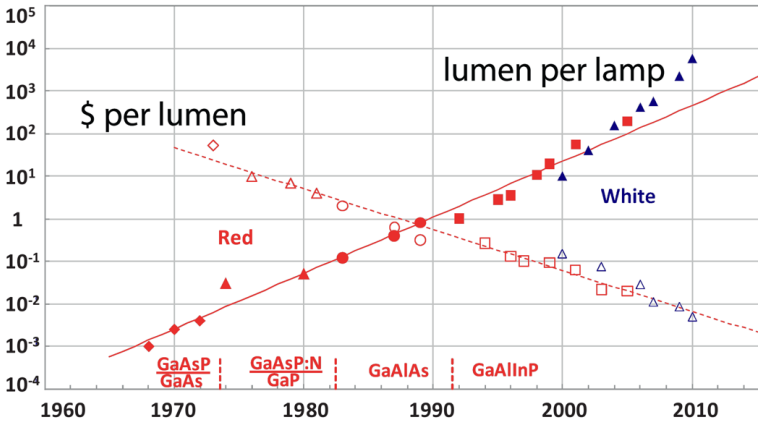

Fig. 1.1.: The progress of LED development over the past few decades in terms of cost ($ per lumen) and output (lumen per lamp). (Reprinted from [4], with the permission of John Wiley and Sons)

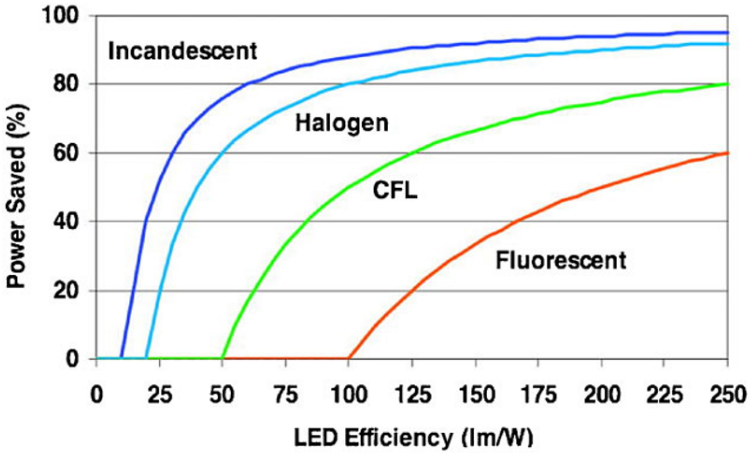

Fig. 1.2.: Energy saving of LED over traditional lighting. (Reprinted from [3], with the permission of John Wiley and Sons)

2. **Efficiency Droop**: The LED can only maintain a high efficiency at low current injection. Efficiency drops significant as the drive current increase. Figure1.3 shows the external quantum efficiency of LED at different drive current, at various operating temperatures. As seen from the figure, the efficiency of LED peaks at a certain current (around 10 mA), and drops significantly at high

current. Efficiency droop at high current leads to higher cost of LED for high power applications. Researches have identified Auger recombination as the main cause for droop. Careful design of LED can suppress, yet not complete overcome the effect of droop, For general lighting applications, LED drive current must be > 350mA. Efficiency droop poses significantly challenge upon LED designing.

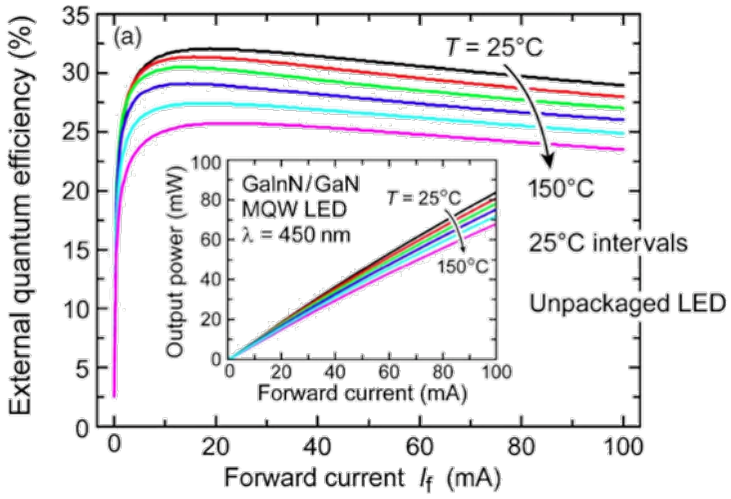

Fig. 1.3.: LED external quantum efficiency at different drive current for different temperature. (Reprinted from [5], with the permission of AIP Publishing)

3. **Green Gap**: The high efficiency only applies to blue LED. For green/yellow LED, efficiency is significantly lower (see Fig. 1.4). Improving the efficiency of green LED, thus reducing the green gap, would open up color-mixing approach and as a result, avoiding ∼25% energy loss due to the Stokes shift in the phosphor conversion process. Several key factors contribute to the existence of "green gap", including degraded material quality at high In%; stronger polarization field at higher In%, leading to stronger quantum-confined Stark effect and less electron-hole overlap; and more severe droop in Green LEDs. Fundamental research in droop mitigation strategies should benefit both blue and green LEDs.

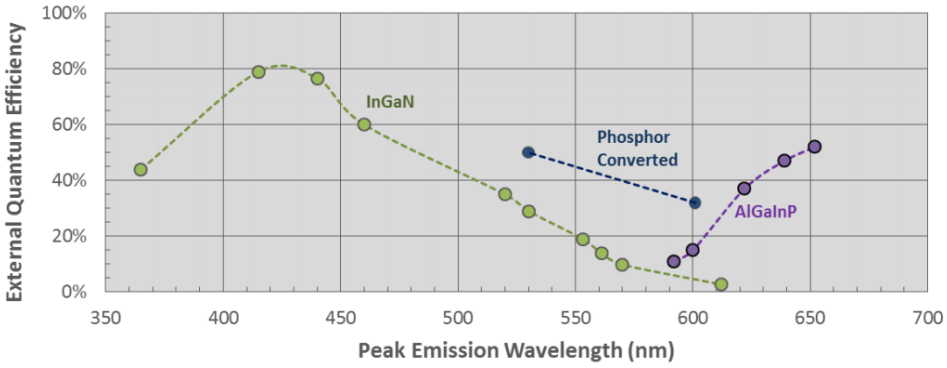

Fig. 1.4.: External quantum efficiency of state-of-the-art LEDs vs. wavelength. (Reprinted from [6], with the permission of ISHS)

4. **Non-uniform Emission Distribution**: Another design challenge in LED is to have uniform light emission throughout the device. Modern commercial LED designs employs multiple quantum wells from by InGaN/GaN as recombination centers. However, typical LED emits most light from 1-2 QWs closest to the p-side, as will be shown in later chapter. This is because the very different transport properties for electrons and holes. Concretely, holes are much more difficult to transport across the MQW and thus heavily concentrated on the p-side, leading to a skewed recombination distribution. Higher hole concentration also leads to more pronounced droop effect.

1.2 The Need For Transport Modeling

The core structure of an LED is known as the active region. The active region (shown in Fig. 1.5) typically consists of several layers of materials, each a few nanometers in thickness, that forms a quantum structure acting as recombination centers for electrons and holes. A key objective for LED design is to properly guide electrons and holes into the active region and maximize the rate of which they recombine via radiative processes. Carrier transport (as well as recombination) in the active region, therefore, is at the heart of LED operation. Fundamental understanding of

LED transport physics holds key to understanding, thus solving, the most important design challenges such as efficiency droop and non-uniform emission distribution.

Once electrons and holes are guided to and confined in the active region, they may go through two types of recombination processes: non-radiative and radiative. Two main non-radiative recombination processes in LEDs are Shockley-Read-Hall (SRH) and Auger recombination. Although non-radiative and radiative recombination are separate physical processes, they are intertwined such that designing LED often needs to consider their competing effects and trade-offs.

In real LED design optimization, it is not feasible to evaluate every design experimentally. Cost is the main reason prohibiting such practice. Another main reason is difficulties in result interpretation due to many interfering effects such as structure changes, crystal growth, electrical and optical response. Therefore, theoretical models essential to guide experimental device design.

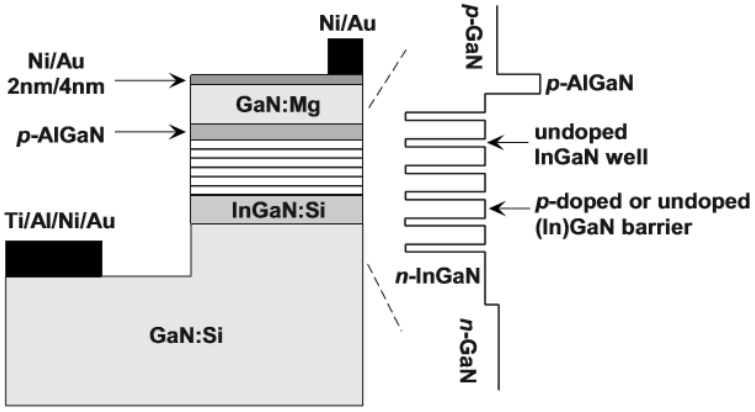

Fig. 1.5.: Schematic of an example LED structure. Figure on the right is a schematic band diagram of the active region. (Reprinted from [7], with the permission of AIP Publishing)

1.3 Predictive Model Must Include Critical Physics

An effective model not only has to describe the behaviors of existing, known devices, but also able to predict and characterize new devices. Such a predictive model is required to be built upon sound physical foundations which allows known physics to be extrapolated into uncharted design territories. The LED active region, although a small scale compared to full device, includes most of the critical transport-related physics of the entire device. Figure 1.6 depicts those relevant processes that govern the LED operation.

The gray lines in Fig. 1.6 are the local band edges (conduction band minimum and valance band maximum). The band diagrams come from full self-consistent simulation results (rather than schematic such as shown in Fig.1.5) that will be described in details in later chapters. These band diagram represents a commercial $GaN/In_{0.13}Ga_{0.87}N$ 6QW LED, which is used as a prototypical device in development of this thesis, and will be shown in several figures that follows.

The 6 QWs in the active region form confined quantum-mechanical states, whose locations and energy levels are marked by the red-dashed lines in the figure. Electrons provided by n-contact on the left and holes provided by p-contact on the right are injected into the center region to recombine. The electrons and holes injected from the contacts are transported into multiple QWs through thermionic emission, coherent tunneling or incoherent scattering processes (such as electron-electron and electron-phonon scattering). Once confined in the quantum states, electrons and holes may undergo three types of recombinations: Shockley-Read-Hall (SRH), Radiative and Auger. There processes are all critical to LED operation and must be captured in a physically-consistent way in a predictive model.

1.4 Why We Need A Better Model

LED modeling is a multi-scale multi-physic problem that involves several physical aspects [8]: carrier transport (electron, hole transport and recombination in the

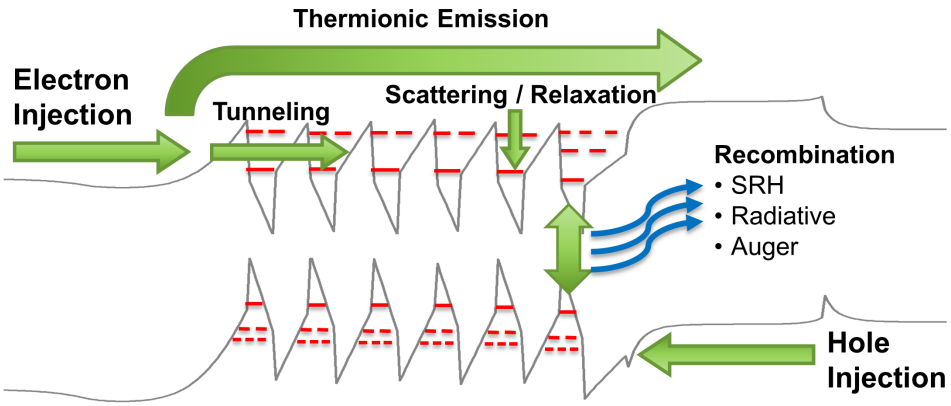

Fig. 1.6.: Relevant physical processes that govern the operation of an LED. These must be covered in a good predictive transport model.

active region), current spreading (current distribution in and near the contact), heat transport and light extraction (ray-tracing or wave propagation). The scope of this thesis is limited to the first part: carrier transport in the LED active region.

In terms of carrier transport, there are distinct physics in different regions with the device. The n and p-doped contacts consists of a single, uniform material (GaN) with small variation in local potential profile and it therefore follows classical physics which can be well-captured by drift-diffusion model. However, the centerpiece of the active region consists of multiple quantum-wells (QW), each only a few nanometers in length. Within such scale, semiclassical transport do not apply since the device physics is dominated by quantum mechanics.

1.4.1 Traditional Approach Towards Transport

Traditional drift-diffusion based modeling approach has been well-established in LED. It is also the backbone of most commercial LED simulation softwares such as Crosslight [9], Silvaco [10] and STR [11].

The electron and hole transport in an LED is depicted in figure 1.7. Electrons are injected from the n-type emitter on the left, and holes are injected from the right from

the p-type collector. Electrons and holes falls into the quantum wells and recombine there. Most of the existing transport modeling tools are based on the semiclassical approach, where the electron and hole transport are described with the drift-diffusion equations:

$$J_{n_{3D}} = -q\mu_n \nabla\phi + qD_n \nabla n - R_n \tag{1.1}$$

$$J_{p_{3D}} = q\mu_p \nabla\phi - qD_p \nabla p - R_p \tag{1.2}$$

Here μ_n/μ_p are the electron/hole mobility, D_n/D_p are the electron/hole diffusion coefficients. R_n/R_p are the recombination coefficients and describes the carrier capture into the quantum well. These transport equations describe only the unbounded carriers (denoted with subscript '3D') which flow above the quantum wells. These '3D' carriers are treated by classical physics and their density are calculated according to its Fermi-Dirac distributions.

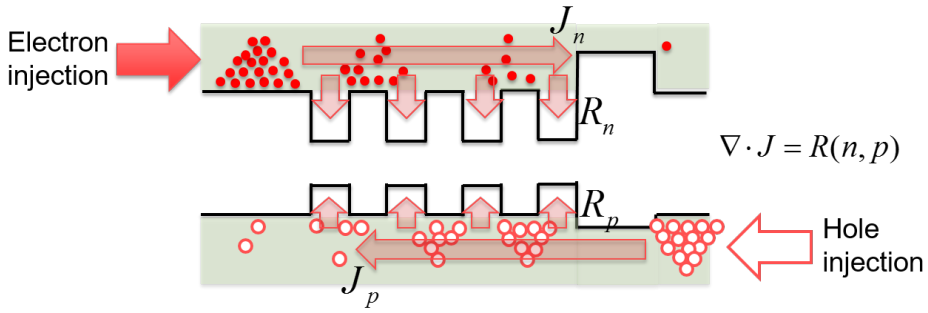

Fig. 1.7.: Schematic of carrier flow above QW in the LED. These carriers are treated with classical physics in most LED models.

However, the same semiclassical treatment of carriers cannot be applied to the carriers in the QWs. As depicted in figure 1.8, the QW carriers are confined within narrow regions of a few nm. As a result, the states are quantized along the confinement direction. To calculate the bounded carrier density in a QW (denoted with subscript '2D'), Schrödinger equation is solved for that QW in order to obtain the allowed energy levels and wavefunction of electrons:

$$(H_0 + q\phi)\Psi = E\Psi \tag{1.3}$$

H_0 is the Hamiltonian of the QW, Ψ is the electronic wavefunction. Once all the eigen energies ϵ and wavefunctions Ψ are obtained for that particular QW, the electron density is calculated by summing up all the occupied states:

$$n_{2D} = \sum_{n,k} |\Psi^2_{n,k}| \cdot f(\epsilon_{n,k}) \tag{1.4}$$

$f(\epsilon_{n,k})$ is the distribution function of an electron (or hole) at energy $\epsilon_{n,k}$.

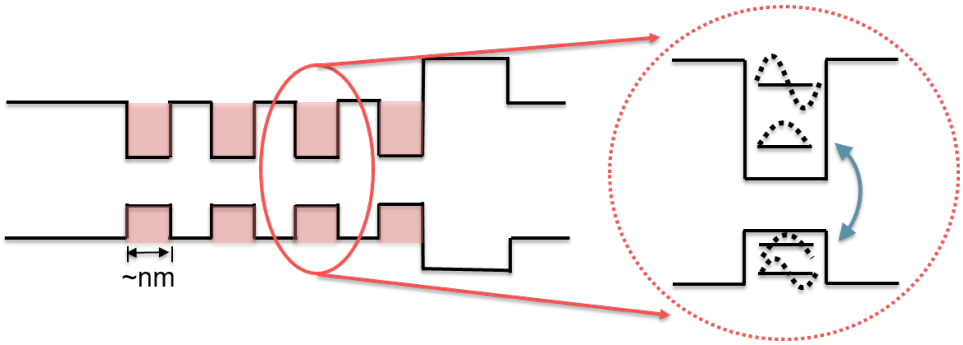

Fig. 1.8.: Schematic of electron wavefunction inside QW of an LED. The allowed energy and wavefunction of the confined electrons obtained by solving the Schrödinger equation.

1.4.2 The Limitations of the Traditional Approach

The semiclassical model is based on the a few fundamental assumptions. The first and foremost is that electrons and holes in the quantum region (QWs) can be partitioned into '2D' density—population that is energetically bound to region, and '3D' density—popluation that has sufficient kinetic energy to escape into the surrounding regions. These two types of carrier densities have fundamentally different behaviors and thus require different treatments as described previously.

The second assumptions is that the bound carriers are sufficiently strongly-confined in the QWs that the states are localized and stationary. Thus the Schödinger equation can be solved in individual QW to obtain the eigenstates. The states are fully coherent and transport of the bound carriers is absent in the confined direction.

The third assumption is that for the unbound carriers coherence is completely lost and thus carrier transport is described by drift and diffusion. The coupling between bound and unbound carriers is captured via a capture rate. This capture rate acts as a recombination for the unbound carriers and generation for the bound carriers.

There are several fundamental limitations of the traditional approach:

1. There exists no clear energetic distinction between the bound and unbound carriers. Typically the barrier band edges are used as the energy boundary. However such distinction is ad-hoc and ambiguous. Although carriers above the QWs are largely unbounded, they are still affected by the quantum confinement. Likewise, the confined carriers can scatter, gain energy and escape out of the QW. Moreover, the distinction can be drawn arbitrarily, especially for non-regular shaped potential, and such distinction may affect result interpretation significantly.

2. The bounded carriers are not in a closed system. Rather, they transport through the structure via thermionic current and tunneling, and as a results those states have a finite lifetime. Traditional model cannot treat the bounded electrons as in an open system and their transport through an extended structure.

3. The coupling between unbounded and bounded carriers involves more assumptions, such as distinct quasi-Fermi levels for '2D' and '3D' densities. It also involves more parameters, choices of which could be ambiguous and subject to interpretation.

4. The semiclassical model assumes two-band parabolic band structure and does not account for the real, complicated band structure in transport calculation.

1.4.3 The Quantum Picture—What the Device Really Looks Like

Figure 1.9(a) summarizes the physics approach of semiclassical model in a picture: above the barriers, carrier transport calculated with the drift-diffusion model where

all quantum effects are omitted. In each quantum well, quantum states are assumed to be confined and localized, and do not participate in transport. States are represented in the picture with red lines, stating the fact that they have infinite lifetime (zero broadening). Figure 1.9(b) puts the same device in quantum picture, and shows what the states really look like in such device scale. As one can see, a distinct quantum interference pattern is observed throughout the energy range of interest. As a result, no distinction should be made between classical and quantum regions. Coupling between continuum and discrete states occurs naturally. The broadening of states is a direct manifestation of finite carrier lifetime, due to frequent scattering events and coupling to the open leads. Therefore, carrier transport occurs through a complex, extended structure, and is directly influenced by the overall quantum-mechanical properties of the system. The electrons fill all the QW ground states and partially fill the excited states. The hole states are spaced much more closely in energy due to their larger effective mass. The heavy and light hole bands are explicitly coupled in this model due to breaking of translational symmetry. The hole charge density spreads in energy over multiple confined quantum states. The complexity of hole states can only be captured with a sophisticated band model, such as the 20 band tight-binding model ($sp^3d^5s^*$ with spin-obit coupling) used in the bulk of this work.

1.5 Structure of This Document

Clearly, a better model was needed: the quantum effects are very pronounced in the length scale of a typical LED, not only below the energy barriers, but in the entire energy range of relevance. The quantum states are coupled with each other, as well as the device contacts. Transport occur through quantum states in a complex, extended structure.

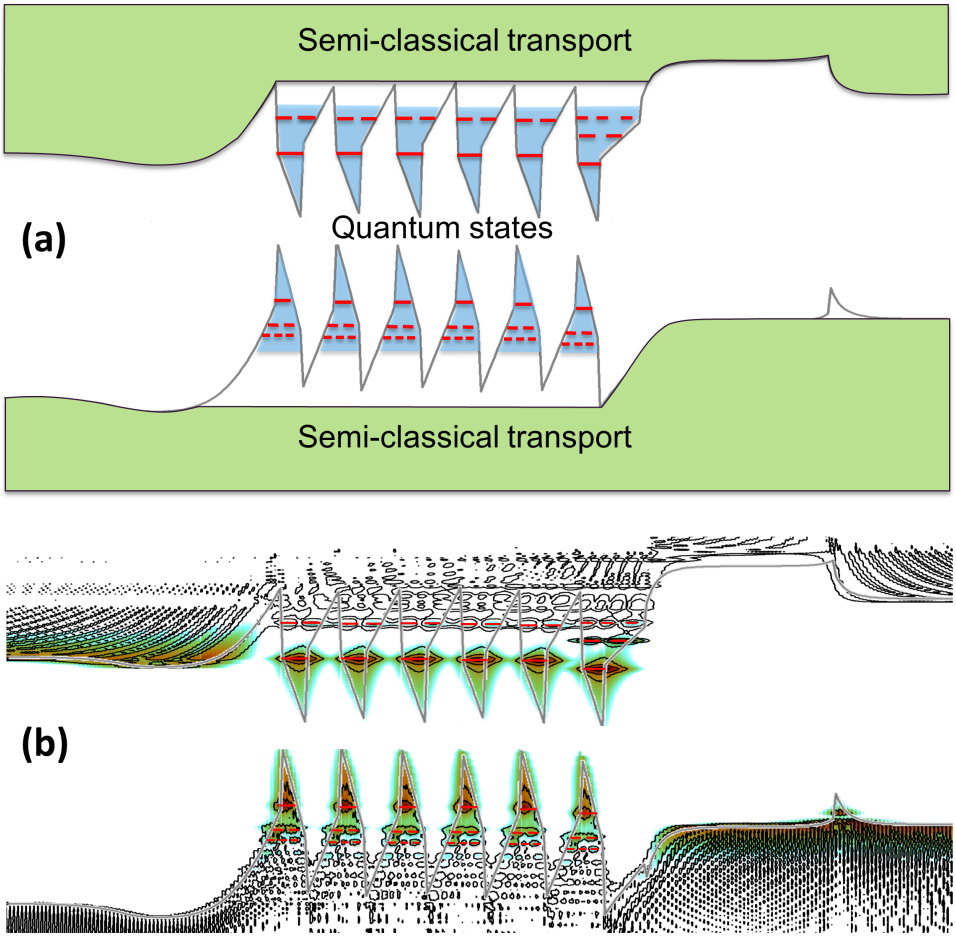

Fig. 1.9.: Comparison between semiclassical and quantum pictures. (a) Semiclassical picture—carrier transport is based on semiclassical drift-diffusion model; quantum states in each individual quantum well are calculated with Schrödinger equation. Quantum states are confined, localized in each individual quantum well, and do not participate in transport. Using lines to represent quantum states indicates that those states have infinite lifetime (zero broadening). (b) Quantum picture—what the quantum mechanical states in the device really look like.

2. THEORY AND MODEL DETAILS

2.1 The Overview

In the previous chapter, it was shown that the active region of a typical LED consists of MQW that contains confined quantum states serving as recombination centers, and highly-doped n/p contacts guiding electrons and holes from the opposite ends into the MQW. Different quantum wells (QWs), each in the few nanometer scale, are coupled to each other through tunneling and thermionic transport. These transport mechanisms as well as carrier capture into the QWs are integral part of the LED operation and need to be well understood. Carrier transport defined in a quantum system is therefore at the heart of the LED operation. Solutions to droop and non-uniform light emission lie in the quantitative understanding of the quantum mechanical carrier transport physics through the complex nanostructure. State-of-the-art modeling approach toward carrier transport in LEDs are based on drift-diffusion-based semiclassical models, which either heuristically patch in quantum effects or simply ignore them.

The Non-Equilibrium Green Function (NEGF) formalism is the accepted state-of-the-art carrier transport theory for a wide range of nanoscale semiconductor devices, such as resonant tunneling diodes [12], field-effect transistors [13], quantum-cascade lasers [14] and solar cells [15]. Under the NEGF formalism, tunneling, thermionic emission [16], scattering [17] and recombination [18] are all treated on the same footing throughout the device. Therefore, it is the ideal candidate for LED modeling, which overcomes limitations of the traditional model laid out in the previous chapter.

NEGF has, however, not been extensively employed for the modeling of opto-electronic devices, due to a variety of issues. NEGF tends to be computationally very expensive, especially for realistically extended complex devices that include in-

coherent scattering or require full 2D or 3D modeling. There is no accepted NEGF physics-based self-energy that leads to full thermalization in high carrier density device regions. NEGF has been previously applied to modeling LED devices [18] [19], however, due to the computation challenges, they are typically limited in terms of device scale (simplified device structure, band structure) and applicability. Multi-scale approaches combining drift-diffusion and NEGF have been proposed recently [20], however in such approaches, transport is based on semiclassical physics with NEGF correction and appropriate relaxation of carriers in the leads and QWs is not included.

The goal of this chapter is to establish some background knowledge by introducing the NEGF formalism and key terminologies. It also aims to point out the main limitations that prevents directly applying the NEGF formalism to our LED modeling problem. This chapter also serves as the foundation for our novel modeling approach, which will be introduced in details in the next chapter.

2.2 NEGF Transport in "Quasi" One-Dimensional System

The main target device in this modeling work is the MQW LED, depicted in Fig. 1.5. The structure is grown layer-by-layer via MOCVD (Metal-Organic Chemical Vapour Deposition) process, with each layer of material being a few nanometers in thickness. The thickness of each layer is much smaller compared to the area of cross section ($\sim \mu m^2$). The electrons and holes are injected from bottom and top and transport vertically. Therefore, we could treat the vertical transport problem as a one-dimensional transport problem, and applying periodic boundary conditions in the transverse direction (thus the notion "quasi-1D"). Here longitudinal (or vertical) represents the direction of current flow and transverse (or cross-sectional) represents the direction perpendicular to current flow.

The device Hamiltonian is denoted by $H_{Device}(\boldsymbol{k}, E)$, with E being the total energy and $\boldsymbol{k}$ is the transverse momentum. Atomistic tight-binding (TB) representation is adopted in this work, where Hamiltonian is represented by localized orbital basis [21],

and transverse periodicity of the Hamiltonian matrices is incorporated with the Bloch theorem. $H_{Device}(\boldsymbol{k}, E)$ is a block tri-diagonal matrix, with each block being a slab of atomic unitcell in the longitudinal direction. For Wurtzite structure grown along the c-axis, each slab is just one atom (either anion or cation) represented by a matrix of size $N \times N$, where N is the number of basis orbitals.

2.2.1 Coherrent Transport

Quantum kinetics equations

The equation of motion for electrons is described by the Dyson equation [22]:

$$G^R(\boldsymbol{k}, E) = (E - H_{device}(\boldsymbol{k}, E) - \Sigma^R(\boldsymbol{k}, E))^{-1} \tag{2.1}$$

G^R is the Retarded Green's function, which describes the quantum-mechanical properties of the system.

Σ^R is the retarded self-energy. It is the sum of all self-energies from interactions between the system and various forms of internal or external physical processes:

$$\Sigma^R = \Sigma^R_S + \Sigma^R_D + \Sigma^R_\eta \tag{2.2}$$

$\Sigma^R_{S/D}$ represent self-energies coupling the device to the two contacts, "source" and "drain" (or "emitter" and "collector", as is the more familiar terminology for diode). Σ^R_η includes any scattering processes involved (will be discussed later). For coherent transport, Σ^R_η does not exist and the only self-energies existing are due to the contacts.

The advanced Green's function is closely related to the retarded Green's function by:

$$G^A(\boldsymbol{k}, E) = \left(G^R(\boldsymbol{k}, E)\right)^\dagger \tag{2.3}$$

Physically the Green's function $G(x, x')$ could be viewed as the response wave-function at x when a unit excitation is applied at position x'. Therefore, the retarded

and advanced Green's functions represent two different boundary conditions: G^R corresponds to outgoing waves and G^A corresponds to incoming waves.

The retarded and advanced Green's functions contain information about the dynamics (states) of the device. To obtain the kinetics (filling of states), the correlation Green's functions are calculated following Keldysh equation in its discretized form [23]:

$$G^<(\boldsymbol{k}, E) = G^R(\boldsymbol{k}, E)\Sigma^<(\boldsymbol{k}, E)G^A(\boldsymbol{k}, E) \tag{2.4}$$

$$G^>(\boldsymbol{k}, E) = G^R(\boldsymbol{k}, E) - G^A(\boldsymbol{k}, E) + G^<(\boldsymbol{k}, E) \tag{2.5}$$

Physical quantities

The spectral function $A(\boldsymbol{k}, E)$ of the system can be obtained directly from its retarded Green's function:

$$A(\boldsymbol{k}, E) = i\big(G^R(\boldsymbol{k}, E) - G^A(\boldsymbol{k}, E)\big) \tag{2.6}$$

The local density of states (LDOS), which contains information regarding the spacial and energy distribution of available states, is obtained by taking the diagonal of spectral function:

$$LDOS(x, \boldsymbol{k}, E) = A(x, x' = x; \boldsymbol{k}, E) \tag{2.7}$$

The electron density is obtained directly from the correlation function $G^<$:

$$n(x) = \frac{-2i}{A\Delta}\sum_{\boldsymbol{k}}\int\frac{dE}{2\pi}G^<(x, x' = x; \boldsymbol{k}, E) \tag{2.8}$$

Here "$A\Delta$" can be viewed as the normalization volume for each discretized lattice point, where Δ is the discretization spacing and A is the cross section area. The integral inside equation 3.7 can be viewed as the density matrix at momentum $\boldsymbol{k}$ and

energy E. Factor of 2 accounts for spin degeneracy. If we replace the $G^<$ with A (eq. 2.6), then we get the electron density of states instead of density.

The general current density formula can also be obtained from $G^<$ as (without magnetic field):

$$2\pi J(x;\boldsymbol{k},E) = -\frac{ie\hbar}{2m}\Big[(\nabla-\nabla')G^<(x,x';\boldsymbol{k},E)\Big]_{x'=x} \tag{2.9}$$

m is the free electron mass. ∇ and ∇' are gradient operators on x and x' respectively.

In the nearest-neighbor TB representation, and integrating over energy and momentum, the spatial-resolved total current density is obtained:

$$J_L = \frac{2e}{\hbar A}\sum_{\boldsymbol{k}}\int\frac{dE}{2\pi}2Re\big\{tr[t_{L;L+1}G^<_{L+1;L}(\boldsymbol{k},E)]\big\} \tag{2.10}$$

The position x dependency in previous equations is replaced here by the layer index L, and $t_{L;L+1}$ is the Hamiltonian coupling element between two neighboring layers. $tr[...]$ indicates a trace over the orbital indices.

If we only consider coherent transport, then the current density at any particular $(\boldsymbol{k},E)$ must be uniform across the device. This is because the total current (summing over all $(\boldsymbol{k},E)$) must be conserved throughout the device, and since there is no internal or external processes allowing an electron to gain energy or momentum, the current density $J(\boldsymbol{k},E)$ must be constant. Therefore, in the coherent transport regime, one only needs to calculate the "terminal" current—current density right at the contact. It can be derived that

$$J = -\frac{2e}{\hbar A}\sum_{\boldsymbol{k}}\int\frac{dE}{2\pi}tr\{\Gamma^R_S G^R\Gamma^R_D G^A\}(f_S-f_D) \tag{2.11}$$

where

$$\Gamma^R = i\big(\Sigma^R-\Sigma^{R^\dagger}\big) \tag{2.12}$$

and $f_{S/D}$ is the Fermi function at the two contacts. The quantity inside the $tr\{...\}$ is the transmission coefficient [24].

2.2.2 Incoherrent Transport

When scattering processes are present, they are included as self-energies:

$$\Sigma^{<} = \Sigma_1^{<} + \Sigma_2^{<} ... + \Sigma_n^{<}$$

where 1...n are different scattering processes, including charged-impurity, surface roughness, acoustic phonon, optical phonon, electron-photon, electron-electron scattering. Calculating scattering self-energies is an involved, computational expensive process, and needs to be done in a self-consistent fashion involving Green's functions as illustrated in the fig. 2.2.

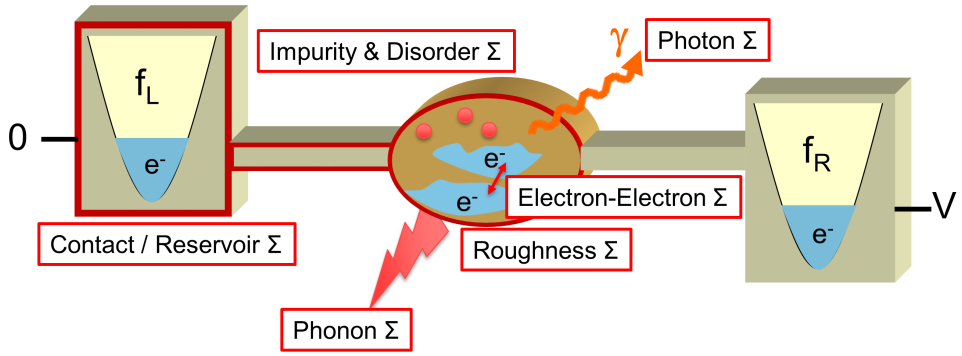

Fig. 2.1.: Device treatment in NEGF simulation, including contact and various scattering mechanisms.

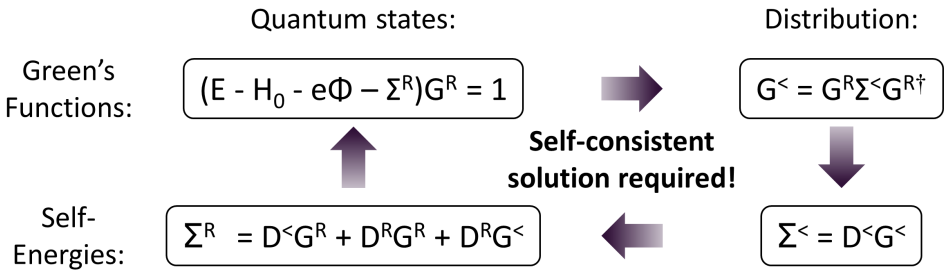

Fig. 2.2.: Schematic of self-consistent calculation process for incoherent transport.

2.2.3 Incoherrent Transport—Phenomenological Scattering Model

Büttiker Probe

One convenient way to include the effect of scattering in NEGF is using a phenomenological approach, known as the Büttiker Probe (BP). In this approach, scattering processes are treated on the equal footing with the emitter and collector contacts. Instead of being attached to the two ends of the devices, the probes are attached to each atomic/orbital site within the device. The retarded and lesser-Green's function are still similar to the ballistic case, with an the additional self-energy taking into account the local probes:

$$G^R = (E \cdot I - H - \Sigma_e^R - \Sigma_c^R - \Sigma_B^R P)^{-1}$$

$$G^< = G^R(\Sigma_e^< + \Sigma_c^< + \Sigma_{BP}^<)G^{R\dagger}$$

In contrast to the contact self-energies, which contain only one non-zero block, the BP self-energies is a block-diagonal matrix, and each block in the diagonal describes connection to one atomic site. Similar to contacts, the probes also assume a local equilibrium distribution with an artificial Fermi level, and additional solver must be included to to balance out these Fermi levels and ensure overall current conservation in the device. The main strength of Büttiker Probe is its computational efficiency.

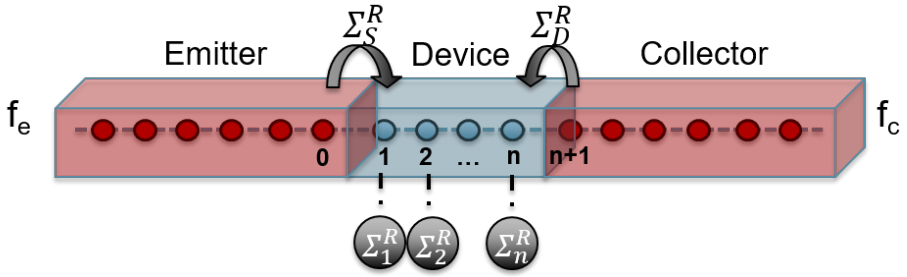

Fig. 2.3.: Schematic of a device connected to two contacts, emitter and collector. Scattering process is mimicked with local contacts known as "Büttiker probe".

With careful adjustment, it can almost mimic any type of scattering within the device,

even electron-electron scattering, without the need for the self-consistent calculation described previously.

2.3 NEGF's challenges in LED Modeling

The NEGF formalism has proven successful in predicting the electrical and/or optical characteristics for a wide range of devices, including resonant tunneling diodes [12], field-effect transistors [13], quantum-cascade lasers [14] and solar cells, and has been accepted as the state-of-the-art transport theory in such devices. There has been significant effort in enabling full quantum transport simulation for LEDs, most notably [18] [19] [25]. However, quantum transport simulation was not able to evolve from a proof-of-concept into the accepted state-of-the-art for LEDs. NEGF tends to be computationally very expensive, especially for realistically extended complex devices that include incoherent scattering, and there is no accepted NEGF physics-based self-energy that leads to full thermalization in high carrier density device regions. The main challenges of full NEGF based quantum transport simulation in LED are summarized below.

The main challenge in NEGF modeling of realistic device is solving self-energies that describe various scattering processes, since it is usually requires a self-consistent calculation, which is a highly computationally intensive process. Therefore, assumptions are usually made to simplify the calculation and ease the numerical load, include self-consistent Born approximation for the many-particle interaction, eliminating off-diagonal elements in scattering self-energy, and decoupling between energy-momentum/momentum-momentum whenever possible. Even with those assumptions, device simulation with NEGF with scattering is still too numerically expensive compared to semiclassical-based modeling approach, and as a result NEGF has only been applicable to a limited range of devices, with limited dimension (a few nm) of the active modeling region.

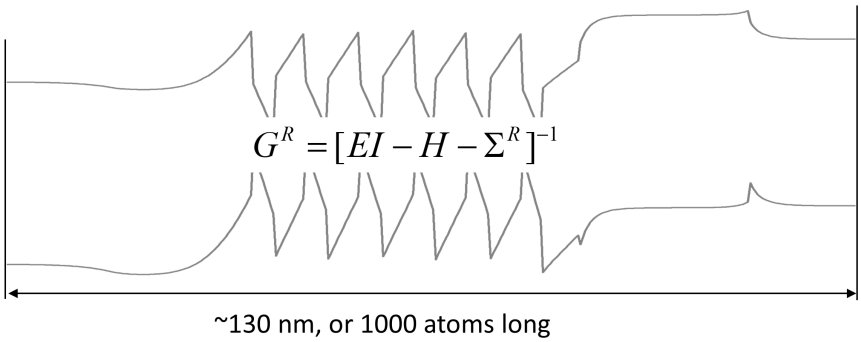

Fig. 2.4.: Typical length scale of critical carrier transport region in an LED.

Compared to a typical device where NEGF is applicable, modeling in LED is far more computationally restrained. A typical LED structure is shown in figure 2.4. The active modeling region in this LED is around 130 nm, or roughly 1000 atoms long. Using a 20 orbials per atom representation (which is the standard used in this work), a single matrix has dimension of 20,000 x 20,000. Inverting such a matrix along, as is needed in calculating G^R, is already numerically challenging, consider matrix inversion scales as $\mathcal{O}(n^3)$. Solving a coherent transport problem (no scattering) would need to invert and store such matrices tens of thousands times.

LED also includes a wide range of tightly-coupled physical processes. In each of the quantum-well region, electrons in the QW states interacts with ones in the continuum states via phonon scattering, including both acoustic and polar-optical phonons. Also, due to the nature of nitride family of materials, electrons interacts strong with the lattice via impurity scattering and piezoelectric scattering. More importantly, electrons and holes exist simultaneously in the device, and they go through various recombination processes, like Shockley-Read-Hall (SRH), radiative, and Auger recombinations, these processes require additional scattering self-energies to capture. Lastly but perhaps most importantly, strong electron-electron scattering exists in the device, due to the high carrier densities in the confined region in the order of $10^{12} cm^{-2}$. With such high density, electron-electron could dominate over all other processes, and currently there is no computationally feasible methods to solve electron-electron scattering rig-

orously. Because of these limitations, there have been no feasible quantum model for commercial LED devices. Therefore, we need a new approach.

3. MULTI-SCALE-EQUILIBRIUM-NONEQUILIBRIUM MODEL

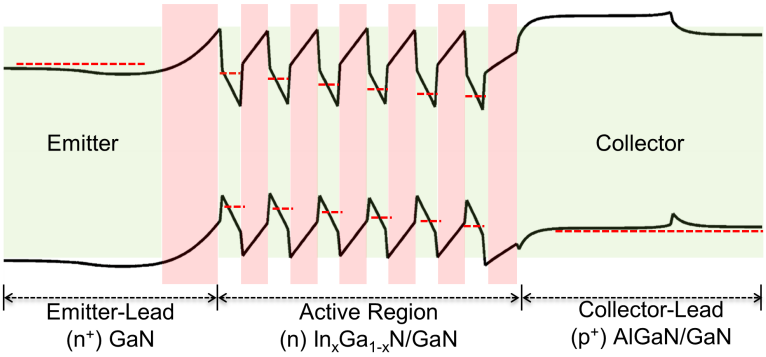

Fig. 3.1.: LED structure that consists of an n-GaN layer; a low-doped active region made of InGaN/GaN MQW; an AlGaN electron-blocking layer and a p-GaN layer. The distiction between equilibrium (eq–green) and non-equlibrium (neq–red) regions are marked as different colors. Each eq-region has a unique quasi Fermi level, for holes and electrons as indicated by a red dashed line. The Fermi level drops across the device are depicted not to scale to emphasize that they are different from one QW to the next.

3.1 Device Partitioning Overview

The new model developed in this work is the Multi-scale-Equilibrium-noneQuilibrium (MEQ) approach. The method is designed for specific families of devices structures similar to the MQW LED. The methodology is based on carrier scattering versus carrier tunneling oriented partitioning of the device and the method originally published

in Ref. [17] is extended to include multiple carrier reservoirs in the interior of the device.

Figure 3.1 shows the band diagram of a typical AlInGaN LED active region. The LED structure contains an n-doped GaN layer, MQW consisting of barriers/wells made of $GaN/In_xGa_{1-x}N$, $Al_xGa_{1-x}N$ electron-blocking-layer (EBL) and a p-doped GaN layer. The Hamiltonian of the total device H_{Device} is divided as:

$$H_{Device} = H_{LL} + H_{B1} + H_{W1} + H_{B2} + H_{W2} + ... + H_{W6} + H_{B7} + H_{RL} \tag{3.1}$$

Here, LL and RL refer to the left GaN and right AlGaN plus GaN regions, respectively. Barriers (wells) are indicated with B (W) followed by the corresponding numerical label shown in Fig. 3.1. The device Hamiltonian is represented with atomistic tight-binding (TB) with 20 orbitals per atom ($sp^3d^5s^*$ representation including spin-orbit interaction). TB Parameters for GaN, InN and AlN were each fitted following Ref. [26] and alloy parameters are interpolated with virtual crystal approximation (VCA). TB parameter fitting process and results will be discussed in more detail later on.

The n- and p-layers are highly doped to provide electrons/holes to the MQW, and the MQW forms quantum-mechanical density of states (DOS) serving as recombination centers for electrons and holes. The n/p layers and the QWs contains high carrier density, typically on the order of $10^{12}cm^{-2}$. In those regions, carrier scattering is strong and they are separated by large tunneling barriers, and therefore, we consider each of them a local equilibrium carrier reservoir with a unique local quasi Fermi level ('equilibrium' or 'eq' region, marked green in Fig. 3.1). Individual reservoirs are connected to one another via tunneling and thermionic emission. The tunneling barriers are treated as non-equilibrium (or 'neq', marked red in Fig. 3.1) regions where coherent quantum transport is solved with the NEGF method.

In each reservoir, an imaginary optical potential (η) is included in the diagonal of the Hamiltonian [12] to mimic the effect of scattering. The effect of η leads to broadened states in the quantum wells, and the broadening width is chosen to match the experimentally measured [27] photoluminescent (PL) emission width of $\sim$100meV.

For comparison and when indicated explicitly, an artificially small broadening of 10 meV is assumed as well. η assumes a constant, energy independent value for energies above the bulk band-edge, and decays exponentially into the bandgap with a decay length of 50 meV [28]. This exponential decay properly accounts for transport through the band tail states [29].

One important thing to point out is that the η here plays an equivalent role as Büttiker probe discussed in section 2.2.3. However, the main difference is the η value at each spacial location, momentum and energy is not chosen to conserve current. Rather an "averaged" value is used because of thermal equilibrium assumption. Only carrier distribution, no current is calculated in the equilibrium regions.

3.2 Green's Function and Self-Energy

The retarded Greens function of the device is solved by recursively inverting

$$G^R_{Device} = [E - H_{Device} - \Sigma^R_S - \Sigma^R_D - \Sigma^R_\eta]^{-1} \tag{3.2}$$

following the partitioning discussed above and in the lower part of Fig. 3.2. Here Σ^R_η is the self-energy matrix containing the η, so it has the same dimension as the Hamiltonian. Figure 3.2 shows the Hamiltonian matrix pattern with η. As discussed early and also in the previous chapter, the Σ^R_η encompasses all scattering mechanisms, specifically impurity, acoustic phonon, polar optical phonon and electron-electron scattering. The value of η is related to the scattering rate of the system via $S \approx 2 \cdot \eta/\hbar$ [30]. The η self-energy takes on the following form

$$\Sigma^R_\eta(x, \boldsymbol{k}_{||}, E) = \begin{cases} \eta \cdot e^{-\frac{E_C(x,\boldsymbol{k}_{||})-E}{\lambda}}, & \text{below conduction band} \\ \eta, & \text{above conduction band} \end{cases} \tag{3.3}$$

It is constant above the band edge and exponentially decaying below it. λ is the decay length (η-tail), which is a parameter that can be obtained through experimental measurement of density of states, or calculated with NEGF include scattering. Here we assume it's electron so E_C is the conduction band minimum. Note the same

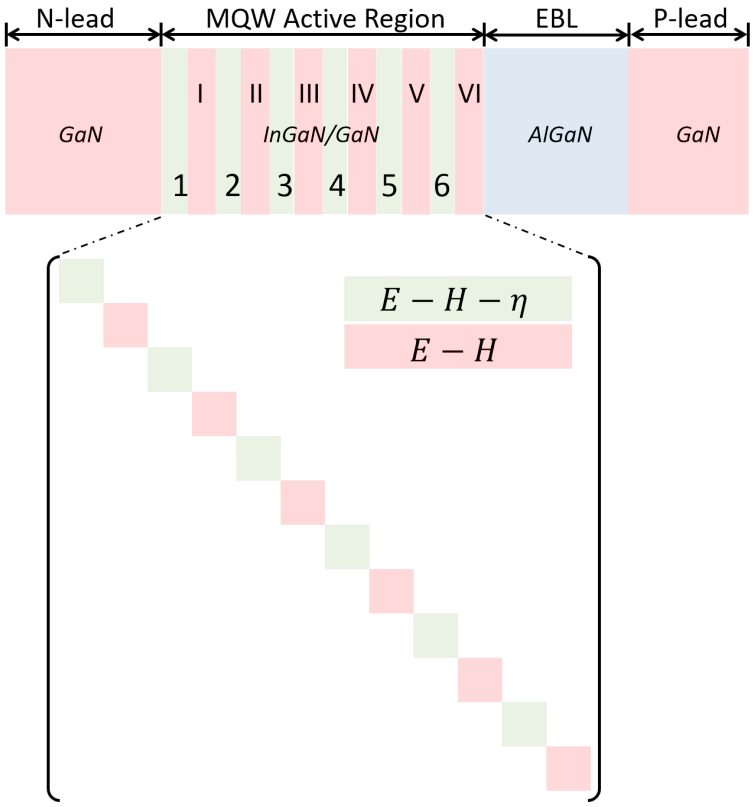

Fig. 3.2.: (a) Schematic diagram of the LED structure modeled in this work. It is the same structure as show in Fig.3.1 but different view. (b) Schematic drawing of inverse of GR matrix form in the active region, showing scattering (η) was only included in the QWs where thermal equilibrium assumption was applied. Barriers were treated as coherent transport regions.

formula with trivial changes can be applied to the holes. Also note that the band edges here is not only position dependent, but also dependent on in-plane momentum $k_{||}$. To get the band extrema at each $k_{||}$, it requires additional bandstructure calculation on individual equilibrium regions. Calculating momentum-dependent band edges will be discussed in more details in a later chapter.

The lesser Greens function is solved individually for each region. In equilibrium regions 'i', the lesser Greens function is solved with the local Fermi level μ_i

$$G^{<}_{i,eq.} = -f(\mu_i)(G^R_i - G^{R\dagger}_i) \tag{3.4}$$

For nonequilibrium regions j, the lesser Greens function is solved with

$$G^{<}_{j,noneq.} = G^R_j \Sigma^{<}_j G^{R\dagger}_j \tag{3.5}$$

G^R_j , the submatrix of G^R_{Device} in the region j, and $\Sigma^{<}_j$ is the sum of the contact self-energy of region j due to its coupling with the neighboring equilibrium regions $j-1$ and $j+1$

$$\Sigma^{<}_j = -f(\mu_{j+1})(\Sigma^R_{j+1} - \Sigma^{R\dagger}_{j+1}) - f(\mu_{j-1})(\Sigma^R_{j-1} - \Sigma^{R\dagger}_{j-1}) \tag{3.6}$$

Here, f denotes the Fermi distribution function. The left and right connected self-energies Σ^R_{j+1}, Σ^R_{j-1} appear in the standard recursive Greens function algorithm and describe the coupling of region j with its neighboring regions (see e.g. Ref. [17]). Note that in this way, all quantum wave effects of surrounding regions are included in the solution of the j-th region [12].

3.3 Basic Physical Quantities

There are three main quantities of interested, which are provided by this model: charge density, coherent current, recombination current.

3.3.1 Charge Density

Calculation of charge density follows formula 3.7. With the multi-region approach, the lesser Green's function is solved in each sub-region following equation 3.4 and used to calculate density in that sub-region. Combining density in individual sub-region results in the density profile in the overall device. For each equilibrium region 'j', the

charge density is a simple function that depends on the Fermi level 'μ_j' of that region alone:

$$n_{j,eq.} = \frac{-2i}{A\Delta} \sum_{\boldsymbol{k}} \int \frac{dE}{2\pi} diag\{A_{eq,j}(\boldsymbol{k}, E)\} \cdot f(\mu_j) \tag{3.7}$$

The electron density in the equilibrium region is the simple product of the Fermi function with its local density of states. It is important to point out that calculating the equilibrium density only requires the diagonal of the Green's function G^R, which has crucial implication on computation reduction compared to the standard NEGF model, this pointer will be further illustrated later. The non-equilibrium region density, on the other hand, will depend on both Fermi levels—of the equilibrium region to its left and right. This can been seen from eq. 3.6.

3.3.2 Coherent Current

The current is calculated only in the non-equilibrium regions. To understand why it only necessary to calculate current in the non-eq (barrier) regions, it is useful to borrow the analogy of a water tank. Consider a giant swimming pool filled with fresh water. To keep the water fresh, two pipes are connected to the pool: the one on the left injects fresh water; the one on the right drains the water from the tank. If the pipes' diameter is small enough compared to the size of the pool, then it is a fairly good assumption to consider the water inside the tank to be still—any water added to the tank will be quickly brought to equilibrium with the rest, and any water extracted out of the tank via the second pipe not significant enough to disrupt the equilibrium. Here the electrons in each eq. region is assumed to be in thermal equilibrium, with a well-defined local Fermi level (water level), and the coherent currents through two neighbor barriers play similar roles as the two water pipes.

The coherent current through an non-equilibrium region is calculated with standard NEGF. For an non-eq region 'j'

$$J_{j,noneq.} = \frac{-2e}{\hbar A}\Sigma_{\boldsymbol{k}} \int \frac{dE}{2\pi} tr\big\{(\Sigma^R_{j-1} - \Sigma^{R\dagger}_{j-1})G^R_j(\Sigma^R_{j+1} - \Sigma^{R\dagger}_{j+1})G^{R\dagger}_j\big\}\big\{f(\mu_{j-1}) - f(\mu_{j+1})\big\} \tag{3.8}$$

The term inside the "tr{...}" is the transmission coefficient ($\boldsymbol{k}$, E dependent) of the sub-region 'j'

$$T_{j,noneq.} = (\Sigma^R_{j-1} - \Sigma^{R\dagger}_{j-1})G^R_j(\Sigma^R_{j+1} - \Sigma^{R\dagger}_{j+1})G^{R\dagger}_j \tag{3.9}$$

The current equation in 3.8 is same as the Laudau formalism. It is important to note the transmission formula eq. 3.9 captures both the thermionic emission (current flow above barrier) as well as ballistic tunneling (current below the barrier).

3.3.3 Recombination

Recombination is a crucial process in LED, and the driving force of LED operation. In traditional semiclassical-based models, recombination is included as a rate term in the continuity equation. There are mainly three recombination mechanisms in an LED: Shockley-Read-Hall (SRH), radiative and Auger. The total recombination current is the sum of all three components

$$J^R_{total} = qV(J_A + J_B + J_C) \tag{3.10}$$

The traditional model calculates recombination with a simple heuristic formula known as the 'ABC' model

$$J^R_{total} = qV(An + Bn^2 + Cn^3) \tag{3.11}$$

Or some forms of its variation [31]. The 'ABC' model has been well-established [32] and proven to fit well to the experiment data with well-chosen A, B, C coefficients [33].

Non-radiative recombination in the NEGF formalism do not currently exist. Radiative recombination in NEGF has been previously developed [15] and applied to

LED simulation [18] [19] [34]. It requires to calculate the scattering current due to the photon field, which similar to any other scattering, is an computational intensive process. Therefore, the lack of physical and efficient recombination models is a major drawback of applying NEGF formalism on LED.

A main strength of this model is its compatibility of heuristic recombination model within a (more rigorous) NEGF framework. This is the first known NEGF model that incorporates all three recombination mechanisms in an efficient way, and (as will be shown later) achieving accurate simulation results on realistic LED device.

The recombination current in each equilibrium region 'j', is calculated from its equilibrium electron and hole density via a slightly modified set of ABC equations

$$\begin{aligned} J_{j,A} &= A_j \cdot \frac{n_j p_j - n_{intri}^2}{\tau_N(p_j + p_1) + \tau_P(n_j + n_1)} \\ J_{j,B} &= B_j \cdot n_j p_j \\ J_{j,C} &= C_j \cdot (n_j^2 p_j + p_j^2 n_j) \\ J_{j,total}^R &= J_{j,A} + J_{j,B} + J_{j,C} \end{aligned} \tag{3.12}$$

Here n_{intri} is the intrinsic carrier concentration, and τ_N and τ_P are the electron and hole lifetimes. The values of n_1 and p_1 correspond to electron and hole concentrations calculated when the quasi-Fermi energy is equal to the trap energy. This set of recombination equations couple electron and hole density into a rate equation, similar to the semiclassical model treatment, while preserving the quantum characteristics of physical quantities. This also allows the LED internal quantum efficiency (IQE) to be calculated, and capturing the efficiency droop, as will be demonstrated later.

3.4 Current Conservation—Tying It All Up

In any transport model, one of the fundamental requirement is current conservation, that charge carriers cannot be created or destroyed out of nowhere.

$$\frac{\partial(n,p)}{\partial t} = \nabla \cdot J + G(n,p) - R(n,p)$$

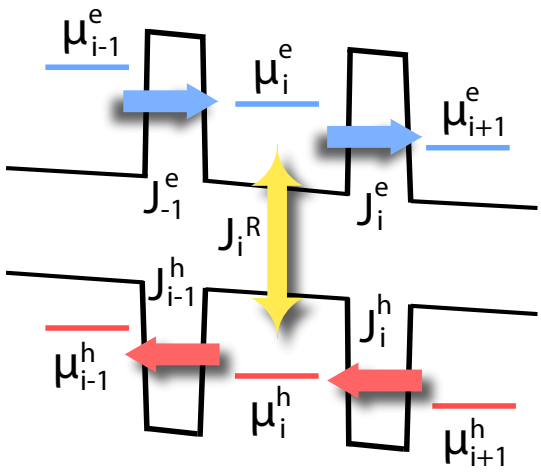

Fig. 3.3.: Schematic diagram of various current component in a single section of LED. Local Fermi levels need to adjust so that overall current is converbed.

n and p are electron and hole charge density, J is the spatial current, G and R are generation and recombination. For steady state, $\frac{\partial(n,p)}{\partial t} = 0$, and consider charge recombination to be the main mechanism (in LED). Then the equation becomes:

$$\nabla \cdot J = R(n,p)$$

Which states that the net current must be balanced out by recombination. In our model, the LED is divided up into multiple equilibrium and non-equilibrium regions and various quantities are calculated separately. Therefore, a key component of the model is to ensure that current is conserved throughout the device, i.e, net current is balanced out by total recombination current. Consider a separate quantum well (with index 'i') shown in figure, it is separated from two neighbor quantum wells 'i-1' and 'i+1' by two barriers. For the i-th QW, the net current components flowing into and out of the region are J^e_{i-1}, J^h_{i-1}, J^e_i, J^h_i, which are the electron/hole current flowing between the i-th QW and its left/right neighbors. In order for current to conserve in

the i-th QW, both the net electron current and hole current must be balanced by the recombination current:

$$J_{i-1}^{e} - J_{i}^{e} - J_{i}^{R} = 0$$

$$J_{i-1}^{h} - J_{i}^{h} + J_{i}^{R} = 0$$

and the total recombination current:

$$J_{i}^{R} = J_{i}^{SRH} + J_{i}^{Rad} + J_{i}^{Auger}$$

which consist of SRH, radiative, and Auger component explained in the previous section. Note the sign convention here: positive current indicates electron current flowing into the region (or hole current flowing out of the region). For an LED consist of N quantum wells, a set of 2N equations must be satisfied simultaneously:

$$\Delta J_{1}^{e} = J_{0}^{e} - J_{1}^{e} - J_{1}^{R} = 0$$

$$\Delta J_{2}^{e} = J_{1}^{e} - J_{2}^{e} - J_{2}^{R} = 0$$

$$\vdots$$

$$\Delta J_{N}^{e} = J_{N-1}^{e} - J_{N}^{e} - J_{N}^{R} = 0$$

and

$$\Delta J_{1}^{h} = J_{0}^{h} - J_{1}^{h} - J_{1}^{R} = 0$$

$$\Delta J_{2}^{h} = J_{1}^{h} - J_{2}^{h} - J_{2}^{R} = 0$$

$$\vdots$$

$$\Delta J_{N}^{e} = J_{N-1}^{h} - J_{N}^{h} - J_{N}^{R} = 0$$

All the recombination current components depends only on the equilibrium density n_i and p_i in the QW, which is depends directly on the quasi-Fermi levels of that QW μ_i^e and mu_i^h. Current flowing between i-th QW and its left neighbor i-1 is the ballistic current calculated with the Landau formula which depends on the quasi-Fermi levels of the two QWs $\mu_i^{e/h}$, $\mu_{i-1}^{e/h}$ and the transmission of the barrier separating those two

QWs. Therefore, each equation in previous set is a function of three quasi-Fermi levels.

$$\Delta J_1^e(\mu_0^e, \mu_1^e, \mu_2^e) = 0$$

$$\Delta J_2^e(\mu_1^e, \mu_2^e, \mu_3^e) = 0$$

$$\vdots$$

$$\Delta J_N^e(\mu_{N-1}^e, \mu_N^e, \mu_{N+1}^e) = 0$$

and

$$\Delta J_1^h(\mu_0^h, \mu_1^h, \mu_2^h) = 0$$

$$\Delta J_2^h(\mu_1^h, \mu_2^h, \mu_3^h) = 0$$

$$\vdots$$

$$\Delta J_N^e(\mu_{N-1}^h, \mu_N^h, \mu_{N+1}^h) = 0$$

The 0-th and N+1-th equilibrium regions are connected to the source and drain, with their quasi-Fermi levels fixed at the source and drain Fermi levels:

$$\mu_0^e = \mu_0^h = E_{Fs}$$

$$\mu_{N+1}^e = \mu_{N+1}^h = E_{Fd}$$

Therefore, solving for current conservation is equivalent to solving for 2N quasi-Fermi levels $\mu_i^{e/h}$ with 2N equations $\Delta J_i^{e/h}(\mu_{i-1}^{e/h}, \mu_i^{e/h}, \mu_{i+1}^{e/h}) = 0$, with $i = 0...N$.

A standard method for solving systems of non-linear equations is the Newton-Raphson method. Newton-Raphson method is known for its efficient convergence to a root, given a sufficiently good initial guess. Here we have applied Newton-Raphson method to solving the current conservation problem. The current conservation problem is equivalent to finding a set of quasi-Fermi levels μ_i^e and mu_i^h where $i = 1...N$, such that the set of functions

$$\Delta J_i(\mu_1^e...\mu_N^e, \mu_1^h...\mu_N^h) = 0, i = 1...N$$

We let $\boldsymbol{\mu}$ denote the entire vector of quasi-Fermi levels, F_i denote ΔJ_i and $\boldsymbol{F}$ denote the entire set of functions $\Delta J_{i=1...N}$, and expand the functions in the vicinity of $\boldsymbol{\mu}$ using Taylor series (neglecting second and higher order terms):

$$F_i(\boldsymbol{\mu} + \delta\boldsymbol{\mu}) = F_i(\boldsymbol{\mu}) + \sum_{j=1}^{j=N} \frac{\partial F_i}{\partial \mu_j} \delta\mu_j$$

The matrix of partial derivatives in equation above forms the Jacobian matrix $\boldsymbol{J}$:

$$J_{ij} \equiv \frac{\partial F_i}{\partial \mu_j}$$

And the equation can be written as:

$$\boldsymbol{J} \cdot \delta\boldsymbol{\mu} = -\boldsymbol{F}$$

and

$$\boldsymbol{\mu}_{new} = \boldsymbol{\mu}_{old} + \delta\boldsymbol{\mu}$$

Therefore, the Newton-Raphson method takes a set of quasi-Fermi levels as initial guess, approximate the set of current conservation equations with linear approximation, and produces a step $\delta\boldsymbol{\mu}$ along the function's fastest descendant direction towards its root. As one can see, one of the main challenges of solving Newton-Raphson equation is to evaluate the Jacobian efficiently. In this method, all the Jacobian elements can be evaluated analytically, and computation of Jacobian is no issue. So the main challenge in this method is to having a stable algorithm that ensures Newton-Raphson method produces good convergence behavior for a wide variety of cases. These challenges, along with solutions will be covered in the next section.

4. NUMERICAL CHALLENGES AND SOLUTIONS

So far, a multi-scale, numerically efficient method has been described to solving transport problem in LED. The method circumvented the need to compute expensive scattering energies self-consistently, by separating treatment of quantum wells and barriers as equilibrium and non-equilibrium (ballistic) regions, and adding phenomenological scattering ($i \cdot \eta$) in equilibrium regions to mimic all scatterings. Electron-hole interaction are described on a masoscopic scale by using the 'ABC' formalism. Physically quantities such as current and density are all balanced out to ensure current conservation throughout the device. However, such a multi-scale model still faces plenty of challenges modeling realistic LED structures. The three main challenges are: solving G^R with matrix inversion, properly resolving resonance states, and globally convergent behavior of Newton-Raphson equations.

4.1 Solving for G^R

One of the major computation bottleneck is solving the matrix inversion for the retarded-Green's function G^R. To see why, consider the quantum transport region of a typical LED structure as shown in figure 2.4, it is roughly 130 nms long, consisting of about 1000 atoms. In the tight-binding representation of such a device, we apply the 20-band model ($sp^3d^5s^*$ with spin-orbit coupling) to describe each atom. Therefore, the Hamiltonian matrix size is roughly 20,000 by 20,000. Solving for G^R would require inverting a matrix same size with the Hamiltonian. Since matrix inversion scales as $\mathcal{O}(n^{2\ 3})$, depending on the sparsity and algorithm used, to invert such a matrix in full would require anywhere between 400 million operations up to 8 trillion operations! What's more, to store such a single matrix would require 6 Gigabytes of memory!

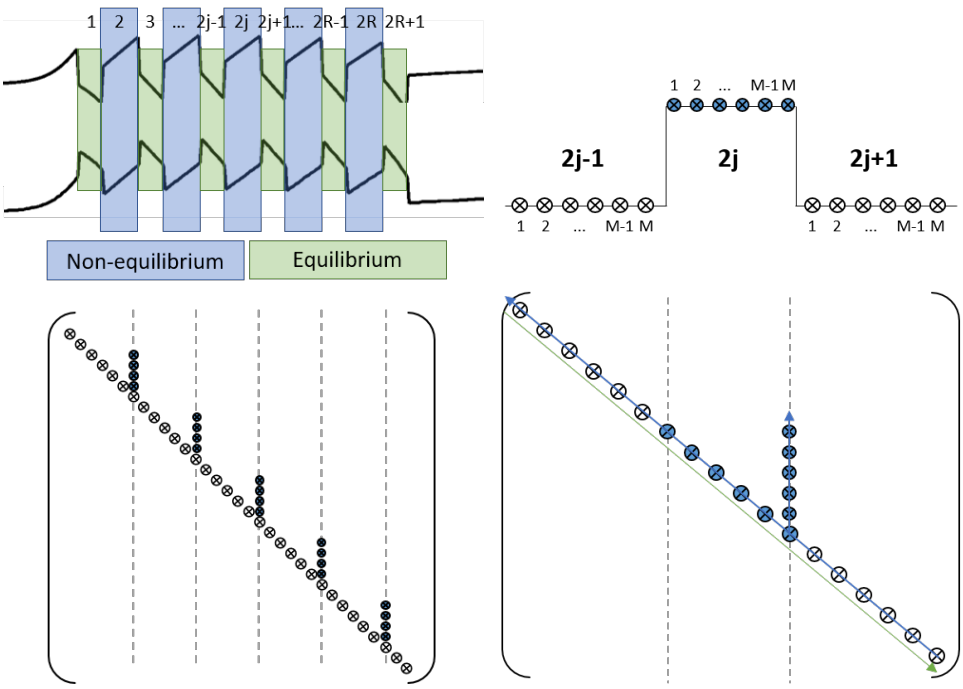

Fig. 4.1.: Schematic of structure treatment and G^R matrix sparsity distribution for (left) MQW (right) single barrier in an LED

4.1.1 Generalized Recursive Green Function Algorithm

The recursive Green function algorithm is the most computationally efficient algorithm for solving the G^R matrix for equilibrium-non-equilibrium type of transport problem [17]. It has been successfully applied to modeling of resonance tunneling diodes (RTD) structures in the past and well-matching experimental results [12] [16]. However, for the LED structure modeled in this thesis, the algorithm cannot be directly applied due to the structure being much more complicated than RTD. Therefore instead, the algorithm needs to be enhanced.

The figure above shows a schematic of an LED, divided up into equilibrium and non-equilibrium regions as described with the model. Assume there are N total number of non-equilibrium regions, with index 2...2N; and N+1 total number of equilibrium regions, with index 1...2N+1, as marked in the figure. For simplicity, assume

every region has same number of lattice sites, denoted as M. Then we focus on a single barrier with index 2j, along with its left and right neighbors, which are reservoirs with index 2j-1 and 2j+1. (figure below) For this particular barrier, its density is calculated from the $G^<$ equation shown previously (link), where only the diagonal elements $G^<_{i,i}$ are needed. And they are calculated as follows:

$$-i \cdot G^{2j,<}_{i,i} = f^{2j-1} \cdot (A^{2j}_{i,i} - A^{2j,r}_{i,i}) + f^{2j+1} \cdot A^{2j,r}_{i,i}$$

where

$$A^{2j}_{i,i} = i(G^{2j,R}_{i,i} - G^{2j,A}_{i,i})$$

$$A^{2j,r}_{i,i} = G^{2j,R}_{i,M} \Gamma^{2j+1}_{M,M} G^{2j,A}_{M,i}$$

It is easy to see that in order to calculate the charge density in a non-equilibrium region, only the diagonal elements $G^{2j,R}_{i,i}$ and the right-most column $G^{2j,R}_{i,M}$ are required in the retarded Green's function. To calculate the density in the equilibrium region 2j-1, as shown in equation (link), only the diagonal elements of A^{2j-1}, and therefore, of $G^{2j-1,R}$ are needed. NEGF current is calculated only in the non-equilibrium regions. For region 2j, the transmission coefficient is calculated as:

$$T^{2j} = Tr\{\Gamma^{2j-1} G^{2j,R} \Gamma^{2j+1} G^{2j,A}\}$$

Because Γ^{2j-1} has only one non-zero element at (1,1) and Γ^{2j+1} only has one non-zero element at (M,M), it is easy to show that to calculate the transmission, thus the ballistic current, only a single element $G^{2j,R}_{1,M}$ is needed, which is the top element in the right-most column of retarded Green's function. The recombination current is calculated with the 'ABC' formulas, knowing the equilibrium densities. No additional elements in the Green's function are needed, besides ones for the equilibrium density. Figure 4.1 shows the sparsity pattern of the G^R matrix for the entire device, including only elements that are necessary. The diagonal elements are needed throughout the device, in both equilibrium and non-equilibrium regions, and only a few segments are needed for the off diagonal elements, that covers only the non-equilibrium regions.

The generalized recursive Green function algorithm, therefore, calculates only those elements.

Similar to the RGF algorithm, this algorithm a two-step process. The first step, also know as the forward process is to calculate g^{LR}, the left-connected Green's function:

$$g^{LR}_{i,i} = (E - H_{i,i} - t_{i,i-1} g^{R}_{i-1,i-1} t_{t-1,i})^{-1}$$

For the first element (connected to the emitter contact), the third term in the bracket becomes the non-zero element in the self-energy of the emitter contact, which can be calculated analytically assuming semi-infinite contact. The second step, or the backward process calculates the G^R, the exact Green's function that is connected to both left and right contacts. The diagonal elements:

$$G^{R}_{i,i} = g^{LR}_{i,i} + g^{LR}_{i,i} t_{i+1,i} G^{R}_{i+1,i+1} t_{i,i+1} g^{LR}_{i,i}$$

The off diagonal elements at the K-th column:

$$G^{R}_{i,K} = g^{LR}_{i,i} (-t_{i+1,i}) G^{R}_{i+1,K}$$

For the first element (connected to the collector contact), $G^R_{N,N}$ is calculated as

$$G^{R}_{i,i} = (E - H_{i,i} - t_{i,i-1} g^{R}_{i-1,i-1} t_{t-1,i} - \sigma^{R}_{D})^{-1}$$

where σ^R_D is the non-zero element in the self-energy of the collector contact.

4.1.2 The impact of g-RGF algorithm

The previously described numerical load of solving the retarded Green's function is dramatically reduced with the g-RGF algorithm. (show comparison with a table) Solving full inversion of the 20,000 by 20,000 matrix would require at least 400 million (or n^2) operations. With the new algorithm, the operation is reduced to 400k (or $1000 \cdot 20^2$), or a factor of 3,000 improvement. To store the partially inverted matrix would require only 1.9 MB of memory, just a fraction of what is needed for storing the full matrix.

4.2 Integral Resolution

One of the key ingredients of successful simulation of resonance structure was proper resolution and numerical integration over energy and the transverse momentum [35] [36]. As for modeling LED, properly locating and resolving resonance states was in particular a challenge. Calculating common quantities of interest, such as density and current involves integration of energy-resolved data (DOS and transmission) over a significant energy range (a few hundred meVs). Consider a resonance state of 1 meV width, in order to have 10 points resolution within this resonance state, one would need 10000 points for an energy range of 1 eV. This is both impractical and unnecessary, because most of the "smooth regions" do not need such resolution. To make matters worse, for hole dominant devices, the location of hole resonances with respect to the band edge is dependent on the transverse momentum. Therefore, to simulate holes, one would need a different energy mesh for every k point. This session describes a systematic way that efficiently addresses the integration issue.

4.2.1 Energy Integration

Resonance Locating

The energy integration strategy is illustrated in fig. 4.2. We start with a rough set of data (left) with many resonances. Although the peak locations are unknown, the peak broadenings are relatively predictable. Since in our model, the scattering mechanisms are mimicked with optical potential ($i\eta$), and the value of the optical potential are constant and predetermined, the minimum broadening of each resonance states would be 2 times the imaginary part of optical potential, and this is the maximum resolution the system needs. Therefore, the remaining job is to find the location of resonance peaks. An elegant way of solving this issue is to solve the following nonlinear equation [37]:

$$f(E) = det(E - H - \Sigma(E)) = 0$$

Since the transmission resonances corresponding to the the zeros of the retarded-Green's function, the problem of resonance finding is equivalent to solving the non-linear equation with an eigenvalue problem. Efficient eigenvalue algorithms such SINL (shift and invert nonsymmetric Lanczos algorithm) can be applied to solving this issue, since one is only interested in a small subspace of eigenvalues around the energy window of interest. However, for the LED structure, it was found out that such resonance-finding algorithm is still far too expensive, mainly for two reasons:

1. Device size: device Hamiltonian matrix is roughly 20,000x20,000 in size.

2. Number of resonances: there are six quantum wells in both conduction and valance band, and a few hundred total resonances peaks. That means the non-linear equations needs to be solved thousands of times.

Instead, a much simple way of resonance-locating was applied, which is to directly locate them on the data by taking the first- and second-derivative on the data (such as $J(E)$). This way of resonance finding almost takes no extra computation, however, has two drawbacks:

1. Accuracy: since the peaks in the initial data set are not well-solved, using the peak as the resonance is not accurate. However, as one can see, this algorithm is an adaptive method, and the peak locations will significantly improve with iterations.

2. Missing broadening information: the resonance-finder not only gets the peaks location by solving for the zeros in the GR. Since the GR matrix is an complex quantity, the energy at resonances also contains an imaginary part, which represents the peak broadening. Such information is not present in this algorithm. Yet, as explained earlier, the minimum broadening, thus the resolution requirement is known, and we can live with not knowing about the exact broadening of individual states with small penalty.

Grid Construction

After the peaks of the resonances are located, the next step is to construct a smooth function $f(E)$ (mid figure) using the peaks $(E_1, E_2...E_n)$ and the minimum broadening (Γ). $f(E)$ is a superposition of a series of Lorentian functions, each one centering at a resonance peak, with broadening of (Γ):

$$f(E) = L(E_1, \Gamma) + L(E_2, \Gamma) + ... + L(E_n, \Gamma)$$

where

$$L(E_n, \Gamma) = \frac{1}{\pi} \frac{0.5\Gamma}{(E - E_n)^2 + (0.5\Gamma)^2}$$

The integral $I(E)$ of function $f(E)$ can be analytically solved (see fig.4.2). Afterwards, divide $I(E)$ into even steps $I(E_1^{'}), I(E_2^{'}, ..., I(E_N^{'}$, and the energy points corresponding to the even steps on $I(E)$ are points on the new grid $E_1^{'}, E_2^{'}, ..., E_N^{'}$.

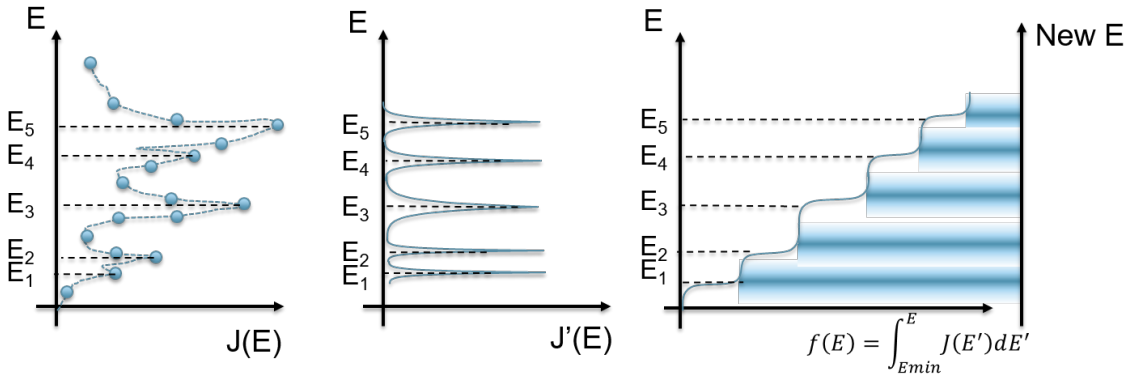

Fig. 4.2.: Energy-mesh refinement process: (left) coarsely sampled energy-resolved data to start with (mid) data is fitted with a set of Lorentzian functions, each centered around a peak. (right) integral function is computed for the fitted data, then a new mesh is created by projecting on to the integral function at equal integral steps, then extracting the energy values corresponding to these steps.

Figure 4.3 shows the whole energy refinement process. One thing to point out is in device simulation, there are multiple quantities that need to be well-resolved, two of the most important ones are $J(E)$, the energy-resolved current and $n(E)$, the

energy-resolved density. Therefore, the energy mesh algorithm in the model offers a range of refinement target choices, including transmission, current, density, DOS and local Fermi levels, or any combination of these. Also, because of the multi-scale model that divides up different regions, these quantities are locally defined when being used as mesh-refinement target, so to ensure local resonances are all resolved.

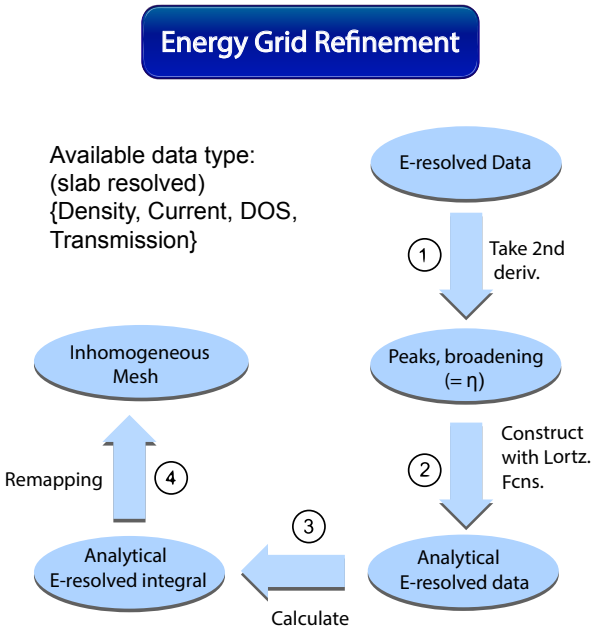

Fig. 4.3.: Flow chart of the energy-grid refinement process.

Below is an example to illustrate the energy-refinement process. The example structure consists of $InGaN/\mathbf{GaN}/InGaN/\mathbf{GaN}/InGaN$ (bold font indicates barrier). The two barriers and the middle QW are lowly doped, the left lead is n-type and the right lead is p-type. 20 band tight-binding model is applied ($sp^3d^5s^*withSO$). The two leads and the QW are assumed to be in equilibrium. Figure **??** shows the electron density w.r.t position and energy. Color indicates the magnitude of density. By design, this test structure contains multiple resonance states, as indicated by "dark puddles" in the figure. Figure **??** shows the energy-resolved current $J(E)$ using the meshing-algorithm describe earlier, and all the resonance peaks are well-

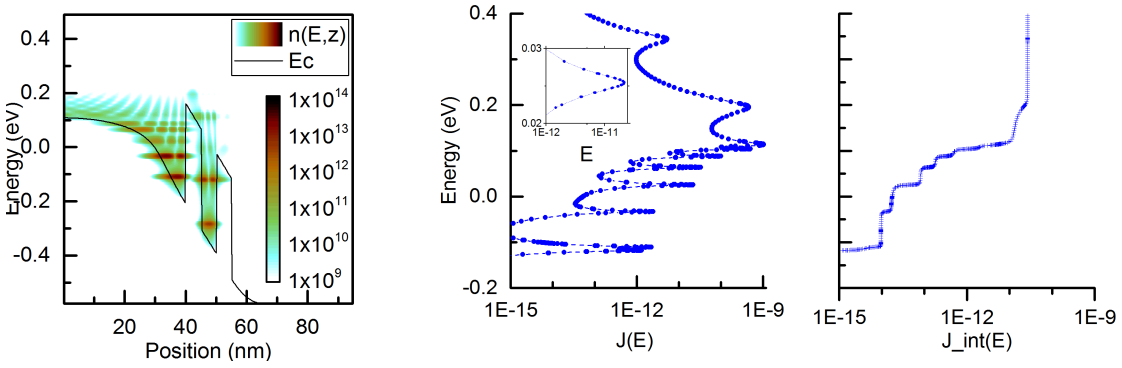

Fig. 4.4.: (left) Electron distribution in space and energy. (mid) Energy-resolved current through the first barrier. (right) Integration of current as a function of energy. All the peaks in the current are well resolved.

resolved, with a few extra points centering around the peak. The right figure shows the energy-integrated current as function of energy:

$$J_{int}(E) \sim \int_{E_{min}}^{E} J(E^{'})dE^{'}$$

Each step in the curve indicates the energy at which current is contributed, and all the steps are well-resolved, resulting in a smooth integral throughout the entire energy range.

4.2.2 Momentum Integration

Another level of integration is the momentum in the transverse plane $k_{||}$. As mentioned earlier, the hole bandstructure are highly anisotropic, which requires independent energy grid for different $k_{||}$. Yet what is often neglected is that different $k_{||}$ also have different contribution to current. Consider the current in the device:

$$J \propto \int dE \int kdkT(E,k)[f_L(E) - f_R(E)] = \int dE \int kdkJ(E,k)$$

where k is the transverse momentum, E is total energy, T is transmission coefficient and $f_{L/R}$ are Fermi functions in the left/right contact. This is assuming weak momentum angle dependence and therefore one can apply cylindrical integration for the

transverse momentum. One can separate out the energy dependence by performing integration over energy first:

$$J(k) = \int dE J(E,k)$$

such that

$$J \propto \int dk \cdot kJ(k)$$

The physical meaning of $J(k)$ is the current contribution at each momentum, and one can see that it is strongly dependent on the quantity $kJ(k)$. A good k mesh should refine around the range of k which contributes most to the physical results, and $kJ(k)$ is a good metric for measuring contribution of individual momentum.

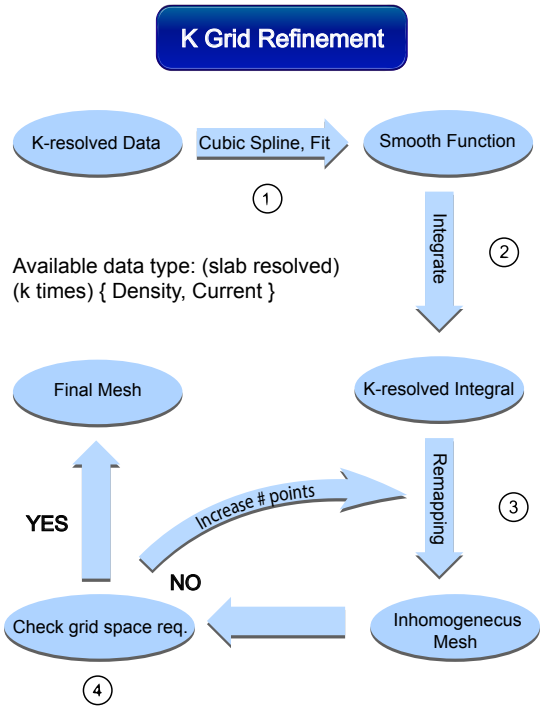

Fig. 4.5.: Flow chart of the k-grid refinement process.

k Mesh Refinement Process

Figure 4.5 outlines for the k-mesh refinement process. The grid construction algorithm is quite similar to the one for energy mesh. However, the main difference is that we can no longer approximate the grid peaks with a Lorentzian function, because we do not know beforehand what each quantity's k distribution looks like. Therefore, the first step is to construct a smooth function from the k-resolved data, such that its integral is also a smooth function. Cubic spline interpolation is applied to fitting the data. After the smooth function and its integral is obtained, the same remapping process used for the energy-mesh can be applied here. Because the k mesh usually contains far fewer points compared to the energy mesh, an extra step is applied to check the minimum grid space in the k mesh, and adding on more k points until it is fulfilled.

Sample results

The same previous example is used here to illustrate the k-mesh refinement process. Figure 4.6 shows the momentum dependence of current $J(k)$, and the major of current is carried by charge carriers with low in-plane momentum, as is clearly shown in the figure. The left figure show adaptive process of k-mesh generation, starting with a homogeneous mesh (blue curve), and eventually the k-mesh skews towards the lower part of $|k|$, to account for its relative higher importance in terms of current carrying. Figure 4.6 shows the $kJ(k)$ as function of $|k|$, whose importance was discussed earlier. Since $J(k)$ is an decreasing function of $|k|$ and $|k|$ itself is an increasing function of $|k|$, $kJ(k)$ will have a maximum that is away from Γ point ($|k| = 0$). As a result, the adaptive refinement process will place more emphasis at k range close to the peak.

Figure 4.7 shows the $J(k)$ for holes. In contrast to electron case, the hole current peaks at momentum near, but not exactly at the Γ point. As a result, $kJ(k)$ has

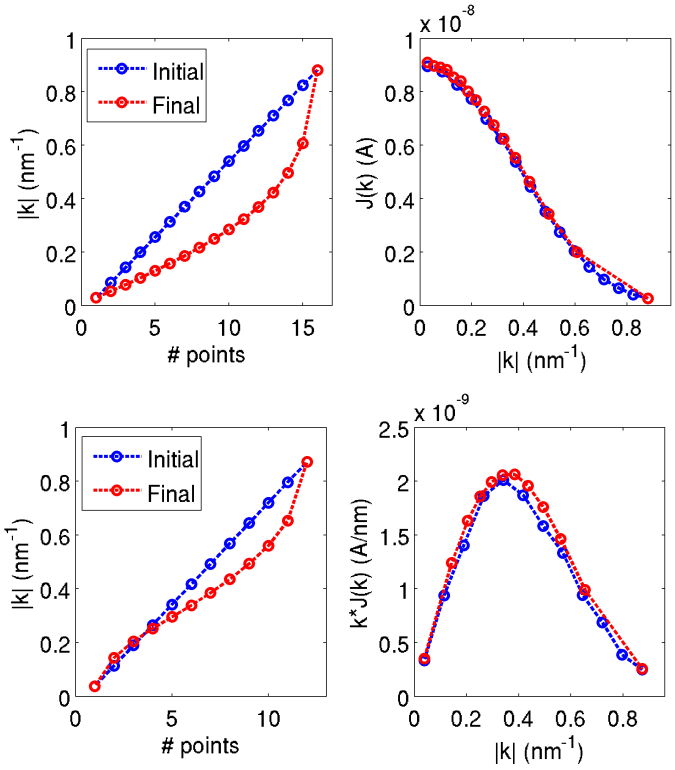

Fig. 4.6.: Refined k-mesh for electrons, using (upper) $J(k)$ and (lower) $k \cdot J(k)$ as refinement targets. The $J(k)$ decays with increasing momentum, resulting in a peak in $k \cdot J(k)$.The k-mesh refinement capture the peaks in both cases.

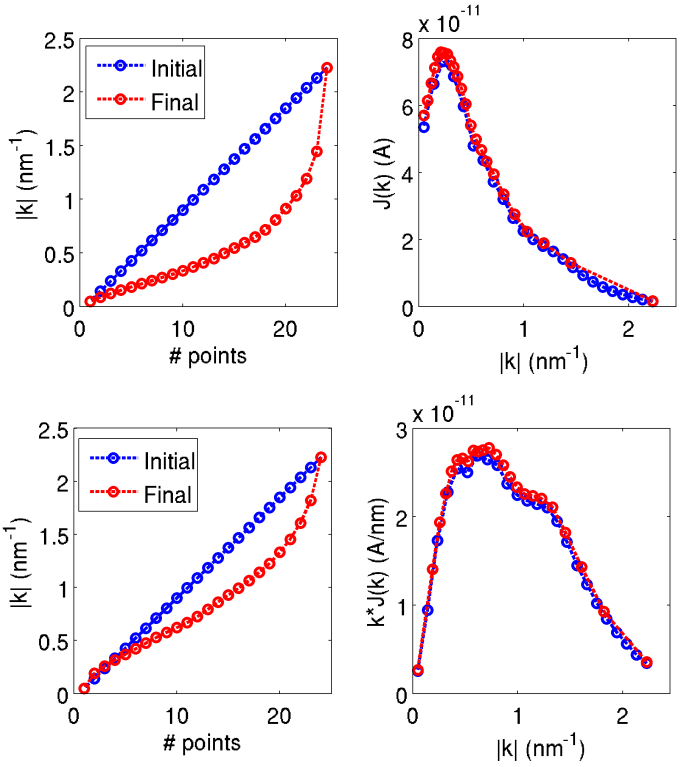

Fig. 4.7.: Refined k-mesh for electrons, using (upper) $J(k)$ and (lower) $k \cdot J(k)$ as refinement targets. The $J(k)$ do not decay with k monotonically but rather has a peak. As a result the $k \cdot J(k)$ function contains two bumps. The k-mesh refinement capture the peaks in both cases.

a double bump as shown in figure (kjkhole). The adaptive process automatically accounts for the double bump feature and places more points around those peaks.

5. RESULTS AND DISCUSSION

5.1 Simulation of a Commercial LED Device and Comparison with Experiment

The model is applied to simulate a commercial LED device. Figure 1 shows the conduction and valance band diagram of the LED active region and breakdown of the device structure: it consists of (from left to right): 15.5nm n-GaN layer doped at $1 \times 10^{18} cm^{-3}$, 20.7nm n-type GaN layer doped at $4 \times 10^{18} cm^{-3}$, 6 repeating $In_{0.13}Ga_{0.87}N/GaN$ (3.1nm/4.6nm) quantum wells unintentionally n-doped at $2 \times 10^{15} cm^{-3}$ and a 24.8nm $Al_{0.12}Ga_{0.87}N$ electron-blocking layer followed by a 15.5nm GaN layer, both p-doped at 4×10^{19} cm-3. The detailed material composition, dimension and doping are marked on Fig. 5.1.

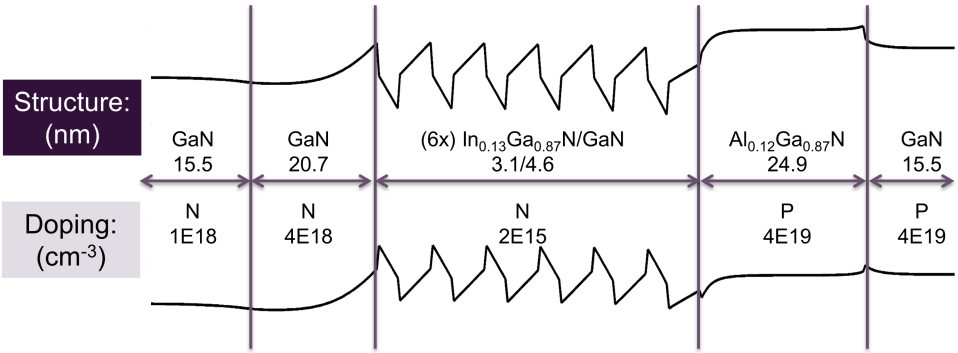

Fig. 5.1.: Device structure and doping profile for a prototypical commercial LED active region.

As for the simulation parameters, the scattering parameter η is chosen so that broadening of quantum states match the experimentally measured [27] photoluminescent (PL) emission width of $\sim 100meV$. The exponential decay length (λ) is 50 meV, chosen to match experiment in Ref. [28]. Parameters A, B, C are chosen

from fitting to experimental IQE data provided by Lumileds. The parameters are summarized in the following table.

Parameter (unit)	Value
$\eta(eV)$	0.1
$\lambda(eV)$	0.05
$A(s^{-1})$	2×10^{6}
$B(cm^{3}s^{-1})$	1.48×10^{-11}
$C(cm^{6}s^{-1})$	1.6×10^{-30}

Table 5.1: Summary of default simulation parameters used for device in Fig. 5.1

The colormap in Fig. 5.2 shows the position and energy distributed charge densities calculated from equation 3.7 with $\boldsymbol{k} = 0$. Setting $\boldsymbol{k} = 0$ enables visualization of the spatial and energetic distribution of the quantum states without loss of generality. Also shown in the same figure are the contour lines of density of states (DOS), which map out the available states for electrons to fill in. The bulk-based conduction and valence band edges are also included in the plot as a guide to the eye. Given typical InGaN LED barrier widths is in the range of a few nanometers, the states in the QWs are broadened and coupled to each other. Carriers are distributed in energy across these broadened states. A distinct quantum interference pattern is observed throughout the energy range of interest, and as a result, no distinction should be made between 'classical' and 'quantum' regions. Coupling between continuum and discrete states occurs naturally. The broadening of states is a direct manifestation of finite carrier lifetime, due to frequent scattering events and coupling to the open leads. Stronger scattering, mimicked by a higher value of η, leads to larger state broadening and more cross-coupling between various states, as shown by comparison of Fig. 5.2 (a) and (b). Therefore, carrier transport occurs through a complex, extended structure, and is directly influenced by the overall quantum-mechanical properties of the system. The electrons fill all the QW ground states and partially fill the excited

states. The hole states are spaced much more closely in energy due to their larger effective mass. The heavy and light hole bands are explicitly coupled in this model due to breaking of translational symmetry. The hole charge density spreads in energy over multiple confined quantum states. The local band edges are also shown in figure. Alloy values are interpolated from the bulk constituents [38]. In a semiclassical model, carrier transport is strictly limited by the energy position of the band edges, allowing transport to only occur above E_C and below E_V. However, at this nanoscale new quantum states emerge that render local bulk band edges almost irrelevant.

Figure 5.3 plots the position-resolved electron and hole density (integrated over all energy and momentum) throughout the device. Band profile and local Fermi levels are plotted as well. For electrons, the Fermi level drop is very small (25 meV) across the sequence of 6 QWs: the electrons are well transported across the finite superlattice. In contrast the total Fermi level drop for the holes is 176 meV. Clearly the holes are not well transported across the superlattice and large Fermi level drops are needed to support required total current conservation. As a result, electron density is well spread throughout the device, while the hole density distribution is heavily skewed towards the p-side QWs.

Figure 5.4(a) compares the I-V characteristics between simulated results and experimental data. The experimental structure contains series resistance in the order of a few $m\Omega \cdot cm^2$, mainly induced by the contacts and spreading layer. The simulations presented here use a fitted series resistance of 2.0 $m\Omega \cdot cm^2$. A well-known challenge of traditional semiclassical-based model is the difficulty in obtaining a good I-V match with experiment. Typically a much larger bias voltage is needed to model the same current as measured in experiments. Methods such as including large alloy fluctuation in the simulation [39] have been proposed as remedies. However, the unrealistic turn-on voltage under normal conditions is an indication of missing critical transport physics in semiclassical models. Therefore, a key achievement here is that the turn-on voltage of this complex device is modeled correctly with our approach. LEDs typically operate under a variety of temperatures, and a realistic operating condition [40]

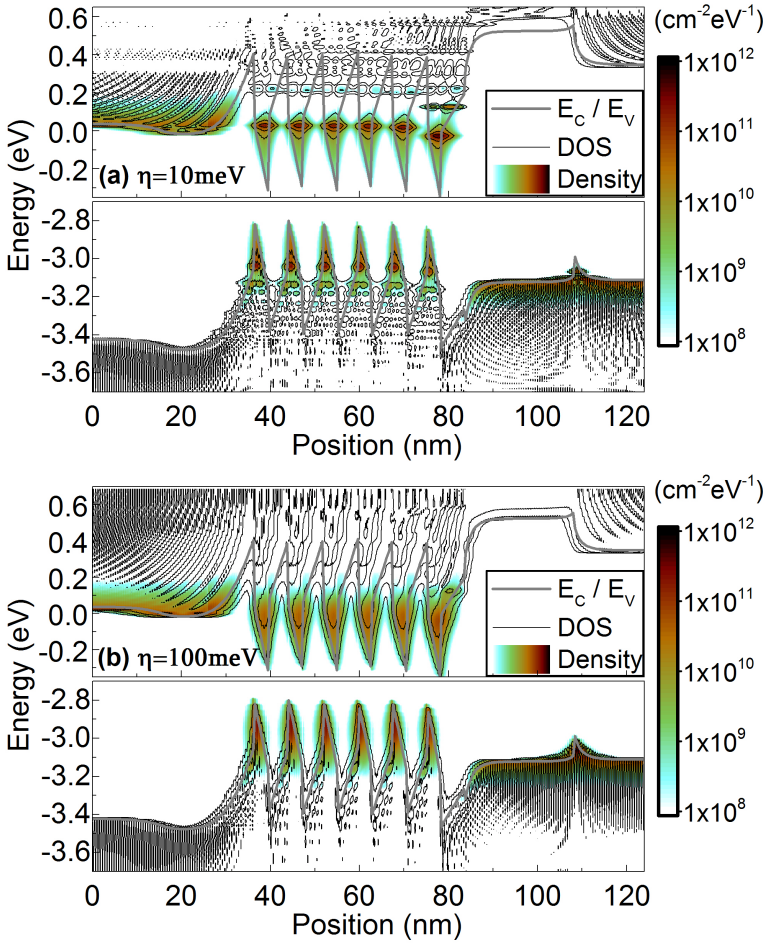

Fig. 5.2.: Energy-resolved electron, hole density of states (contour lines) filled with electrons and holes (color contours). The bulk-based conduction and valence band edges serve as a guide to the eye and only enter the calaculation in the definition of the empirical scattering strength . (a) $\eta = 10meV$ is a typical broadening in GaAs and InP based devices. (b) $\eta = 100meV$ is a broadening that corresponds to experimental optical linewidth measurements.

for LEDs is above 85 $°C$. Figure 4(b) presents the I-V characteristics in a temperature sensitivity study in the range 320-400K. As the carrier temperature inside the LED increases, the I-V curves shifts to a lower voltage. Higher temperatures lead to broadened energy-distribution of carriers in the quantum wells, leading to higher

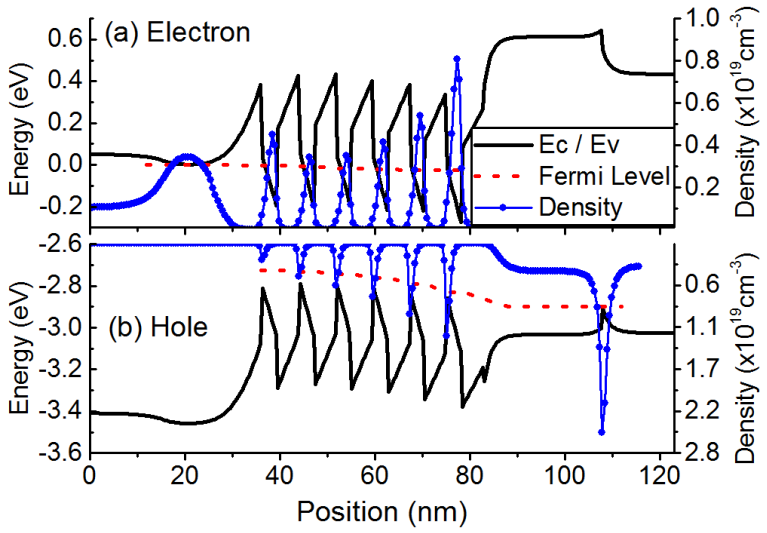

Fig. 5.3.: (a) Conduction band profile with electron density and (b) valance band profile with hole density at $2.9V$ bias. Note that local Fermi levels (red dashed lines) are only defined in the leads and QWs, where those regions are treated as thermal equilibrium, the continued lines across barriers are meant to guide the eye. The Fermi level drops across the device is $25meV$ for electrons and $176meV$ for holes.

thermionic emission at the same voltage. As a result, higher temperatures leads to earlier device turn on. The simulation result shows the LED forward voltage at 20 A/cm^2 varying between 2.75 to $2.92V$ under a realistic temperature variation of $80K$. An assumed electron temperature of $360K$ results in the best match with our experimental data, which in turn agrees reasonably well with experimental evidence [40] that the electron temperature in LEDs is typically above 85 $°C$.

Figure 5 shows the photon emission current density from each quantum well at different drive currents. The main observation is that the largest contribution to the photon current occurs at the last QW nearest the p-GaN ($\sim 24\%$ at $137A/cm^{-2}$). In contrast, each of the middle 4 QWs contributes about 15% to the photon current and the QW at the n-GaN end contributes only 12%. This is in good general agreement with previous experiments [41] [42], although the experimentally observed emission profile is even more skewed towards the last two QWs. Such a phenomenon is an

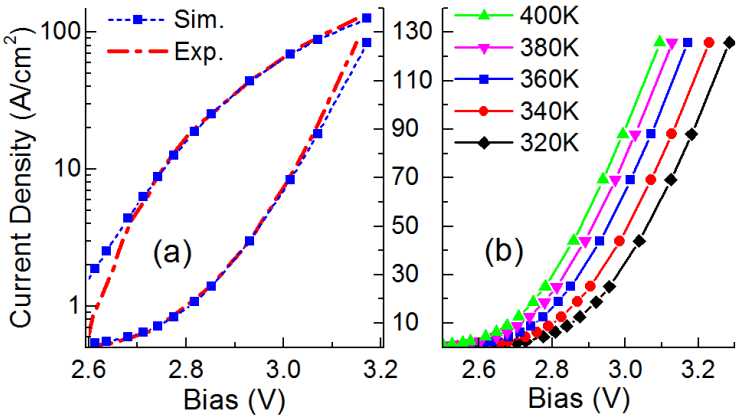

Fig. 5.4.: (a) I-V comparison with experiment shows good quantitative match on a linear and a log scale. A $2.0m\Omega \cdot cm^2$ series resistance, $360K$ electron temperature, and $100meV$ spectral broadening in the quantum wells are assumed in the simulation. (b) I-V Simulations for various temperatures ranging from 320 to $400K$ result in a variation of $\sim 170meV$ in turn-on voltage.

indication of non-uniform carrier distribution among QWs at high bias, and has a negative impact on the LEDs quantum efficiency.

Figure 5.6 plots the percentage of the radiative recombination versus different drive current (square markers). This ratio is equivalent to the internal quantum efficiency (IQE). The simulation data predicts a droop roll-over at around 6 A/cm^2. The relative importance of leakage (overflow of carriers out of the entire active region) vs. Auger to efficiency droop has historically been a debating point. Conclusions from traditional modeling methods tend to vary depending on simulation condition and choice of parameters [43], although recent experiment identifies Auger as the main mechanism responsible for efficiency droop in InGaN/GaN LED [44], the contribution of leakage current is much less than 1%, while Auger recombination accounts for majority of IQE droop, in agreement with experimental findings [44].

Figure 5.7 shows the energy-resolved current density at zero in-plane momentum ($J(E, \boldsymbol{k} = 0)$) across each barrier. For comparison, two cases with different scattering

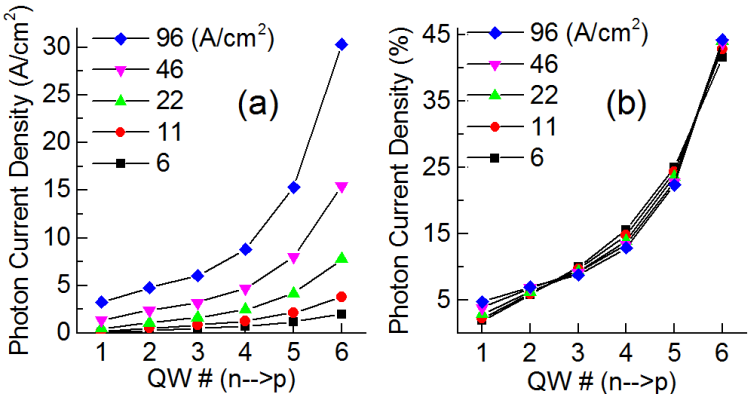

Fig. 5.5.: Photon current density profile for each quantum well with the total current density as a parameter. (a) absolute photo emission current, (b) relative photon emission current. At high current densities the QW closest to the p-side dominates the light emission, which matches experimental observation.

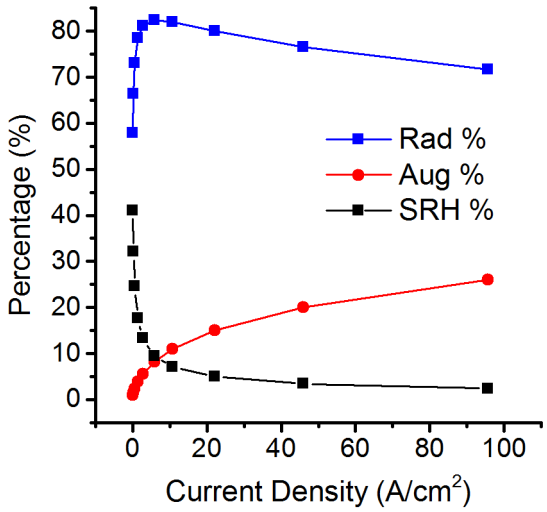

Fig. 5.6.: Relative contribution of radiative recombination, Auger recombination, and carrier leakage to the total current density.

rates ($\eta = 10$ and $100meV$) were plotted together. At $\eta = 10$, corresponding to the lower scattering rate, the current density exhibits distinct (and broadened) reso-

nances, and the current contribution from various confined states can be identified. At $\eta = 100meV$, which is a broadening that corresponds to experimental optical linewidth measurements, the various states become cross-coupled and the current density curves become smooth. $\eta = 100meV$ leads to the quantitative agreement with experimental data shown later. What is more noteworthy, however, is that vast majority of current contribution comes from tunneling under the barriers. The barrier height is marked with a gray dashed line in each figure, and from that one can see, the bulk of current occurs below the barrier both in the conduction band and in the valance band. After current contributions from all in-plane momentum are summed up, almost all the electron current (99%) and the majority of hole current (75%) are contributed through tunneling. Tunneling, especially for holes, are typically ignored or treated as a patched-in effect in the traditional semiclassical-based tools. Tunneling is naturally included, and treated in the same footing with thermionic emission in our quantum-based tool.

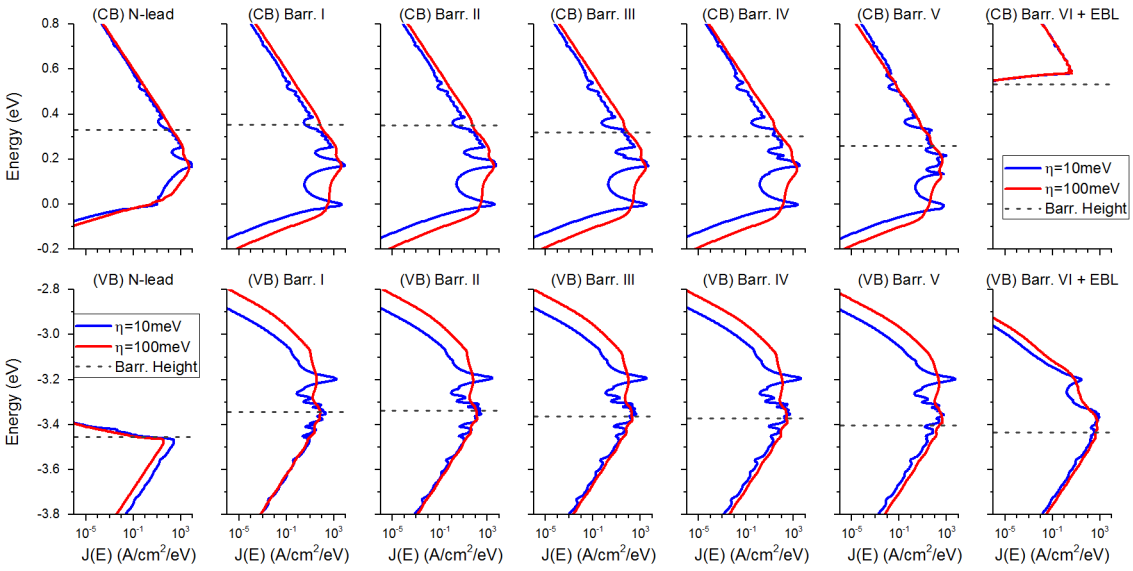

Fig. 5.7.: Energy-resolved current density at zero in-plane momentum $J(E, \boldsymbol{k} = 0)$ for each individual barrier (indexing follows Fig. (ref)). Barrier heights are marked with grey dashed lines in each figure. $\eta = 10$ vs. $100meV$ are plotted together for comparison.

5.2 Trend Analysis w.r.t. Barrier Width

Simulation results on a prototypical LED device in the last section has shown excellent agreement with experimental I-V (Fig. 5.4) and revealed interesting device characteristics such as the p-side dominated light emission. Now we apply the same simulation technique to study several device performance trends.

The first trend to study is how the device performance is influenced by barrier thickness. In this simulation study, the device structure in the previous section is simulated with different GaN barrier thicknesses: 3.6, 4.7, 5.7, 7.3 nm. All 6 barriers maintain the same thickness.

Figure 5.8 compares the band diagrams and density for different barrier thicknesses W (only 3 shown to save space). In each figure, the electrons are on top and holes are at the bottom. Intuitively, increasing W should reduce overall carrier density, since

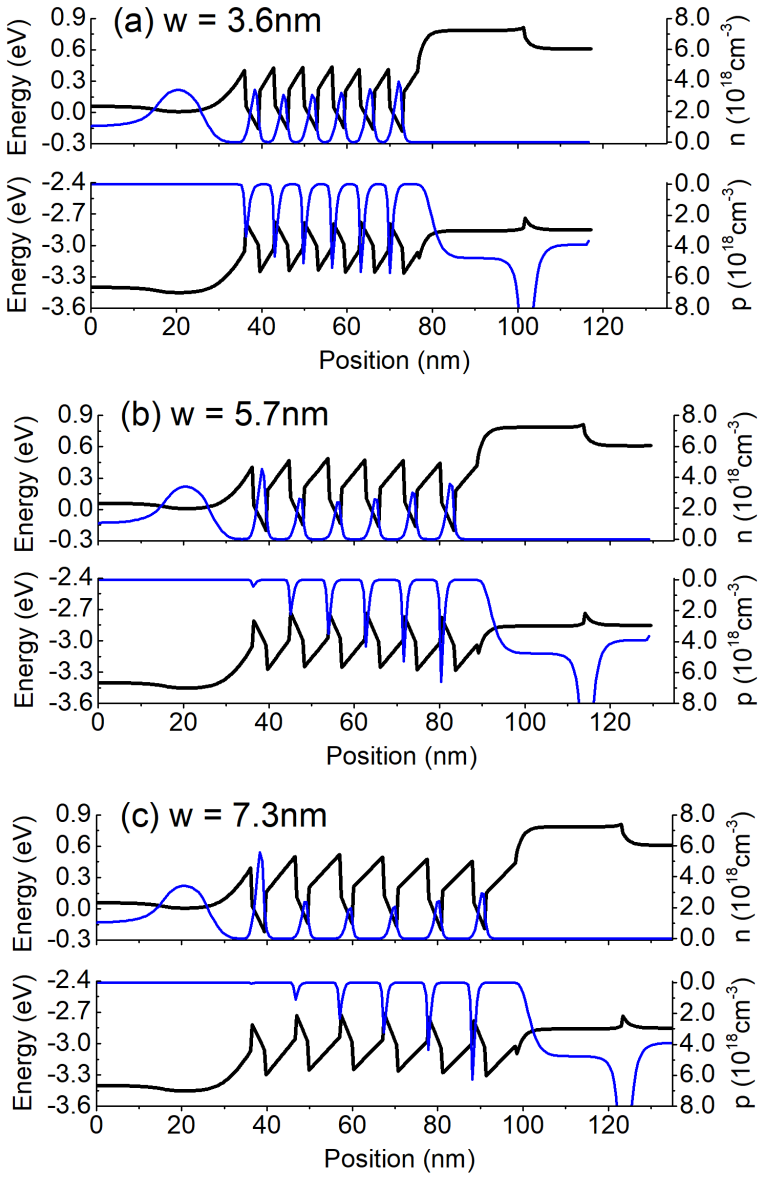

Fig. 5.8.: Band profile and carrier density for different barrier widths (a)3.6nm (b)5.7nm (c)7.3nm. For each plot, the top half is conduction band (black line) with electron density (blue line) and bottom half is valance band (black line) with hole density (blue line)

barriers impedes carrier flow. This is indeed what the model produced. However, what's more interesting is that thicker barriers do not reduce electron and hole density

in the same way. For electrons, the density drops across all but the first QW (there is a slight increase in electron density in the first QW). The carriers among different QWs are more of less evenly distributed, and do not change for different W. For holes, W has a more significant effect on carrier distribution. Increasing W does not affect hole density in all QWs the same way. Concretely, it has much bigger impact on the left (n side) QWs compared to the right (p side) ones. Increasing W makes it significantly harder for holes to make it to the n-side, and as a result, making the overall hole density distribution much more skewed towards the p-side. This is again due to the fact that hole transport is much harder compared to electrons. This point is much better illustrated when looking at the density averaged over each QW shown in Fig. 5.9.

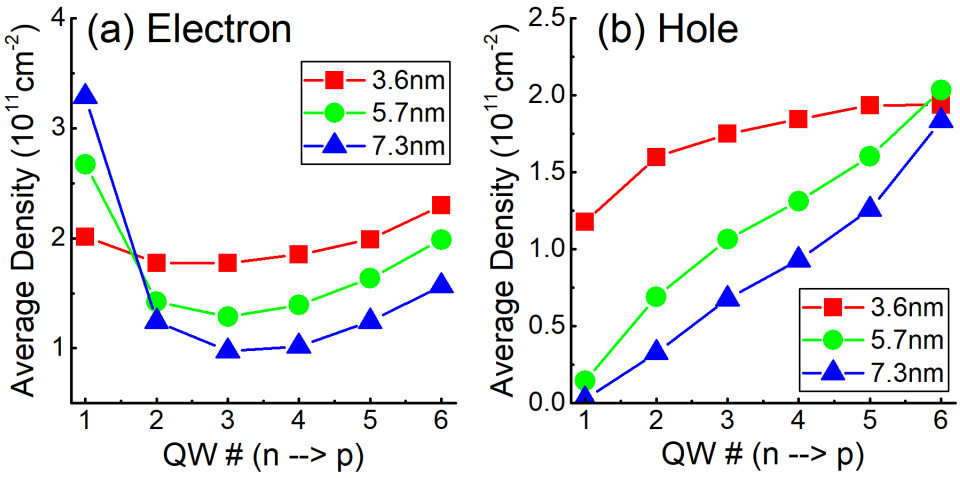

Fig. 5.9.: Average (a) electron (b) hole density in each QW for various barrier widths.

The density distribution among QWs has a direct impact on the emission profile. Figure 5.10 compares the photon emission current per QW for various device structures. The left subplot plots the magnitude of the photon current while the right one shows the percentage of photon emission. As barrier increases, the overall photon emission decreases, due to the reduction of overall charge density (both electrons and holes). However, emission profile (5.10(b)) become more skewed towards p-side as

barrier thickness increases, due to the more skewed density profile seen in Fig. 5.9(b) with larger barrier thickness.

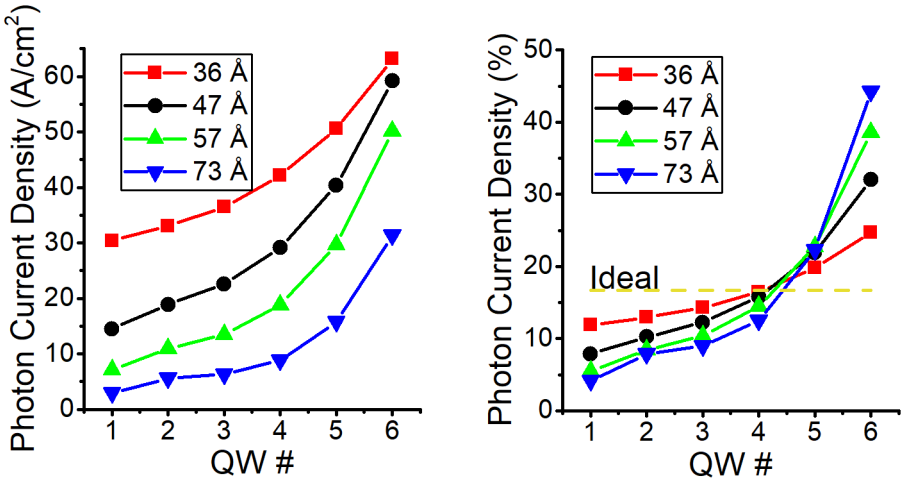

Fig. 5.10.: Photon current density per QW for LED with various W. Bias voltage fixed at 2.9V for all cases. (a) absolute photo emission current, (b) relative photon emission current.

Figure 5.11 compares the I-V and IQE for various barrier width. As barrier width increases, the I-V shifts to the right. This could be understood with the analogy of a simple level arm: as the device length increases (arm length), it requires a high bias (θ) to achieve the same amount of drive current (work). The degree of shift in I-V is more of less linear with the increase in barrier thickness.

Increasing W also has an impact on the IQE, as shown in Fig. 5.11(b). The peak value of IQE reduces slightly with larger W, the reduction at higher current is much more significant. Namely, the barrier width has a more pronounced impact on the degree of droop than peak efficiency. Thick barriers degrades the IQE, because it significantly reduces amount of hole making to the n-side, thus reducing the overall recombination efficiency, and the higher the bias, the more pronounced this effect becomes.

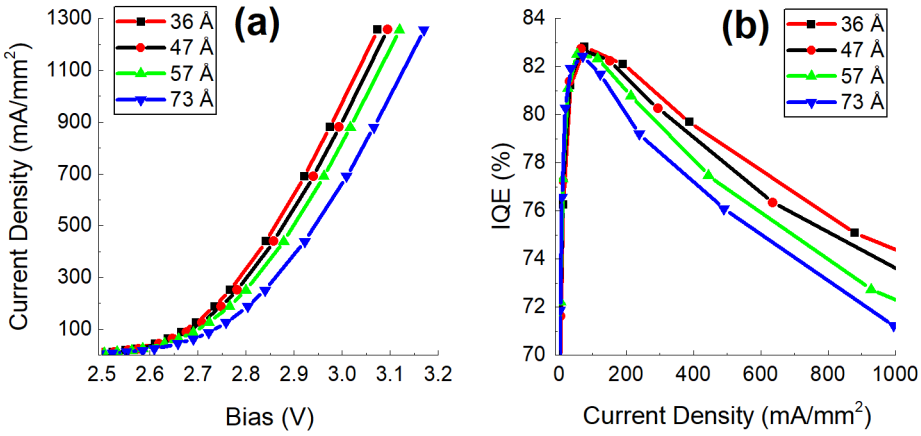

Fig. 5.11.: (a) I-V (b) Internal quantum efficiency for LED with various barrier widths.

5.3 Trend Analysis w.r.t. Al%

The electron block layer (EBL) is a key design component of the LED, and subject to many previous experimental and simulation studies. It is typically made of a thick layer of AlGaN, whose role is to prevent the electron leakage out of the active region.

Figure 5.12 compares the carrier density for various aluminum concentrations (Al%). The most obvious effect of a higher Al% is the rising of the EBL height. However, there are a couple of more subtle changes, which might have a bigger impact on device performance than EBL height, and are rarely discussed in the literature. The first effect is the increase of polarization field between the AlGaN/GaN interfaces, which leads to ① increase in the valance band dip at $x \approx$110nm, and ② increased height of the first hole barrier at $x \approx$80nm.

The direct effect of raised p-side hole barrier is reduction of hole density in the MQW, as can be seen from Fig. 5.12 and 5.13. However, the reduction of hole density also leads to the band profile the MQW region to shift upwards. This in turn leads to a reduction in the electron density.

Due to its impact on carrier density, varying Al% in the EBL has a big effect of I-V characteristics and efficiency. Figure 5.14 shows I-V and IQE for Al% ranging

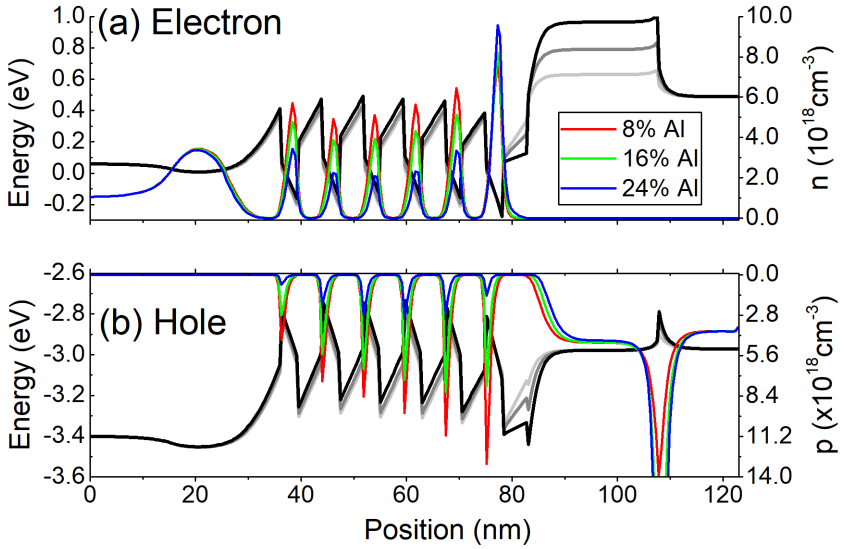

Fig. 5.12.: (a) Electron density with conduction band (b) Hole density with valance band for LED with various Al% in the EBL. Density is plotted with different colors for various Al%, and band diagrams are plotted in gray-black lines with increasing darkness for increasing Al%.

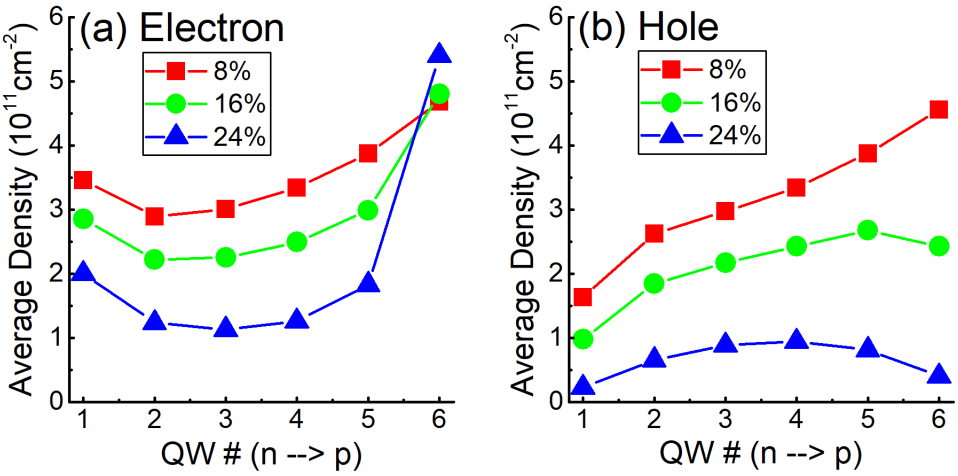

Fig. 5.13.: Average (a) electron (b) hole density in each QW for various Al%.

from 8 to 24%. Increasing the Al concentration shifts to the I-V to higher bias (right) range. This is because increasing Al% significantly impedes hole injection, which in

turn reduces electron density to maintain charge neutral, and as a result causes a dramatic reduction in overall recombination current.

In terms of efficiency, result shows with higher Al %, the overall IQE decreases, as seen from Fig. 5.14 (b). This is somewhat of a surprising finding, since one would normally expect increasing Al% to prevent electron leakage and improve efficiency. However, as Fig. 5.15 shows, higher Al% leads to higher proportion of Auger recombination in the total current. As for leakage, results from previous section has already shown leakage current was already negligible, thus increasing Al% beyond 12% has yields no improvement at all.

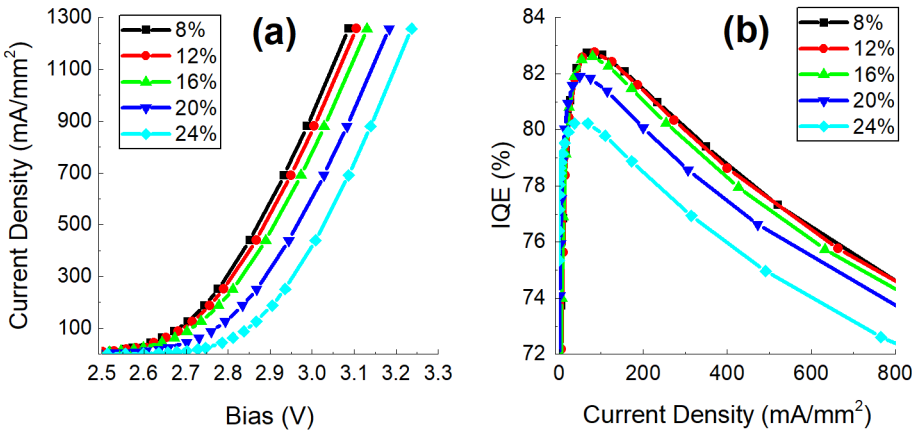

Fig. 5.14.: (a) I-V (b) Internal quantum efficiency for LED with various Al% in the EBL in the range between 8 and 24%.

The next question is: how much of impact does the EBL have at all in reducing leakage? Figure 5.16 compares the proportion of electron leakage current in terms of percentage of total electron current density for Al% between 0 and 8%. In the normal range of LED drive current, leakage has small contribution. Even without EBL (0% Al), the maximum amount of electron leakage current does not exceed 2% total current. Figure 5.17 shows I-V and IQE for Al% below 8% and they both vary very little.

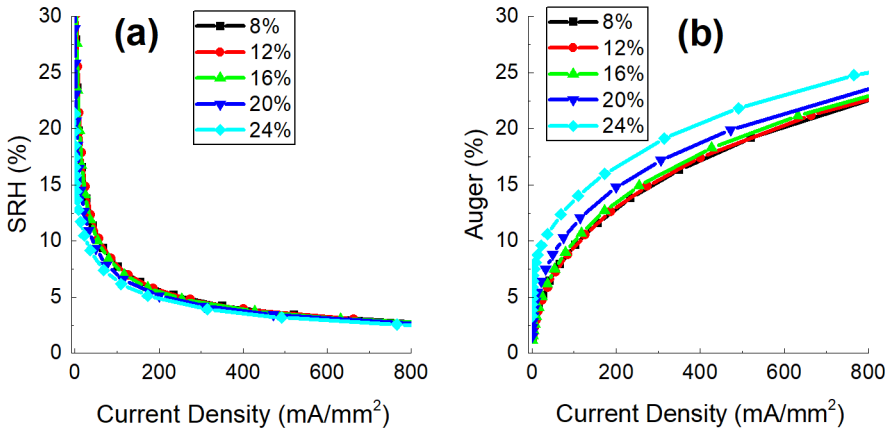

Fig. 5.15.: Percentage of (a) SRH (b) Auger recombination w.r.t the total current for LED with various Al% between 8 and 24%.

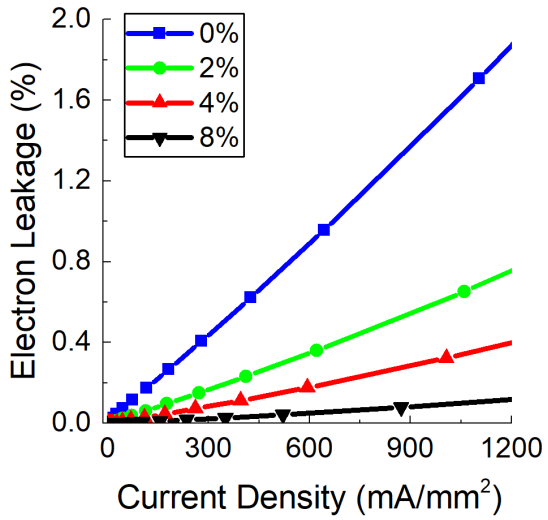

Fig. 5.16.: Percentage of electron leakage current w.r.t total electron current for 0-8% Al in the EBL.

Lastly, the I-V results for two different Al concentrations 12% and 24%, are compared with experiment and plotted in Fig. 5.18. The experimental device structures contain propitiatory designs, which was not assessable to me. However, I was told by

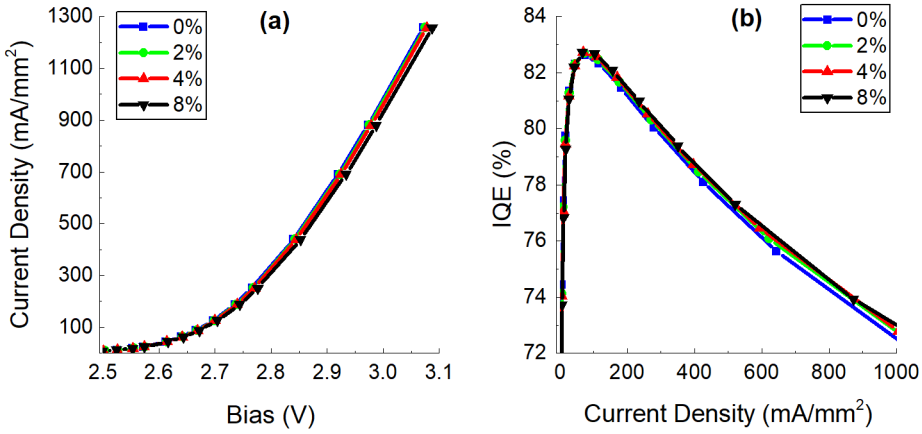

Fig. 5.17.: (a) I-V (b) Internal quantum efficiency for LED with various Al% in the EBL in the range from 0-8%.

collaborator from Lumileds that the devices were very similar to the simulated device structure, and the Al% in those structures are certain. Experimental result for 12% case contained 6 samples, and the average results with error bars were plotted. Despite the constrains, the simulation trends matches well with experiment: increasing the Al% shifts the I-V to the right (higher bias). The magnitude of such shift also matches reasonably well: 0.14V for simulation and 0.26±0.05V for experiment.

5.4 Summary

In this chapter, the model has been demonstrated by simulating an realistic LED structure. The I-V characteristics matches well with experimental results. Simulation also reproduced the experimental observation that emission is higher at the p-side QWs. The LED's characteristics have been studied under variation of temperature and Indium content, and the trends matches intuition and experimental observations. A few limitations of the model has been pointed out. First, contact resistance is not inherent in the model, and only until contact resistance was included could the simulation result match experiment. Secondly, the recombination model only accounts

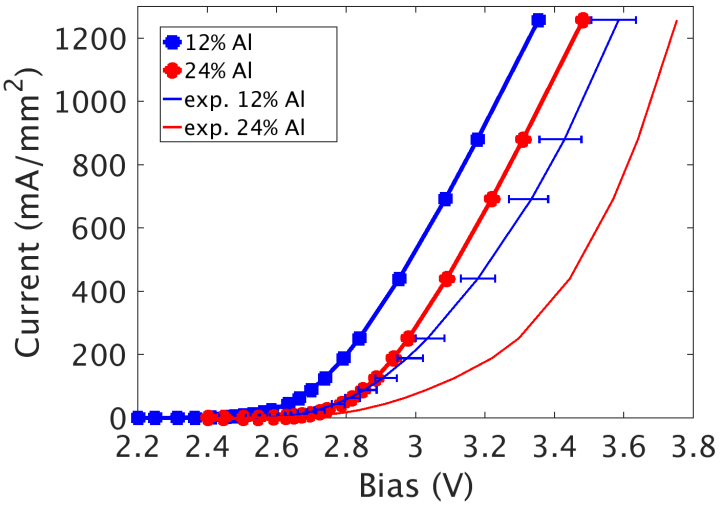

Fig. 5.18.: I-V comparison between simulation results and experiment on similar device structures. Two Al%: 12% and 24% were studied. Experimental result for 12% case contained 6 samples, and the average results with error bars were plotted.

for the average of QW density, not the spatial distribution nor the spectral features. Having a more sophisticated recombination model is necessary to predict the emission spectrum and the correct emission magnitude under strong polarization field. The next chapter will discuss some future work geared towards enhancing the model and making it truly predictive.

6. MODEL EXPANSION

In previous chapters we presented a new approach towards LED model based on the multi-scale equilibrium-nonequilibrium model. The physics foundation of the model were explained in chapter 2 and 3. The model was then applied to simulate a commercial GaN/InGaN in chapter 5 and results agreed well with experiment, serving as a validation of the model. In this chapter, we will discuss about an augmentation to the method: including long range coupling to account for hot-carrier effects.

6.1 Long Range Coupling—Assessment of Hot Carrier Contribution

Our model so far has assumed complete thermalization in the equilibrium region. However, there are two important questions that has not been answered. how good is the equilibrium assumption? How much does hot carriers contribute to the current? In this section, the previous model is augmented to include nonlocal quantum effects, i.e., tunneling through multiple barriers. Figure 6.1 describes the model setup: for comparison purposes, the exact same device structure as the previous one (Fig. 5.1) was used here. The simulation parameter such as recombination rates and temperature were kept same as well. Previously for each 'eq' region, we calculate its net current by including the coherent current through its left and right neighbor barrier ('neq' region) , as well as its recombination current. The main addition here is that for each equilibrium region, we also calculate the tunneling current through multiple neighbor barriers, and include it in the net current formula. As shown in Fig. 6.1, the electron current through barrier No. 4 include four contributions, each successive one covering a longer range: '4 to 4' (single barrier), '3 to 4' (two barrier), '2 to 4' (three barrier), and '1 to 4' (four barrier). We applied the same coherent current formula to calculate transmission (eq. 3.9) and current density (eq. 3.8) through multiple barri-

ers, just as for single barrier. It is important to point out here however, for multiple barriers, the transmission must "go through" scattering regions, i.e., we include the scattering self-energy Σ_η^R in the G^R. Since the transmission method only applies to coherent transport, which does not allow for scattering, this treatment strictly speaking is not allowed. However, here since we are interested in understanding the effect of long-range tunneling (not in its full strict form), including Σ_η^R in G^R can give us a good enough estimation, since Σ_η^R serves as a damping of coherent current.

Numerically, long-range coupling requires the off-diagonal in G^R to be extended, see Fig. 6.2. This requires $\sim 30\%$ extra computation compared to the standard MEQ model.

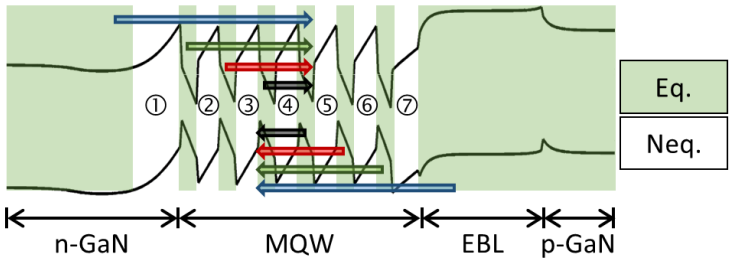

Fig. 6.1.: LED structure considered in this work. Equilibrium (eq - green) and non-equilibrium (neq - white) regions are highlighted. As an example, various tunneling paths through barrier No. 4 are illustrated with arrows of different colors.

Figure 6.3 compares various current components marked in Fig. 6.1 through barrier No. 4. Comparison, two scattering strengths were used in the simulation: a weak scattering case with $\eta = 0.01eV$ is presented along with the normal case ($\eta = 0.1eV$). Subplots (a), (b) are electron current density and (c), (d) are hole current density. Heights of barrier 4 are also marked on the figures with grey dashed lines. At $\eta = 0.01eV$, current density peaks are broadened, but various resonance states are still distinctly visible, at $\eta = 0.1eV$, resonance peaks are much broadened and as a result, states are cross-coupled together and the curve became smooth. Majority of transport occurs under the barriers, and subsequent longer range currents

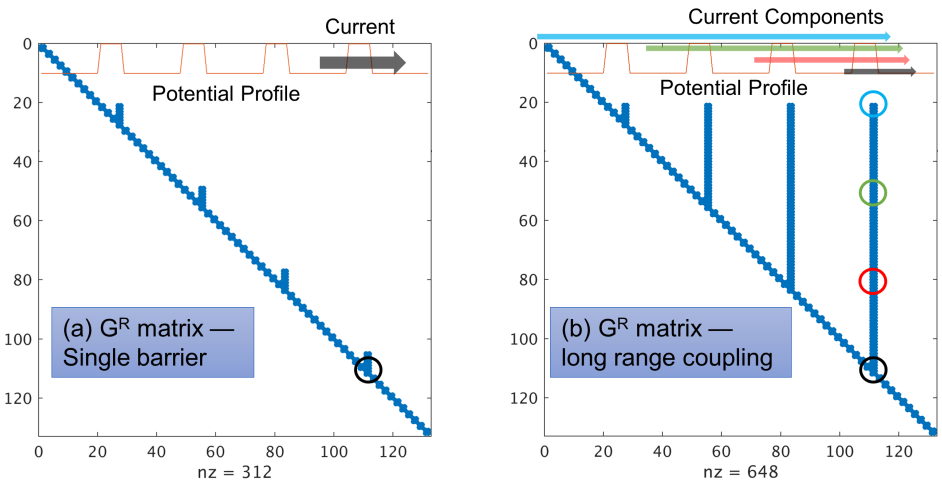

Fig. 6.2.: Comparison of G^R pattern for (a) regular MEQ model (b) MEQ with long range coupling. This is a toy device structure with 4 barriers, shown at the figure top. Long range coupling requires the off-diagonal G^R to be extended. Various current components going through the last barrier are marked with arrows with different colors, and the its responsible G^R element marked with circle of same color.

reduces in magnitude dramatically for both scattering rates. Thermal current above the barriers do not decay significant with coupling range, which is expect since the barrier do not have nearly as much effect on long-range thermionic emission as on long-range tunneling. However, their contribution to the total current is still small.

Figure 6.4 compares overall I-V and IQE results for regular MEQ (short) model and MEQ with long-range coupling. For larger η (more scattering), the forward voltage shifts to lower range (left), and the IQE droop became slightly worse. This is because higher scattering reduces the mobility in the QWs, which leads to more holes piling up at the p-side (see Fig. 5.3). The I-V at 0.1eV agrees quantitatively with experimental results (same results as in Fig. 5.4). The effect of long-range coupling is only observable at low scattering case ($\eta = 0.01eV$). As at high scattering, the long range current components become significantly reduced (Fig. 6.3). The extra contribution from long range components to the total current shifts the forward voltage slightly to the left (Fig. 6.4(a)). Long-range current also contributes to the

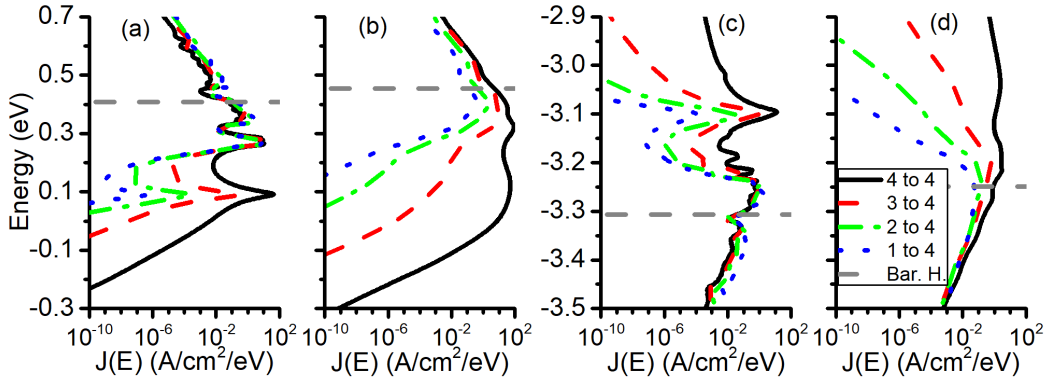

Fig. 6.3.: Energy-resolved current density across the middle barrier (No. 4 in Fig. 6.1) for η=0.01 (a,c) and η=0.1eV (b,d). Barrier heights are marked with grey dashed lines.

electron leakage over the EBL, as well as hole leakage through the n-GaN layer, and thus it slightly reduces efficiency at high current (Fig. 6.4(b)). However, the overall contribution of leakage is relatively small even the worse case ($\sim 1\%$).

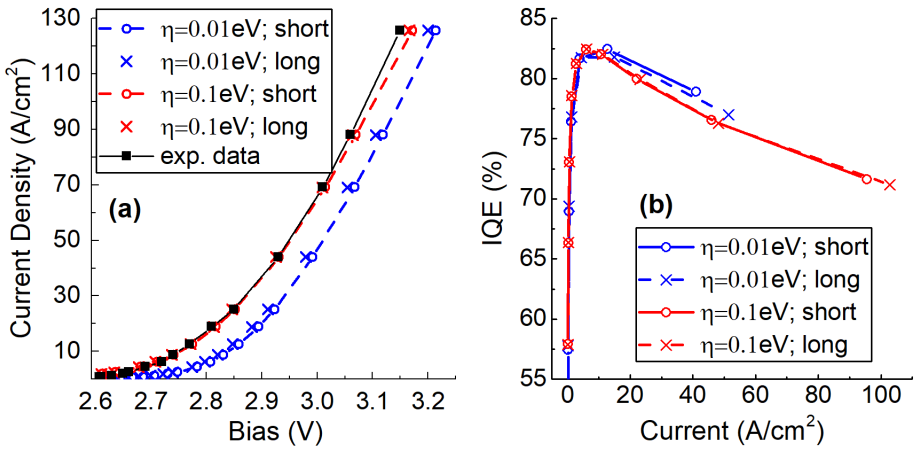

Fig. 6.4.: (a) I-V characteristics with different scattering strengths (η) and tunneling ranges (short vs. long). Larger η suppresses long-range tunneling and long range tunneling can be neglected. The simulated I-V at $\eta = 0.1$eV agrees quantitatively with experimental results (black squares). (b) Internal quantum efficiency (IQE) with different scattering strengths (η) and tunneling ranges. Carriers become more thermalized at higher values of η. This yields charge accumulation at p-side and causes the efficiency droop.

7. ADVANCED RECOMBINATION MODEL—QUANTUM MECHANICAL RADIATIVE RECOMBINATION MODEL

The ABC model has been well-established and widely used in LED modeling. Previously in chapter 3 we discussed its origin and applicability. Up till this point, all the simulation results are obtained using the ABC model for recombination calculation. The strength of the ABC model is its simplicity, allowing us to derive recombination rates easily from known quantities, with very little extra computation. Despite its simplicity, it has worked relatively well in producing qualitative trends, as shown previously in chapter 5. However, the ABC model is a heuristic model, with the 'A','B','C' parameters typically fitted, and it lacks basic physics insights especially the quantum-mechanical nature of electron-hole recombination. Moreover, the coefficients are typically derived from some specific sets of experiments and treated as "one size fits all". The SRH and Auger components currently lack a proper model implementation in NEGF [19], and are typically ignored in any NEGF-based modeling of LED [19] [18]. Improving the non-radiative models beyond the "ABC" approach is beyond the scope of this work. NEGF-based modeling of radiative recombination, however, is already established [15] and adopted previously [19] [18] [34]. This chapter will discuss about developing an efficient, quantum-based modeling approach for radiative recombination. Section 7.1 will list the shortcomings and key missing physics of the "ABC" model. Section 7.2 introduces our model and section 7.3 will show simulation results comparison between the new and previous model.

7.1 Shortcomings of the "ABC" Model

Radiation recombination in the "ABC" model is a simple equation $R_{rad} = B \cdot n \cdot p$, which calculates the recombination rate from direct product of electron and hole charge density. It is illustrated in Fig. 7.1, where electrons and holes are treated as classical particles sitting in conduction and valance bands, and their recombination is a perfect one-to-one relation. However, such an approaches misses three key physics, illustrated in Fig. 7.2.

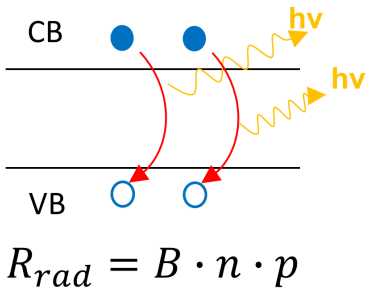

Fig. 7.1.: Illustration of radiative recombination in the "ABC" model. Electrons and holes are treated as classical particles sitting in conduction and valance bands, and their recombination is a perfect one-to-one relation.

Electron and holes in real semiconductor resides in non-flat bandstructures, and their recombination must follow energy selection and momentum selection rules. The energy difference between electron and hole must be same as the photon energy, as illustrated in Fig. 7.2(a) The emission spectrum, which is of key experimental interest, is a direct manifestation of bandstructure details and not provided by the "ABC" model.

Photon emission also must occur between an electron-hole pair of the same momentum, illustrated in Fig. 7.2(b), since the overall momentum during the transition must be conserved and the photon momentum is negligible. Under low carriers, carrier momentums vary in a small range, thus electron and hole momentum typically aligns well. However, momentum mismatch is more pronounced in high carrier den-

sity conditions. Because electron and hole effective mass differ significantly (for GaN: $m_e^* \sim 0.2$ and $m_{hh}^* \sim 1.2$), and an equal energy window could result in large difference in momentum span, as illustrated in Fig. 7.2(b). As a result, under high carrier conditions, the recombination efficiency must decline. Since momentum selection is not covered in the "ABC" model, it typically over-estimates the recombination rate in high density regime, and thus underestimates droop.

The recombination strength depends on the overlap of electron and hole wavefunction. This is especially important for polar materials like nitride, since strong polarization leads to strong band-bending and pronounced quantum-confined Stark-effect (QCSE), as illustrated in Fig. 7.2(c). Increasing the size of the QW can lead to further spatial separation between electron and hole wavefunction, which reduces their recombination strength. This is another missing physical effect in the "ABC" model.

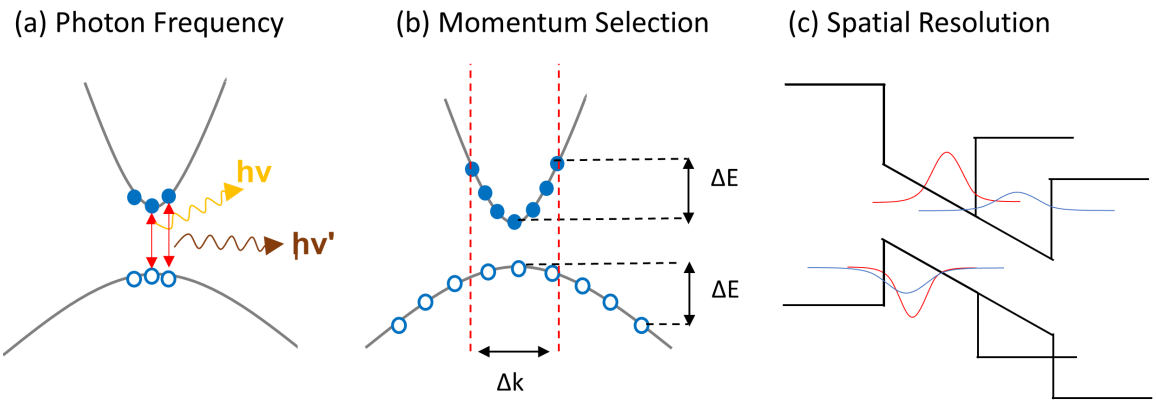

Fig. 7.2.: Illustration of key physics missing in the "ABC" model: (a) emitted photon frequency; (b) momentum selection that limits the transition between electron-hole pair; and (c) spatial resolution of electron and hole wavefunctions.

7.2 Efficient NEGF-based Approach

Electron-Photon Interaction Hamiltonian

Consider a device illuminated by light, the Hamiltonian under the presence of a photonic field is:

$$\hat{H} = \frac{1}{2m_0}(\hat{\boldsymbol{p}} + e\boldsymbol{A})^2 + U \tag{7.1}$$

Here m_0 is the free electron mass, $\boldsymbol{A}$ is the vector potential of the electromagnetic field representing the photon, and U is the periodic crystal potential.

Treating the electromagnetic potential as a perturbation, expand equation 7.1 and dropping second order terms, we can write it as:

$$\begin{aligned} \hat{H} &= \hat{H}_0 + \delta\hat{H} \\ \hat{H}_0 &= \frac{\hat{\boldsymbol{p}}^{\boldsymbol{2}}}{2m_0} + U \\ \delta\hat{H} &= \frac{e}{m_0}\boldsymbol{A}\cdot\hat{\boldsymbol{p}} \end{aligned} \tag{7.2}$$

$\hat{H}$ is split into two terms where $\hat{H}_0$ is the normal Hamiltonian without perturbation, and $\delta\hat{H}$ is the perturbation Hamiltonian due to photonic field.

Dipole Matrix Element

The coupling strength between two states under the photonic field is measured by the dipole matrix element M^{γ} coupling them:

$$M^{\gamma}_{\alpha,L;\alpha',L'}(\boldsymbol{k}) = \frac{e}{m_0}\boldsymbol{A_0}\big\langle\alpha, L, \boldsymbol{k}|\hat{\boldsymbol{p}}|\alpha', L', \boldsymbol{k}\big\rangle \tag{7.3}$$

Here we assume $\boldsymbol{A}$ has no spatial dependence. L, L' are atom positions and α, α' are orbital indices. For simplicity without losing generality, if we ignore the in-plane momentum (setting $\boldsymbol{k} = 0$), it can be written in terms of the Hamiltonian elements as follows:

$$
\begin{aligned}
M^{\gamma}_{\alpha,L;\alpha',L'} &= \frac{e}{m_0}\boldsymbol{A_0}\big\langle \alpha, L, \boldsymbol{k}=0|\hat{\boldsymbol{p}}|\alpha', L', \boldsymbol{k}=0\big\rangle \\
&= \frac{e}{m_0}\boldsymbol{A_0}\cdot\frac{m_0}{i\hbar}\big\langle \alpha, L, \boldsymbol{k}=0|\big[\hat{\boldsymbol{r}}, \hat{\boldsymbol{H}_0}\big]|\alpha', L', \boldsymbol{k}=0\big\rangle \\
&= A_0\cdot\frac{e}{i\hbar}(\boldsymbol{R}^{L'}-\boldsymbol{R}^{L})[H_0]_{\alpha,L;\alpha',L'}
\end{aligned}
\tag{7.4}
$$

Here we assume the light polarization is purely transverse, and $\boldsymbol{A}_0$ is along the layer direction, thus reduces to a scalar $A_0 = \sqrt{\frac{\hbar}{2\epsilon_0 V\omega_q}}$. Therefore, M^{γ} is only dependent on the transverse momentum (no energy dependence).

Electron-Photon Self-Energy

The interaction between electron and photonic field is treated by electron-photon self-energy in NEGF:

$$
\begin{aligned}
\Sigma^{\lessgtr}_{\gamma}(k;E) =& i\hbar M^{\gamma}(k)[N_{\gamma}G^{\lessgtr}(k;E\mp\hbar\omega_{\gamma}) \\
&+ (N_{\gamma}+1)G^{\lessgtr}(k;E\pm\hbar\omega_{\gamma})]M^{\gamma}(k)
\end{aligned}
\tag{7.5}
$$

N_{γ} is the number of photons which can be described by the Bose-Einstein distribution, and M^{γ} represents the inter-band momentum matrix element between conduction and valance band. The first term (with $\mp$) represents the absorption and the second term (with $\pm$) represents the emission.

In our system, the bandgap (thus the emitted photon energy) is much larger than the energy window of relevance for electron and holes. Therefore, the only transition that is allowed is the inter-band transition between electrons in the conduction band ($G^{CB,<}$), and holes in the valance band ($G^{VB,>}$). Furthermore, we assume emitted photons cannot be re-absorbed ($N_{\gamma}=0$), thus the only interaction self-energy we need to take into account is:

$$
\Sigma^{CB,>}(k;E) = i\hbar M^{\gamma}(k)G^{>}(k;E-\hbar\omega_{\gamma})M^{\gamma}(k) \tag{7.6}
$$

The self-energy $\Sigma^{CB,>}$ describes the process of an electron in the conduction band emitting photon of $\hbar\omega_\gamma$ and recombine with a hole in the valance band.

In the normal NEGF approach, $\Sigma^{CB,>}$ will then be used to self-consistently calculate Σ^R, G^R and $G^<$, since the photon emission process affects the carrier dynamics. However, this approach would defeat the whole purpose of our multi-scale NEGF framework, since the computation cost of self-consistent Born scattering calculation will completely overshadow the numerical efficiency of the MEQ model and thus render it useless. Moreover, the photon emission is a non-local process, which requires to calculate off-diagonal elements in the Green's functions. Making it work with the MEQ approach means diagonal approximation must be adopted, which itself will be a heavy approximation.

In our approach, photon emission process is treated with linear approximation within each quantum well. Namely, the self-energy is calculated in one-shot, and then obtain the photon emission rate by calculating the current due to electron-photon scattering. The emission rate for photon with frequency ω_γ is

$$\begin{aligned} R(\hbar\omega_\gamma) &= \frac{1}{V}\int \frac{kdk}{2\pi}\int \frac{dE}{2\pi\hbar} G^{CB,<}(k;E)\Sigma^{CB,>}(k;E) \\ &\approx -\frac{\hbar}{(2\pi)^2 V}\int kdk M^\gamma(k)^2 \int dE G^{CB,<}(k;E)G^{VB,>}(k;E-\hbar\omega_\gamma) \\ &\approx \beta\cdot\frac{1}{\omega_\gamma}\int kdk\int dE G^{CB,<}(k;E)G^{VB,>}(k;E-\hbar\omega_\gamma) \end{aligned} \tag{7.7}$$

In the above equation, we made the approximation that the dipole matrix element for a particular momentum is spatially invariant, and thus can be separated out of the integral and treated as a constant β. This assumption is based on the fact that because the material composition within the LED do not vary a lot (since the majority composition of $InGaN$ and $AlGaN$ is same as the substrate material GaN), the Hamiltonian coupling element on various lattice sites in equation 7.4 also vary to a small extent.

The total photon emission rate in each QW (denoted by index 'i') $\tilde{R}_i$ is the sum of emission in QW i from all photon frequencies

$$\tilde{R}_i = \sum_{\omega_\gamma} R_i(\hbar\omega_\gamma) = \beta_i \cdot \sum_{\omega_\gamma} \frac{1}{\omega_\gamma} \int kdk \int dE G_i^{CB,<}(k;E) G_i^{VB,>}(k;E-\hbar\omega_\gamma) \quad (7.8)$$

It is interesting to note that our equation 7.8 above resembles very closely to the radiative counterpart in the "ABC" equations 3.12, with β taking the place of the B-coefficient, $G_i^{CB,<}$ representing the electron density n_i and $G_i^{VB,>}$ corresponding to the hole density p_i. This to some extent, is a proof of validity of the ABC model. It is important to note, however, that equation 7.8 is no longer a heuristic but rather a physics-based model which covers all essential physics that were previously missing in the "$B \cdot n \cdot p$" model, as discussed in the beginning of this section.

Process Flow

In our approach, we maintain the equilibrium assumption in the QWs and apply the same multi-eq-neq model to simulating the LED but replacing the radiative recombination with the new model discussed in section 7.2. For each QW "i", we solve equation 7.8 to obtain its total radiative rate. For the radiative coefficients "β_i", we calibrate them against the ABC model upfront, and keep them throughout the rest of simulations. Namely:

$$\begin{aligned} \beta_i &= \frac{B \cdot n_i \cdot p_i}{\tilde{K}} \\ \tilde{K} &= \sum_{\omega_\gamma} \frac{1}{\omega_\gamma} \int kdk \int dE G_i^{CB,<}(k;E) G_i^{VB,>}(k;E-\hbar\omega_\gamma) \end{aligned} \quad (7.9)$$

This way allows us to efficiently apply our model without needing to calculate the matrix product between Green's function and the dipole matrix. Because the ABC model is used to calibrate the β coefficients upfront, it must be assumed that the B-coefficients used are well validated (against experiment or other trusted sources), and that ABC model is applicable to the calibration case, meaning electrons and holes densities are not too high to cause significant momentum mismatch. Once calibrated,

the β coefficient is assumed to be a charge-independent quantities that encompass all material and environmental factors. Then β_i is inserted into equation 7.8, where the integral accounts for photon frequency, momentum-selection and spatial charge overlap. The three-step process is summarized below and depicted in Fig. 7.3.

- Step 1: For each individual QW, evaluate $\tilde{K}$. (eq. 7.9)
- Step 2: calibrate β using the conventional ABC model with verified B coefficient. (eq. 7.9)
- Step 3: Proceed the rest of MEQ model, using eq. 7.8 for radiative current and β calibrated from step 2.

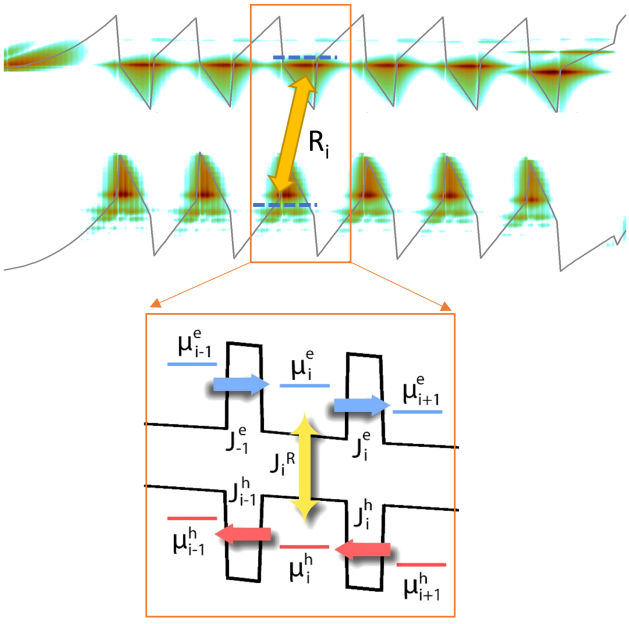

Fig. 7.3.: Process of calculating the position and energy-resolved radiative recombination rate. Each quantum well is still treated as equilibrium where radiative rates are calculated with new model, and Fermi levels in all QWs are adjusted to conserve total current.

7.3 Result Comparison

The new radiative recombination model is applied to simulate the previous device (see Fig. 5.1) and results are compared. Figure 7.4 compares the I-V and IQE using the previous 'ABC' model and the new model labeled as 'Quantum'. I-V comparison between the two models shown in figure 7.4(a) looks nearly identical. This indicates that the transport characteristics are insensitive to the choice of recombination model. It also validates our previous I-V results (ones matching experiment) which were based on the simplistic ABC model.

The IQE (Fig. 7.4(b)) on the other hand, shows some difference between the "quantum" and "ABC" model. The peak values are near identical (because of the calibration step). However, the efficiency calculated with the new model is lower at higher current (more droop). This is because the "ABC" model assumes complete recombination between electrons and holes (perfect momentum matching), while the "quantum" model accounts for momentum and energy selection, which makes a larger difference for higher electron and hole density.

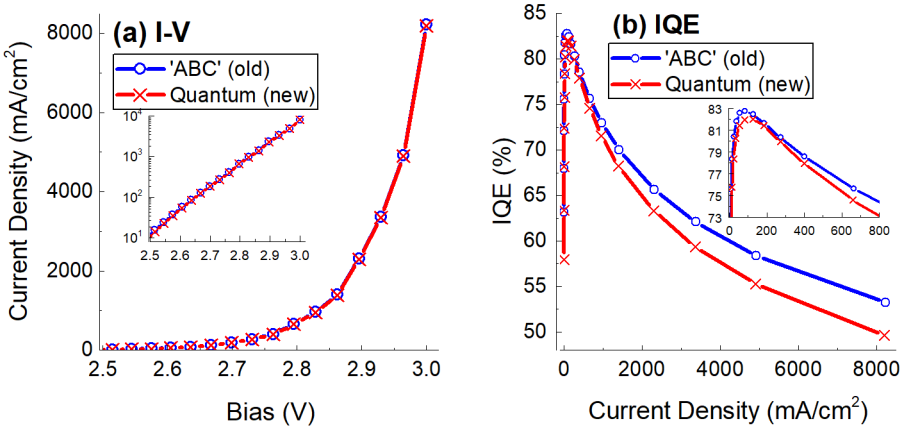

Fig. 7.4.: Comparison of (a) IV and (b) IQE between the old and new recombination models.

To further illustrate the last point, we plot a few quantities at the highest bias point (right-most point in Fig. 7.4(b)) in figure 7.5. Plots (a) and (b) shows the averaged electron and hole density at each QW; (c) and (d) shows the radiative and Auger recombination current at each QW. In each plot, two models are plotted together and the ratio (new/old) is also plotted as dash line (to the right axis). From (a) and (b), one can tell that the new model has resulted in slightly more charge than the old (ratio > 1). This also lead to a slight higher Auger recombination with the new model (also ratio > 1, see (c)). However, only the radiative recombination (d) is lower than the old model (ratio < 1, see (d)). Since our Auger model currently does not account for any carrier selection, more charge (a and b) directly translates to more Auger recombination (c). For radiative recombination with the new model, more charge still leads to more recombination, but the rate of increase is slower than before due to the consideration of momentum selection.

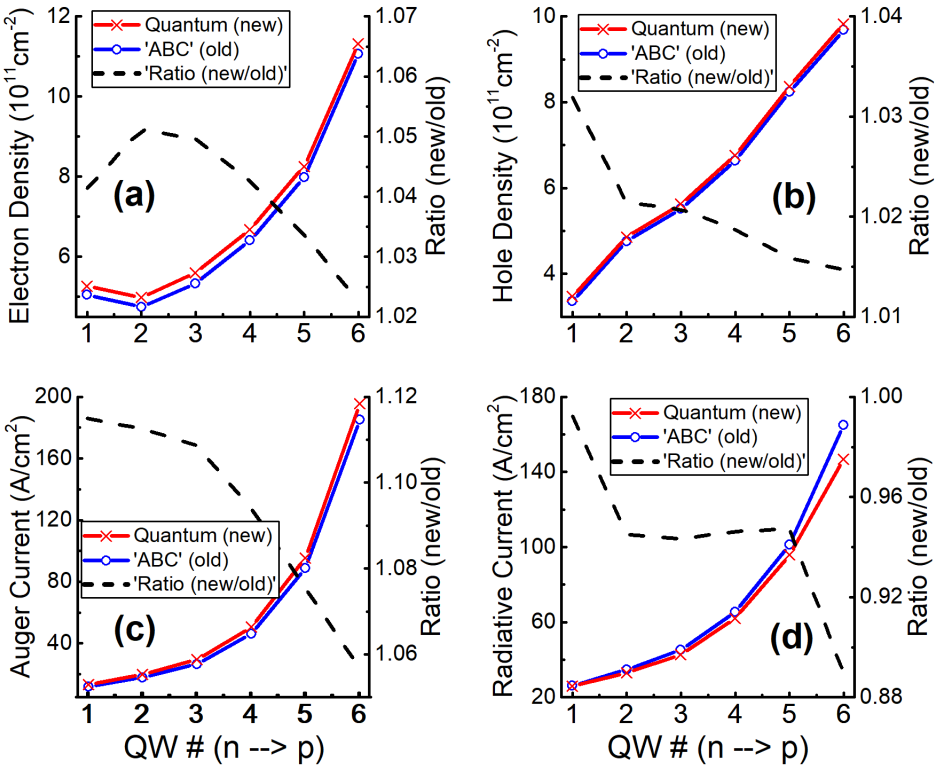

Fig. 7.5.: Comparison of (a) average electron density (b) average hole density (c) Auger recombination (d) radiative recombination between the new model ("quantum", red curve) and the old ("ABC", blue curve). Also in each subplot, the ratio between new and old are plotted (black dashed line) to the right axis.

8. SUMMARY AND OUTLOOK

In this work, a multi-scale transport model was developed for LED, and could potentially be applied to a broader range of optoelectronic devices. It gets the best of both worlds from traditional semiclassical model and status-quo full NEGF model. Here are the three main characteristics:

- NEGF-based – includes coherent quantum effects and relaxation.
- Multi-scale – numerically efficient, solves transport in critical regions only. This allows, for the first time, full atomistic quantum transport simulation on realistic-scaled LED devices.
- Allowing for efficient inclusion of recombination, both in heuristic and QM (radiative only) form.

The model was demonstrated on a realistic, commercial LED device, and quantitatively matched I-V with experiment. We also deduced carrier temperature from simulation, reproduced p-sided emission pattern and validated Auger as the dominant contributor to droop.

We then tapped into device engineering by studying two key design parameters: barrier thickness and EBL Al%. Based on simulation results, a key advice was engineering for better hole transport, and we recommended the optimal barrier thickness and Al%.

There are still a few key areas upon which the current model could improve.

Hot Carrier Due to Auger Recombination

Auger recombination generates excess carriers with high energy, these hot carriers lead to increased leakage current. Our simulation result suggesting EBL plays a

small role (leakage current is low even at low Al%, see fig. 5.16) is indication that we are underestimating the leakage current in some way. Leakage due to Auger recombination is the most likely missing piece.

A simple way to include it is to add hot carrier current terms J^h_{hc} and J^e_{hc} in the current balance equation (see fig. 8.1), and have a physical connection to the Auger recombination.

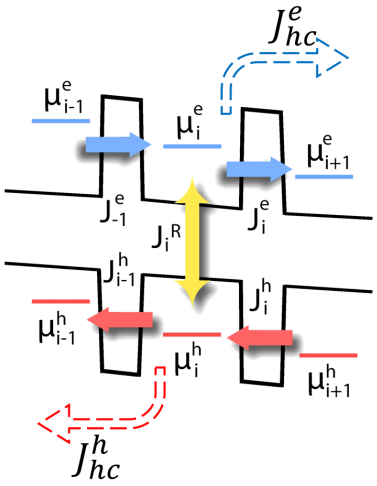

Fig. 8.1.: Illustration of including Auger-caused leakage components in the model.

Improving the "η" Model

In the current approach, η in the scattering self-energy is constant value defined by the band edges, which are parameterized for bulk materials. This leads to DOS in the equilibrium region following the bulk band edges (see fig. 5.2), rather than the shape of quantum states. The η self-energy is pure imaginary, in reality it contains real-part which causes the resonance states to shift in energy and such effect is not captured here. Therefore, an alternative approach is to The alternative approach is use η self-energy that are directly related to (or composed of) the Green's functions.

Improving QW Distribution Function

The current model assumes Fermi distribution in each quantum well. In reality, the carrier distribution could be skewed towards higher energy, due to hot electrons. The skewed carrier distribution could have interesting effects in both transport and optical properties. A proposed solution is to perform a full NEGF simulation on a single quantum well, including all relevant scattering mechanisms (POP, impurity, e-e, etc.), extract the "real" carrier distribution under real conditions, and use them in the equilibrium regions of our model.

Including Scattering in the barriers

The current approach only allows coherent transport in the non-equilibrium regions. In general, the multi-scale framework developed here could allow for inclusion of real scattering in those regions. Scattering could have impact on tunneling current (e.g. via traps) and thermionic emission. To do that, the non-eq regions need to be expanded to solve incoherent transport with the self-consistent Born, with neighboring eq. regions as boundary conditions. Additional simulation flow need to be added, namely iterating SC-born with multi-eq-neq. A rough estimate indicates a minimum of three orders of magnitude of extra computation load is expected, which would make such simulation irrelevant for real industry applications.

Improving the Auger model

In the previous chapter, we've demonstrated an efficient method of quantum-based radiative recombination. In those simulations, Auger recombination is still calculated with the 'ABC. Similar enhancement on the radiative recombination could be applied to Auger. However, the Auger mechanism is much more involved compared to radiative recombination. Instead of interaction between an e-h pair, it involves two more elements: another electron (or hole), an empty electronic (hole) states in higher

(lower) energy. This significantly complicates the matter. On top of that, two types of Auger processes: e-e-h and h-h-e needs to be considered. Direct Auger recombination could be much less important compared to phonon-assisted Auger recombination [45], so photon interaction will probably need to be included as well.

9. SUPPLEMENT INFORMATION

9.1 Data Location

All the simulation data, as well as associated NEMO5 input decks used to generate figures in this thesis are stored in the Klimeck group SVN repository: *https://nanohub.org/groups/klimeck/svn/trunk/StudentData/JunzheGeng/PhD_thesis*.

Figure 4.4: input deck and data stored in */data/meshing/Emesh/*.

Figure 4.6 and 4.7: upper two plots of both figures located in */data/meshing/kmesh/current/*; lower two plots of both figures located in */data/meshing/kmesh/k_kcurrent/*. Figure 4.6 corresponds to the conduction band (data files containing label "CB" or "electron"), and Figure 4.7 corresponds to the valence band (data files containing label "VB" or "hole").

Figure 5.4: all input decks and data associated are located in */data/LED_sweep_temperature/*. IV files are contained in */data/LED_sweep_temperature/IV_collection/*, files names contain "_with_Rs". Experimental I-V data located in */data/Lumileds_data/IV_25_vs_85C.xlsx*.

Figure 5.2, 5.3, 5.5, 5.6, 5.7 all corresponds to the same set of simulations. All data and input decks are located in */data/LED_prototype/*, which contains a shortcut to */data/LED_sweep_temperature/IV_collection/T_360* and a folder *energy_resolved_data* containing energy-resolved data (shown in Fig. 5.2). Figures 5.3, 5.7 corresponds to the same temperature and same bias point, located in folder *2c90V*.

Figure 5.8, 5.9, 5.10, 5.11 correspond to the same set of simulations (sweep barrier width). All data and input decks located at */data/LED_sweep_barr_width/*.

Figure 5.12, 5.13, 5.14, 5.15, 5.16, 5.17, 5.18 correspond to the same set of simulations (sweep aluminum concentration). All data and input decks located at

/data/LED_sweep_Al_pct/. Experimental data in Fig. 5.18 are contained in */data/Lumileds_data/EBL_IV_data.xlsx*.

Figure 6.3 and 6.4 correspond to the same set of simulations (compare long vs. short range). All data and input decks located at */data/LED_compare_eta_coupling_range/*.

9.2 Regression Test Location

Key functionalities related to the LED simulation are covered by the regression test system. As of NEMO5 revision 24753, all "LED feature" related regression tests are located in */regression_test/numerical_test/Philips_LED/V2/*. Under the directory, there are two folders: *Feature* and *Device*.

Feature directory, contains tests on cores features, such as

- Analytical k integration (applicable to effective mass), located in */Feature/analytical_momentum*.
- Electron/hole transport only, located in */Feature/CB_transport_only* and */Feature/VB_transport_only*.
- k-dependent η, located in */Feature/Confined_band_solver*.
- Coherent current calculation, located in */Feature/Current*.
- Current conservation (balancing Fermi levels), located in */Feature/Current_conservation*.
- Density calculation, for wire (no momentum) UTB (1D momentum) and quasi-1D (2D momentum), located in */Feature/Density*.
- Recombination, located in */Feature/Recombination*.
- Poisson, located in */Feature/Self_consistent*.
- Energy mesh refinement on a double barrier structure, located in */Feature/resonance_Emesh*.

- Momentum mesh refinement on a double barrier structure, located in */Feature/resonance_kmesh*.

Device directory, contains tests on typical device structures

- 6-QW GaN/InGaN LED with AlGaN EBL, representing all structures simulated in chapter chapter:results and . Computationally simplified to fixed number of energy and k point; one-shot Poisson calculation with semiclassical initial guess. Located in */Device/6qw_LED_em_analy_k_oneshot/*.

- 6-QW GaN/InGaN LED with AlGaN EBL (same structure as above), but with effective mass band model and analytical momentum integration. This allows much faster computation, suitable to serve as an estimation model, or being used for nanoHUB tool. Located in */Device/6layer_oneshot_poisson/*.

- 4-barrier GaN/InGaN structure; sp^3 tight-binding model; one-shot Poisson calculation with semiclassical initial guess. This is a toy structure for model testing, not real device. Located in */Device/4_barr_LED_toy/*.

LIST OF REFERENCES

LIST OF REFERENCES

[1] "Energy savings forecast of solid-state lighting in general illumination applications," 2014.

[2] J. T. J. N. Roland Haitz, Fred Kish, "The case for a national research program on semiconductor lighting," 2000.

[3] C. Weisbuch, M. Piccardo, L. Martinelli, J. Iveland, J. Peretti, and J. S. Speck, "The efficiency challenge of nitride light-emitting diodes for lighting," *physica status solidi (a)*, vol. 212, no. 5, pp. 899–913, 2015.

[4] R. Haitz and J. Y. Tsao, "Solid-state lighting: 'the case 10 years after and future prospects," *physica status solidi (a)*, vol. 208, no. 1, pp. 17–29, 2011.

[5] M.-H. Kim, M. F. Schubert, Q. Dai, J. K. Kim, E. F. Schubert, J. Piprek, and Y. Park, "Origin of efficiency droop in gan-based light-emitting diodes," *Applied Physics Letters*, vol. 91, no. 18, 2007.

[6] P. M. Pattison, J. Y. Tsao, and M. R. Krames, "Light-emitting diode technology status and directions: opportunities for horticultural lighting," in *ActaHortic.*, no. 1134, pp. 413–426, International Society for Horticultural Science (ISHS), Leuven, Belgium, May 2016.

[7] J. Xie, X. Ni, Q. Fan, R. Shimada, U. Özgür, and H. Morko, "On the efficiency droop in ingan multiple quantum well blue light emitting diodes and its reduction with p-doped quantum well barriers," *Applied Physics Letters*, vol. 93, no. 12, p. 121107, 2008.

[8] S. Y. Karpov, "Modeling of iii-nitride light-emitting diodes: Progress, problems, and perspectives," 2011.

[9] *Crosslight Website*, 2017 (Accessed: 2017-06-19).

[10] *Silvaco Website*, 2017 (Accessed: 2017-06-19).

[11] *STR Website*, 2017 (Accessed: 2017-06-19).

[12] G. Klimeck, R. Lake, R. C. Bowen, W. R. Frensley, and T. S. Moise, "Quantum device simulation with a generalized tunneling formula," *Applied Physics Letters*, vol. 67, no. 17, pp. 2539–2541, 1995.

[13] M. Luisier and G. Klimeck, "Atomistic full-band simulations of silicon nanowire transistors: Effects of electron-phonon scattering," *Phys. Rev. B*, vol. 80, p. 155430, Oct 2009.

[14] T. Kubis, C. Yeh, P. Vogl, A. Benz, G. Fasching, and C. Deutsch, "Theory of nonequilibrium quantum transport and energy dissipation in terahertz quantum cascade lasers," *Phys. Rev. B*, vol. 79, p. 195323, May 2009.

[15] U. Aeberhard and R. H. Morf, "Microscopic nonequilibrium theory of quantum well solar cells," *Phys. Rev. B*, vol. 77, p. 125343, Mar 2008.

[16] R. C. Bowen, G. Klimeck, R. K. Lake, W. R. Frensley, and T. Moise, "Quantitative simulation of a resonant tunneling diode," *Journal of Applied Physics*, vol. 81, no. 7, pp. 3207–3213, 1997.

[17] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, "Single and multiband modeling of quantum electron transport through layered semiconductor devices," *Journal of Applied Physics*, vol. 81, no. 12, pp. 7845–7869, 1997.

[18] S. Steiger, R. G. Veprek, and B. Witzigmann, "Electroluminescence from a quantum-well led using negf," in *2009 13th International Workshop on Computational Electronics*, pp. 1–4, May 2009.

[19] A. Shedbalkar, Z. Andreev, and B. Witzigmann, "Simulation of an indium gallium nitride quantum well light-emitting diode with the non-equilibrium green's function method," *physica status solidi (b)*, vol. 253, no. 1, pp. 158–163, 2016.

[20] M. Auf der Maur, "Multiscale approaches for the simulation of ingan/gan leds," *Journal of Computational Electronics*, vol. 14, no. 2, pp. 398–408, 2015.

[21] J. C. Slater and G. F. Koster, "Simplified lcao method for the periodic potential problem," *Phys. Rev.*, vol. 94, pp. 1498–1524, Jun 1954.

[22] S. Datta, *Electronic Transport in Mesoscopic Systems*. Cambridge Studies in Semiconductor Physi, Cambridge University Press, 1997.

[23] H. Haug and A. Jauho, *Quantum Kinetics in Transport and Optics of Semiconductors*. Solid-State Sciences, Springer Berlin Heidelberg, 2008.

[24] D. Fisher and P. Lee, "Relation between conductivity and transmission matrix," *Phys. Rev. B*, vol. 23, pp. 6851–6854, Jun 1981.

[25] M. Goano, F. Bertazzi, X. Zhou, M. Mandurrino, S. Dominici, M. Vallone, G. Ghione, A. Tibaldi, M. Calciati, P. Debernardi, F. Dolcini, F. Rossi, G. Verzellesi, M. Meneghini, N. Trivellin, C. De Santi, E. Zanoni, and E. Bellotti, "Challenges towards the simulation of gan-based leds beyond the semiclassical framework," 2016.

[26] Y. Tan, M. Povolotskyi, T. Kubis, T. B. Boykin, and G. Klimeck, "Transferable tight-binding model for strained group iv and iii-v materials and heterostructures," *Physical Review B*, vol. 94, Jul 2016.

[27] K. P. O'Donnell, T. Breitkopf, H. Kalt, W. V. der Stricht, I. Moerman, P. Demeester, and P. G. Middleton, "Optical linewidths of ingan light emitting diodes and epilayers," *Applied Physics Letters*, vol. 70, no. 14, pp. 1843–1845, 1997.

[28] O. Ambacher, D. Brunner, R. Dimitrov, M. Stutzmann, A. Sohmer, and F. Scholz, "Absorption of ingan single quantum wells determined by photothermal deflection spectroscopy," *Japanese Journal of Applied Physics*, vol. 37, no. 3R, p. 745, 1998.

[29] M. A. Khayer and R. K. Lake, "Effects of band-tails on the subthreshold characteristics of nanowire band-to-band tunneling transistors," *Journal of Applied Physics*, vol. 110, no. 7, 2011.

[30] A. Wacker, "Semiconductor superlattices: A model system for nonlinear transport," *Physics Reports*, vol. 357, pp. 1–111, 2002.

[31] A. David, N. G. Young, C. A. Hurni, and M. D. Craven, "All-optical measurements of carrier dynamics in bulk-gan leds: Beyond the abc approximation," *Applied Physics Letters*, vol. 110, no. 25, p. 253504, 2017.

[32] J. Piprek, "Efficiency droop in nitride-based light-emitting diodes," *physica status solidi (a)*, vol. 207, no. 10, pp. 2217–2225, 2010.

[33] Y. C. Shen, G. O. Mueller, S. Watanabe, N. F. Gardner, A. Munkholm, and M. R. Krames, "Auger recombination in ingan measured by photoluminescence," *Applied Physics Letters*, vol. 91, no. 14, p. 141101, 2007.

[34] R. Wang, Y. Zhang, F. Bi, T. Frauenheim, G. Chen, and C. Yam, "Quantum mechanical modeling the emission pattern and polarization of nanoscale light emitting diodes," *Nanoscale*, vol. 8, pp. 13168–13173, 2016.

[35] G. Klimeck, "Quantum and semi-classical transport in nemo 1-d," *Journal of Computational Electronics*, vol. 2, no. 2, pp. 177–182, 2003.

[36] R. C. B. C. L. F. Gerhard Klimeck, Roger K. Lake and W. R. Frensley, "Resolution of resonances in a general purpose quantum device simulator (nemo)," *VLSI Design*, vol. 6, pp. 107–110, 1998.

[37] R. C. Bowen, W. R. Frensley, G. Klimeck, and R. K. Lake, "Transmission resonances and zeros in multiband models," *Phys. Rev. B*, vol. 52, pp. 2754–2765, Jul 1995.

[38] F. Bernardini, *Spontaneous and Piezoelectric Polarization: Basic Theory vs. Practical Recipes*, pp. 49–68. Wiley-VCH Verlag GmbH & Co. KGaA, 2007.

[39] T.-J. Yang, R. Shivaraman, J. S. Speck, and Y.-R. Wu, "The influence of random indium alloy fluctuations in indium gallium nitride quantum wells on the device behavior," *Journal of Applied Physics*, vol. 116, no. 11, p. 113104, 2014.

[40] C. A. Hurni, A. David, M. J. Cich, R. I. Aldaz, B. Ellis, K. Huang, A. Tyagi, R. A. DeLille, M. D. Craven, F. M. Steranka, and M. R. Krames, "Bulk gan flip-chip violet light-emitting diodes with optimized efficiency for high-power operation," *Applied Physics Letters*, vol. 106, no. 3, 2015.

[41] Y. C. Shen, J. J. Wierer, M. R. Krames, M. J. Ludowise, M. S. Misra, F. Ahmed, A. Y. Kim, G. O. Mueller, J. C. Bhat, S. A. Stockman, and P. S. Martin, "Optical cavity effects in ingan/gan quantum-well-heterostructure flip-chip light-emitting diodes," *Applied Physics Letters*, vol. 82, no. 14, pp. 2221–2223, 2003.

[42] A. David, M. J. Grundmann, J. F. Kaeding, N. F. Gardner, T. G. Mihopoulos, and M. R. Krames, "Carrier distribution in (0001)ingangan multiple quantum well light-emitting diodes," *Applied Physics Letters*, vol. 92, no. 5, p. 053502, 2008.

[43] J. Piprek, "How to decide between competing efficiency droop models for gan-based light-emitting diodes," *Applied Physics Letters*, vol. 107, no. 3, p. 031101, 2015.

[44] J. Iveland, L. Martinelli, J. Peretti, J. S. Speck, and C. Weisbuch, "Direct measurement of auger electrons emitted from a semiconductor light-emitting diode under electrical injection: Identification of the dominant mechanism for efficiency droop," *Phys. Rev. Lett.*, vol. 110, p. 177406, Apr 2013.

[45] E. Kioupakis, P. Rinke, K. Delaney, and C. Van De Walle, "Indirect auger recombination as a cause of efficiency droop in nitride light-emitting diodes," *Applied Physics Letters*, vol. 98, 4 2011.