33

Quantum Cascade Lasers: Electrothermal Simulation

Song Mei
Yanbing Shi
Olafur Jonasson
and
Irena Knezevic

33.1 Introduction

Quantum cascade lasers (QCLs) are high-power, coherent light sources emitting in the mid-infrared (mid-IR) and terahertz (THz) frequency ranges [1]. QCLs are electronically driven, unipolar devices whose active core consists of tens to hundreds of repetitions of a carefully designed stage. The QCL active core can be considered a superlattice (SL) in which each stage is a multiple quantum well (MQW) heterostructure where confined electronic states with specific energy levels are formed because of quantum confinement. The concept of achieving lasing in semiconductor SLs was first introduced by Kazarinov and Suris [2] in 1971. The first working QCL was demonstrated by Faist et al. [1] two decades later.

QCLs are typically III-V material systems grown on GaAs or InP substrates. Molecular beam epitaxy (MBE) [3] and metal-organic chemical vapor deposition (MOCVD) [4] are the techniques that enable precise growth of thin layers of various III-V alloys. It is also possible to incorporate strain into the structure as long as the total strain in a stage is balanced. Both the precision and the possibility of introducing strain bring great flexibility to the design of the QCL active core, so lasing over a wide range of wavelengths (from 3 to 190 μm) has been achieved. The growth techniques produce high-quality interfaces with atomic-level roughness.

Mid-IR QCLs (wavelength range 3–12 μm) have widespread military and commercial applications. A practical portable detector requires mid-IR QCLs to operate at room temperature (RT) in continuous wave (CW) mode and with high (watt-level) output power. Furthermore, these QCLs must also have high wall-plug efficiency (WPE) (the ratio of emitted optical power to the electrical power pumped in) and

long-term reliability under these high-stress operating conditions. As the stress likely stems from excessive nonuniform heating while lasing [5,6], improving device reliability and lifetime goes hand in hand with improving the WPE.

33.1.1 Lasing in QCLs

In QCLs, multiple conduction subbands are formed in the active core by means of quantum confinement. QCLs are unipolar devices, meaning that lasing is achieved through radiative intersubband transitions (transitions between two conduction subbands) instead of radiative interband transitions (transitions between the conduction and valence bands) in traditional quantum well (QW) semiconductor lasers. As a result, electrons do not combine with holes after the radiative transitions and can be used to emit another photon. In order to reuse electrons, the same MQW heterostructure is repeated many times (25–70) in the QCL active core (the so-called cascading scheme).

Figure 33.1 depicts a typical conduction band diagram of two adjacent stages in a QCL under an electric field. Each stage consists of an injector region and an active region. The injector region has several thin wells separated by thin barriers (10–30 Å), so a miniband is formed with multiple subbands that are close in energy and whose associated wavefunctions have high spatial overlap. Typically, the lowest few energy levels in the miniband are referred to as the injector levels. The injector levels collect the electrons that come from the previous stage and inject them into the active region. The active region usually consists of 2–3 wider wells (40–50 Å) separated by thin barriers. Consequently, a minigap forms in the active region between the upper lasing level (3) and the lower lasing level (2). Another important energy level in the active region is the ground state (1). There is a thin barrier (usually the thinnest among all layers) between the injecting region and the active region, called the injection barrier.

By design, the injector levels are close in energy and strongly coupled to the upper lasing level because of the thin injection barrier. The upper and lower lasing levels have large spatial overlap, which allows a radiative transition between the two levels; the wavelength of the emitted light is determined by the energy spacing between these two levels. The lower lasing level overlaps with the ground state for efficient electron extraction. Electron emission of longitudinal optical (LO) phonons is the dominant mechanism for electron extraction, so the energy spacing between the lower lasing level and the ground state is designed to be close to the LO phonon energy to facilitate extraction. With careful design, the electron lifetime in the upper lasing level is longer than in the lower lasing level, so population inversion can be

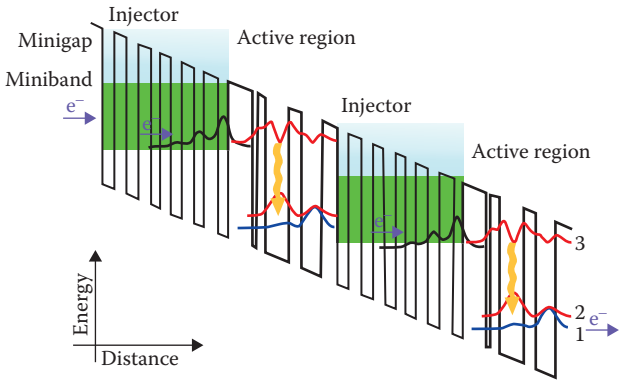

FIGURE 33.1 A typical conduction band diagram of two adjacent QCL stages under an applied electric field. Each stage consists of an injector region and an active region. A miniband is formed in the injector region while a minigap is formed in the active region (between the upper and lower lasing levels). Lasing is associated with a radiative transition from the upper (3) to the lower (2) lasing level. Electrons in the lower lasing level depopulate quickly to the ground level (1) by emission of longitudinal optical phonons.

achieved. After reaching the ground state, electrons tunnel through the injector into the upper lasing level of the next stage, and the process is repeated. Of course, the lasing mechanism description above is idealized. In reality, the efficiency of the radiative transition between the upper and lower lasing levels is very low [7,8].

33.1.2 Recent Developments in High-Power QCLs

In recent years, considerable focus has been placed on improving the WPE and output power of QCLs for RT CW operation. Bai et al. [9] showed 8.4% WPE and 1.3 W output power around 4.6 μm in 2008. Shortly thereafter, Lyakh et al. [10] reported 12.7% WPE and 3 W power at 4.6 μm. Watt-level power with 6% WPE at 3.76 μm and then lower power at 3.39 and 3.56 μm are reported by Bandyopadhyay et al. [11,12]. Bai et al. [13] demonstrated 21% WPE and 5.1 W output power around 4.9 μm in 2011. Much higher WPE and/or output power has been achieved at lower temperatures or at pulsed mode [7,14] near 4.8 μm. A summary of recent developments can be found in review papers [8,15].

While good output powers and WPEs have been achieved, long-term reliability of these devices under RT CW operation remains a critical problem [5,6]. These devices are prone to catastrophic breakdown owing to reasons that are not entirely understood, but are likely related to thermal stress that stems from prolonged high-power operation [5]. This kind of thermal stress is worst in short-wavelength devices that have high strain and high thermal-impedance mismatch between layers [11,12,16,17].

In addition to improved device lifetime, we seek better CW temperature performance (higher characteristic temperatures T_0 and T_1, defined next) [6]. The first aspect is a weaker temperature dependence of the threshold current density. Empirically, the threshold current density (the current density at which the device starts lasing) has an exponential dependence on the operating temperature T: $J_{\text{th}} \propto \exp\left(\frac{T}{T_0}\right)$. Higher characteristic temperature T_0 is preferred in QCL design, as it means less variation in J_{th} as the temperature changes.

Another key temperature-dependent parameter is the differential quantum efficiency (also called the slope efficiency or external quantum efficiency), defined as the amount of output optical power dP per unit increase in the pumping current dI: $\eta_{\text{d}} = \frac{\text{d}P}{\text{d}I} \propto \exp\left(-\frac{T}{T_1}\right)$. The differential quantum efficiency is directly proportional to the WPE (WPE $= \eta_{\text{d}}\eta_{\text{f}}$, where η_{f} is the feeding efficiency). Therefore, the higher the T_1, the closer the η_{d} is to unity, and the higher the WPE. Recently, deep-well structures with tapered active regions have demonstrated significant improvements in T_0 and T_1 with respect to the conventional 4.6-μm device [9], underscoring that the suppression of leakage plays a key role in temperature performance [6,18,19]. Still, the microscopic mechanisms and leakage pathways that contribute to these empirical performance parameters remain unclear.

33.1.3 QCL Modeling: An Overview

Under high-power RT CW operation, both electron and phonon systems in QCLs are far away from equilibrium. In such nonequilibrium conditions, both electronic and thermal transport modeling are important for understanding and improving QCL performance.

Electron transport in both mid-IR and THz QCLs has been successfully simulated via semiclassical (rate equations [20–22] and Monte Carlo [23–26]) and quantum techniques (density matrix [27–33], nonequilibrium Green's functions (NEGF) [34–36], and lately Wigner functions [37]). InP-based mid-IR QCLs have been addressed via semiclassical [38] and quantum transport approaches (8.5-μm [39] and 4.6-μm [35,36] devices). There has been a debate whether electron transport in QCLs can be described using semiclassical models, in other words, how much of the current in QCLs is coherent. Theoretical work by Iotti and Rossi [23,40] shows that the steady-state transport in mid-IR QCLs is largely incoherent. Monte Carlo simulation [41] has also been used to correctly predict transport near threshold. However, short-wavelength structures [9] have pronounced coherent features, which cannot be addressed semiclassically

[33]. NEGF simulations accurately and comprehensively capture quantum transport in these devices but are computationally demanding. Density matrix approaches have considerably lower computational overhead than NEGF but are still capable of capturing coherent transport features. A comprehensive review of electron transport modeling was recently written by Jirauschek and Kubis [42].

Electronic simulations that ignore radiative transitions are applicable for modeling QCLs below or near threshold where the interaction between electrons and the laser electromagnetic field can be ignored. Such simulations are useful for predicting quantities such as threshold current density and T_0. However, in order to accurately model QCLs under lasing operations, the effect of the laser field on electronic transport would have to be included. In some cases, the effects of the laser field can be very strong [39], especially for high-WPE devices, where the field-induced current can be dominant [38]. When included in simulations, the laser field is typically modeled either as an additional scattering mechanism [38,43] or as a time-dependent sinusoidal electric field [39,44]. In this work, we ignore the effect of the laser field on electron dynamics.

Thermal transport in QCLs is often described through the heat diffusion equation, which requires accurate thermal conductivity in each region, a challenging task for the active core that contains many interfaces [17,45–48]. It is also very important to include nonequilibrium effects, such as the nonuniform heat generation rate stemming from the nonuniform temperature distribution [47] and the feedback that the nonequilibrium phonon population has on electron transport [26].

In this chapter, we present a multiphysics (coupled electronic and thermal transport) and multiscale (bridging between a single stage and device level) simulation framework that enables the description of QCL performance under far-from-equilibrium conditions [49]. We present the electronic (Section 33.2) and thermal (Section 33.3) transport models, and then bring them together for electrothermal simulation of a real device structure (Section 33.4). We strive to cover the basic ideas while pointing readers to the relevant references for derivation and implementation details.

33.2 Electronic Transport

Depending on the desired accuracy and computational burden, one can model electronic transport in QCLs with varying degrees of complexity. The goal is to determine the modal gain (proportional to the population inversion between the upper and lower lasing levels) under various pumping conditions (current or voltage) and lasing conditions (pulsed wave or CW). A typical electron transport simulator relies on accurately calculated quasibound electronic states and associated energies in the direction of confinement. Electronic wavefunctions and energies are determined by solving the Schrödinger equation or the Schrödinger equation combined with the Poisson's equation in highly doped systems. Section 33.2.3.1 introduces a $\mathbf{k} \cdot \mathbf{p}$ Schrödinger solver coupled with a Poisson solver. More information about other solvers for electronic states can be found in the review paper [42] and references therein.

The simulations of electronic transport fall into two camps depending on how the electron single-particle density matrix is treated. The diagonal elements of the density matrix represent the occupation of the corresponding levels and off-diagonal elements represent the "coherence" between two levels. Transport is semiclassical or incoherent when the off-diagonal coherences are much smaller than the diagonal terms and can be approximated as proportional to the diagonal terms times the transition rates between states [50]. In that case, the explicit calculation of the off-diagonal terms is avoided and only the diagonal elements are tracked, which simplifies the simulation considerably. However, when the off-diagonal terms are appreciable, transport is partially coherent and has to be addressed using quantum transport techniques.

33.2.1 Semiclassical Techniques

Semiclassical approaches assume that electronic transport between stages is largely incoherent "hopping" transport. The key quantities are populations of electronic states that are confined in the QCL growth

direction, and electrons transfer between them due to scattering events. The scattering rates can be obtained empirically or more rigorously via Fermi's golden rule. Common semiclassical approaches are the rate equations and ensemble Monte Carlo (EMC), the latter solving a Boltzmann-like transport equation stochastically.

33.2.1.1 Rate Equations

In the rate equation approach [20–22], scattering between relevant states, i.e., the injector level, the upper and lower lasing levels, and the ground state, is captured through transition rates. The rates include all relevant (radiative and nonradiative) scattering mechanisms and can be either empirical parameters or calculated [51,52]. The computational requirements of rate equation models are low, so they are suitable for fast numerical design and optimization of different structures [22].

33.2.1.2 Ensemble Monte Carlo

The heterostructure in the QCL active core is a quasi-two-dimensional (quasi-2D) system, where electrons are free to move in the x–y plane, while confined cross-plane in the z direction; the confinement results in the formation of quasibound states and discrete energy levels corresponding to the bottoms of 2D energy subbands. The electron wavefunctions in 3D are plane waves in the x–y plane and confined wavefunctions in z direction. Electronic transport is captured by a Boltzmann-like semiclassical transport equation [23], which can be solved via the stochastic EMC technique assuming instantaneous hops between states in 3D due to scattering [53]. The simulation explicitly tracks the energy level and in-plane momentum of each particle in the simulation ensemble (typically $\sim 10^5$ particles). Tracking in-plane dynamics makes it more detailed than the rate equation model. The transition rates are generally computed directly from the appropriate interaction Hamiltonians and therefore depend on the energy levels as well as the wavefunction overlaps between different electronic states [42,53,54]. EMC allows us to include nonequilibrium effect into transport, which is covered in more detail in Section 33.2.3.

33.2.2 Quantum Techniques

Density matrix and NEGF are the two most widely used techniques to describe quantum transport in QCLs. Recently, a Wigner function approach was also successfully used to model a SL [37].

33.2.2.1 Density Matrix Approaches

In semiclassical approaches, the central quantity of interest is the distribution function $f_n^{E_k}(t)$, the probability of an electron occupying an eigenstate n and having an in-plane kinetic energy E_k. The quantum-mechanical analogue is the single-electron density matrix, $\rho_{nm}^{E_k}(t)$, where the diagonal elements $\rho_{nn}^{E_k}(t) = f_n^{E_k}(t)$ are occupations and the off-diagonal elements $\rho_{nm}^{E_k}(t)$ are the spatial coherences between states n and m at the in-plane energy E_k. When employing semiclassical methods, off-diagonal matrix elements are assumed to be much smaller than diagonal elements. This approximation may fail in some cases, e.g., when two eigenstates with a large spatial overlap have similar energies. This scenario often arises when modeling THz QCLs [28,32,55], but can also come up in mid-IR QCLs [33]. In these cases, semiclassical models fail.

The density matrix models that have been employed for QCL modeling can be categorized into two groups. The first includes hybrid methods, where transport is treated semiclassically within a region of the device (typically a single stage) while the effects of tunneling between different regions, separated by barriers, are treated quantum mechanically using a density matrix formalism with phenomenological dephasing times [28,31,55]. The second group involves completely quantum-mechanical methods that rely on microscopically derived Markovian master equations that guarantee positivity of the density matrix [32,33]. Both methods are more computationally expensive than their semiclassical counterparts, because the density matrix contains many more elements than its diagonal semiclassical analogue.

33.2.2.2 Nonequilibrium Green's Functions

The NEGF technique (see a good overview in Reference [42]) relies on the relationships between single-particle time-ordered Green's functions and correlation functions [34–36]. The correlation function $G^{<}_{\alpha,\beta}(k; t_1, t_2)$, often referred to as the lesser Green's function [44,56], is one of the central quantities and can be understood as a two-time generalization of the density matrix, where k refers to the magnitude of in-plane wave vector. The correlation function contains both spatial correlations (terms with $\alpha \neq \beta$) as well as temporal correlations between times t_1 and t_2 (not included in semiclassical or density matrix models). Typically, the potential profile is assumed to be time independent, in which case the correlation function only depends on the time difference $G^{<}_{\alpha,\beta}(k; t_1, t_2) = G^{<}_{\alpha,\beta}(k; t_1 - t_2)$. Fourier transform over the time difference into the energy domain gives the energy-resolved correlation function $G^{<}_{\alpha,\beta}(k, E)$, which is the quantity that is usually solved for numerically [35,44,56]. The main advantages of the NEGF formalism are that it provides spectral (energy-resolved) information and it includes the effects of collisional broadening (the broadening of energy levels due to scattering), which is particularly important when the states are close in energy. These advantages carry a considerable computational cost, so NEGF calculations are much more time-consuming than density matrix approaches [39].

33.2.3 EMC with Nonequilibrium Phonons

Here, we focus on presenting semiclassical modeling of electron transport in QCL structures via EMC [25,26,57]. The solver consists of two parts: a coupled Schrödinger–Poisson solver and a transport kernel. We solve for the electronic states using the coupled Schrödinger–Poisson solver and feed the energy levels and the wavefunctions of the relevant electronic states to the transport kernel. The transport kernel keeps track of the electron momentum, energy, and distribution among subbands. If the electron density inside the device is high, transport kernel will periodically feed the electron distribution back to the Schrödinger–Poisson solver and update the electronic states. This loop is repeated until the electron distribution converges. By doing so, we solve for both the electron transport and the electronic band structure self-consistently.

Since the active QCL core consists of repeated stages, the wavefunctions in any stage can be obtained from the wavefunctions in any other stage by translation in space and energy. This translational symmetry makes it possible to simulate electron transport in only one generic central stage instead of in the whole QCL core [53]. Typically, electronic states in nonadjacent stages have negligible overlap, which also means that the transition rates between them are negligible. As a result, it is sufficient to limit interstage scattering events to only those between adjacent stages.

Figure 33.2 shows a schematic of three adjacent stages under an applied field. We simulate electron transport in the central stage λ, while nearest-neighbor interstage ($\lambda \rightleftarrows \lambda \pm 1$) and intrastage ($\lambda \rightarrow \lambda$) scattering is allowed. Periodic boundary conditions (PBCs) are applied in the simulation, i.e., whenever one electron scatters from the central stage into the next stage (process ②), an electron scatters from the previous stage into the central stage (process ①) and vice versa (process ③ and process ④). PBCs are justified by the cascading scheme.

33.2.3.1 Electronic Bandstructure Solver

We employ the **k**·**p** method to solve the Schrödinger equation and couple it to a Poisson solver [25,53,57]. The **k**·**p** method is an efficient way to solve for the electronic band structure near the band edges, where the transport happens in QCLs. The **k**·**p** method considers the contribution from the conduction band (C), light-hole band (LH), and the spin-orbit split-off band (SO) (the heavy-hole band (HH) decouples from the other three at the band edge) [58]. The contributions from the LH and SO are especially important for narrow-gap materials, such as InP. Moreover, in modern QCLs, strain-balanced structures have been employed to obtain enhanced performance. In these structures, alternate layers are compressively or tensilely strained while the whole structure is strain free, with carefully designed thickness of each

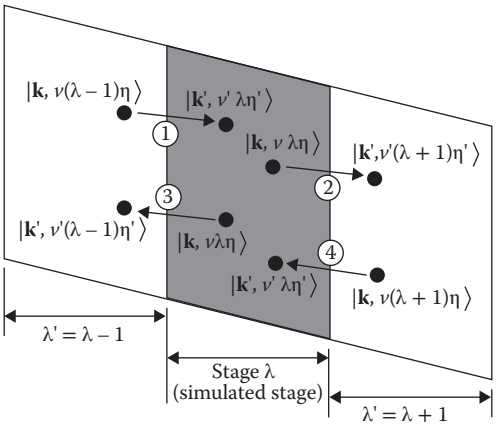

FIGURE 33.2 Schematic of the three simulated stages in a QCL active core under an applied field. Scattering is limited to nearest-neighbor stages and periodic boundary conditions are justified by the cascading scheme; therefore, only three stages are needed in the EMC transport kernel.

layer. The **k·p** method allows for convenient inclusion of the effects of strain on the band structure. The implementation details of the **k·p** solver can be found in Reference [53].

The **k·p** solver can only solve for a finite structure rather than an infinite periodic one. As a result, we need to simulate a finite number of stages and add artificially high barriers to the two ends to confine all the states. If a stage is far enough from the boundaries, the calculated band structure in it should be the same as if we were to solve for the whole periodic structure. Tests have confirmed that three stages, which we also used in EMC, are enough when solving for the electronic states to ensure that the central-stage states are unaffected by the simulation-domain potential boundaries. The states from the central stage are then translated in energy and position to the neighboring stages according to the stage length and the applied electric field.

When we need to solve for electron transport and electronic states self-consistently, it is necessary for the solver to be able to automatically pick out the electronic states belonging to the central stage. One intuitive criterion is to calculate the "center of mass" for each state (the expectation value of the cross-plane coordinate, $\langle z \rangle$) and assign those falling in the central stage to that stage. However, in our three-stage scheme, this method may pick up the states that are too close to the boundary. One can either extend the number of stages in the **k·p** solver to five, so the three stages in the middle are all far from the boundary, or use additional criteria such as that there be more than 50% possibility of finding an electron in the central stage, based on the probability density distribution, or requiring that the location of the probability density peak be in the central stage. Additional criteria requiring strong confinement of states have been explored in [32].

33.2.3.2 Transport Kernel with Nonequilibrium Phonons

The EMC kernel tracks the hopping transitions of electrons between subbands and stages until convergence and outputs the transport information for us to calculate the experimentally relevant quantities such as current and modal gain [53]. In the transport kernel, both electron–electron interactions and electron–LO-phonon interactions are considered. Other scattering processes, such as intervalley scattering, impurity scattering, and interface roughness scattering can be considered under different circumstances [42]. Photon emission is not considered, either. Because EMC tracks individual particles, nonequilibrium electron transport can be automatically captured. (EMC tracks individual simulation particles, each of which might represent thousands of real electrons.)

The most important scattering mechanism in QCLs is electron–LO-phonon scattering, which facilitates the depopulation of the lower lasing level. As shown in Reference [59], phonon confinement has little effect on the electronic transport; therefore, for simplicity, LO phonons are treated as bulklike dispersionless phonons with energy $\hbar\omega_0$. The transition rate between an initial state $\phi_i(z)$ with energy E_i and a final state $\phi_f(z)$ withenergy E_f can be derived from the Fermi's golden rule as follows:

$$\Gamma_{a(-),e(+)} = \frac{e^2\hbar\omega_0 m_f^*}{8\pi^2\hbar^3}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon_0}\right)\int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dq_z \int_0^{\infty} dE_{kf} N_{\mathbf{q}} \frac{|\mathcal{I}_{if}(q_z)|^2}{\mathbf{q}_\parallel^2 + q_z^2}\delta(E_f - E_i \mp \hbar\omega_0), \tag{33.1}$$

where e is the electronic charge while ϵ_0 and ϵ_∞ are static and high-frequency electronic permittivities of the material, respectively. The integrals are over the in-plane kinetic energy E_k of the final state and the cross-plane momentum transfer q_z. $\mathbf{q}_\parallel = \mathbf{k}'_\parallel - \mathbf{k}_\parallel$ is the in-plane momentum transfer.

$$|\mathcal{I}_{if}(q_z)|^2 = \left|\int_0^d dz \phi_f^*(z)\phi_i(z)e^{-izq_z}\right|^2 \tag{33.2}$$

is defined as the overlap integral (OI) between the initial and final states, where q_z is the cross-plane momentum transfer. The integration is over the angle between initial and final in-plane momenta $\mathbf{k}_\parallel$ and $\mathbf{k}'_\parallel$ (θ), cross-plane momentum component of the final state (k_{fz}), and the kinetic energy of the final state (E_{kf}). $N_{\mathbf{q}}$ represents the number of LO phonons with momentum $\mathbf{q} = (\mathbf{q}_\parallel, q_z)$. The expression can be further simplified in the equilibrium case, where $N_{\mathbf{q}}$ follows the Bose–Einstein distribution [53]. In order to model nonequilibrium phonon effects, we numerically integrated the expression using a phonon number histogram according to both $q_\parallel$ and q_z [54].

According to the uncertainty principle, position and momentum both cannot be determined simultaneously. Since our electrons are all confined in the central stage ($\Delta z'$ is finite), the cross-plane momentum is not exactly conserved during the scattering process ($q_z \neq k'_z - k_z$) [60]. This analysis does not affect the momentum conservation in the x–y plane, because we assume infinite uncertainty in position there. Previously, the cross-plane momentum conservation has been considered through the momentum conservation approximation (MCA) [61,62] and a broadening of q_z according to the well width [60]. The MCA forbids a phonon emitted between subbands i and f to be reabsorbed by another transition between i' and f' if $i \neq i'$ or $f \neq f'$, and thus might underestimate the electron–LO interaction strength [54]. The concept of well width is hard to apply in an MQW structure such as the QCL active core [54]. We observe that the probability of a phonon with cross-plane momentum q_z being involved in an interaction is proportional to the OI in Equation 33.2. Figure 33.3 depicts the typical OIs for both intersubband ($i_1 \rightarrow 3$ and $2 \rightarrow 1$) and intrasubband ($3 \rightarrow 3$) transitions. As a result, in each electron–LO-phonon scattering event, we randomly select a q_z following the distribution from the OI (Figure 33.3). Depending on the mechanism (absorption or emission), a phonon with $(\mathbf{q}_\parallel, q_z)$ is removed or added to the histogram according to the 2D density of states (DOS) and the effective simulation area [54]. Once the phonons with a certain momentum are depleted, transitions involving such phonons become forbidden.

In order to couple the EMC solver to the thermal transport solver, we need to keep a detailed log of heat generation during electron transport. In all the relevant scattering events, electron–LO-phonon scattering is the only inelastic mechanism and therefore is the only mechanism that contributes to heat generation. As a result, the total energy emitted and absorbed in the form of LO phonons is recorded during each step of the EMC simulation. The nonequilibrium phonons decay into acoustic longitudinal acoustic (LA) phonons via a three-phonon anharmonic decay process. The formulation and the parameters here follow [63]. The simulation results of EMC including nonequilibrium phonons are shown in Section 33.4.

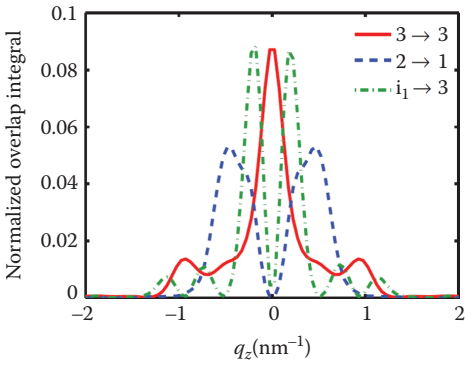

FIGURE 33.3 Normalized overlap integral $|\mathcal{I}_{if}|^2$ from Equation 33.2 versus cross-plane phonon wave vector q_z for several transitions (intersubband $i_1 \rightarrow 3$ and $2 \rightarrow 1$; intrasubband $3 \rightarrow 3$). (Reprinted with permission from Y. B. Shi and I. Knezevic. Nonequilibrium phonon effects in midinfrared quantum cascade lasers. *Journal of Applied Physics,* 116(12):123105, 2014. Copyright 2014, American Institute of Physics.)

33.3 Thermal Transport

The dominant path of heat transfer in a QCL structure is depicted in Figure 33.4. The operating electric field of a typical QCL is high, which means that considerable energy is pumped into the electronic system. These energetic, "hot" electrons relax their energy largely by emitting LO phonons. LO phonons have high energies but flat dispersions, so their group velocities are low and they are poor carriers of heat. An LO phonon decays into two LA phonons via a three-phonon process referred to as anharmonic decay. LA phonons have low energy but high group velocity and are the main carriers of heat in semiconductors [47,63]. If we neglect the diffusion of optical phonons, the flow of energy in a QCL can be described by the following equations:

$$\frac{\partial W_A}{\partial t} = \nabla \cdot (\kappa_A \nabla T_A) + \left.\frac{\partial W_{LO}}{\partial t}\right|_{coll}; \quad \frac{\partial W_{LO}}{\partial t} = \nabla \cdot (\kappa_A \nabla T_A) + \left.\frac{\partial W_e}{\partial t}\right|_{coll} - \left.\frac{\partial W_{LO}}{\partial t}\right|_{coll}, \tag{33.3}$$

where W_{LO}, W_A, and W_e are the LO phonon, acoustic phonon, and electron energy densities, respectively. κ_A is the thermal conductivity in the system and T_A is the acoustic phonon (lattice) temperature. The term $\nabla \cdot (\kappa_A \nabla T_A)$ describes heat diffusion, governed by acoustic phonons. We have also used the fact that the rate of increase in the LO-phonon energy density equals the difference between the rate of its generation by electron–LO-phonon scattering and the rate of anharmonic decay into LA phonons.

In a nonequilibrium steady state, both the LO and LA energy densities are constant, so

$$-\nabla \cdot (\kappa_A \nabla T_A) = \left.\frac{\partial W_e}{\partial t}\right|_{coll}. \tag{33.4}$$

As described in the previous section, the right-hand side of Equation 33.4 is the heat generation rate Q and can be obtained by recording electron–LO-phonon scattering events in electronic EMC [47,64]:

$$Q = \left.\frac{\partial W_e}{\partial t}\right|_{coll} = \frac{N_{3D}}{N_{sim} t_{sim}} \sum (\hbar\omega_{ems} - \hbar\omega_{abs}), \tag{33.5}$$

where $N_{3D} = N_s / D_{stage}$ is the electron density (N_s is the sheet density and D_{stage} is the length of a single stage) while N_{sim} and t_{sim} are the number of simulation particles and the simulation time, respectively,

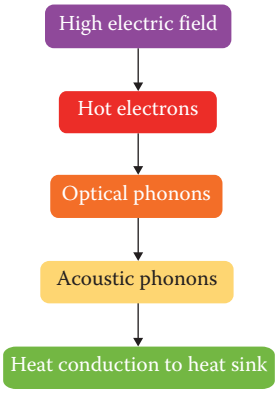

FIGURE 33.4 Flow of energy in a QCL.

and $\hbar\omega_{ems}$ and $\hbar\omega_{abs}$ are the energies of the emitted and absorbed LO phonons, respectively. To solve Equation 33.4, we need information on both the thermal conductivity κ_A and the heat generation rate Q; they are discussed in Subsections 33.3.1 and 33.3.2, respectively.

33.3.1 Thermal Conductivity in a QCL Device

33.3.1.1 Active Core: A III-V Superlattice

The QCL active core is an SL: It contains many identical stages, each with several thin layers made from different materials and separated by heterointerfaces. The thermal conductivity tensor of an SL system reduces to two values: the in-plane thermal conductivity $\kappa_{\|}$ (in-plane heat flow is assumed isotropic) and the cross-plane thermal conductivity $\kappa_{\perp}$. Experimental results have shown that, in SLs, the thermal conductivity is very anisotropic [65] ($\kappa_{\|} \gg \kappa_{\perp}$) while both $\kappa_{\|}$ and $\kappa_{\perp}$ are smaller than the weighted average of the constituent bulk materials [66–70]. Both effects can be attributed to the interfaces between adjacent layers [71,72].

Here, we discuss a semiclassical model for describing the thermal conductivity tensor of III-V SL structures. Note that the model described here is in principle applicable to SLs in other material systems as long as they have high-quality interface and thermal transport is mostly incoherent [48,73–75]. In particular, we focus on thermal transport in III-arsenide-based SLs, as they are most commonly used in mid-IR QCL active cores [48].

Under QCL operation conditions of interest (>77 K, and typically near RT), thermal transport is dominated by acoustic phonons and is governed by the Boltzmann transport equation (BTE). To obtain the thermal conductivity, we solve the phonon BTE with full phonon dispersion in the relaxation-time approximation [48].

33.3.1.2 Twofold Influence of Effective Interface Roughness

To capture both the anisotropic thermal transport and the reduced thermal conductivity in SL systems, we need to observe the twofold influence of the interface. First, it reduces $\kappa_{\|}$ by affecting the acoustic phonon population close to the interfaces [76]. Second, it introduces an interface thermal boundary resistance (ITBR), which is still very difficult to model [65,77]. Common models are the acoustic mismatch model (AMM) and the diffuse mismatch model (DMM) [65,76]; the former assumes a perfectly smooth interface and only considers the acoustic mismatch between the two materials, while the latter assumes complete randomization of momentum after phonons hit the interface. As most III-V-based QCLs are

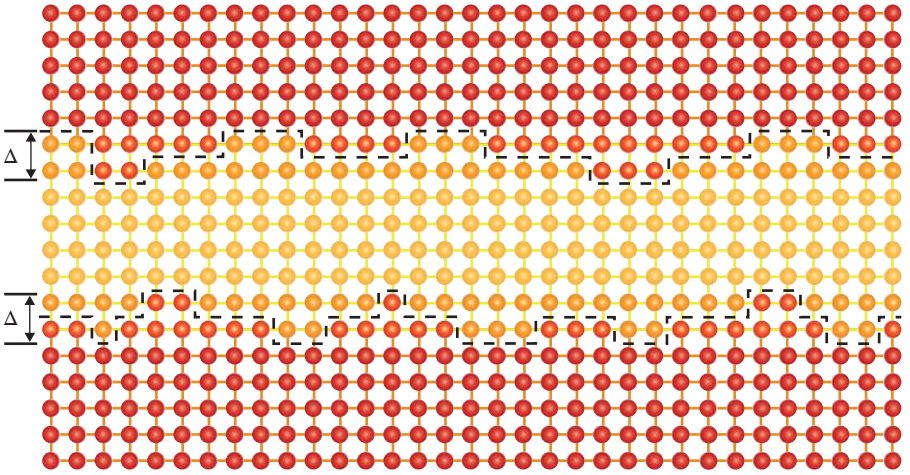

FIGURE 33.5 Even between lattice-matched crystalline materials, there exist nonuniform transition layers that behave as an effective atomic-scale interface roughness with some rms roughness Δ. This effective interface roughness leads to phonon momentum randomization and to interface resistance in cross-plane transport. (From S. Mei and I. Knezevic, *Journal of Applied Physics*, 118, 175101, 2015. With the permission of AIP Publishing.)

grown by MBE or MOCVD, both well-controlled techniques allowing consistent atomic-level precision, neither AMM nor DMM captures the essence of a III-V SL interface. Figure 33.5 shows a schematic of interface roughness in a lattice-matched SL. The jagged dashed boundaries depict transition layers of characteristic thickness Δ between the two materials.

We introduce a simple model that calculates a more realistic ITBR (a key part in calculating $\kappa_\perp$) by interpolating between the AMM and DMM transmission rates using a specularity parameter p_{spec}. The model has a single fitting parameter: the effective interface rms roughness Δ. Since the growth environment is well controlled, using one Δ to describe all the interfaces is justified. We use Δ to calculate a momentum-dependent specularity parameter:

$$p_{\text{spec}}(\vec{q}) = \exp(-4\Delta^2|\vec{q}|^2\cos^2\theta), \tag{33.6}$$

where $|\vec{q}|$ is the magnitude of the phonon wave vector and θ is the angle between $\vec{q}$ and the normal direction to the interface. Consistent with the twofold impact of interface roughness, Δ affects the thermal conductivity through two channels. Apart from calculating the ITBR, an effective interface scattering rate $\tau^{-1}_{\text{interface}}(\vec{q})$ dependent on the same specularity parameter $p_{\text{spec}}(\vec{q})$ is added to the internal scattering rate to calculate modified $\kappa_\parallel$ (see detailed derivations in [48]). By adjusting only Δ, typically between 1 and 2 Å, the calculated thermal conductivity using this model fits a number of different experiments [66,68,69].

33.3.1.3 $\kappa_\parallel$ and $\kappa_\perp$ of a QCL Active Core

Thermal transport inside the active core of a QCL is usually treated phenomenologically: $\kappa_\parallel$ is typically assumed to be 75% of the weighted average of the bulk thermal conductivities of the constituent materials, while $\kappa_\perp$ is treated as a fitting parameter (constant for all temperatures) to best fit the experimentally measured temperature profile [17,78]. We calculated the thermal conductivity tensor of a QCL active core [78] and showed that the typical assumption is not accurate and that the degree of anisotropy is temperature dependent (Figure 33.6).

The ratio between $\kappa_\parallel$ and the averaged bulk value (inset to Figure 33.6) varies between 45% and 70% over the temperature range of interest. $\kappa_\perp$ has a weak dependence on temperature in keeping with the

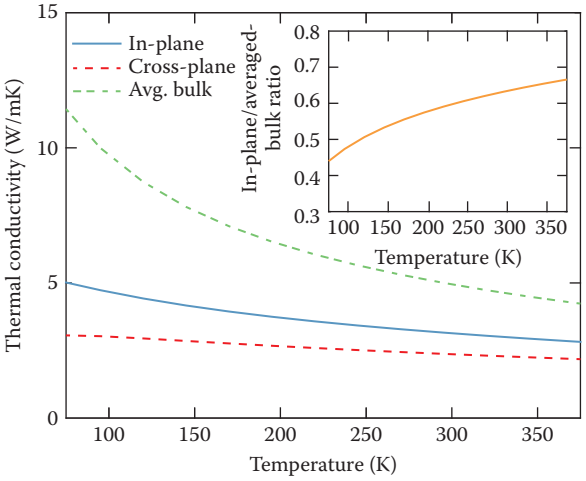

FIGURE 33.6 Thermal conductivity of a typical QCL active region [78] as a function of temperature. A single stage consists of 16 alternating layers of $In_{0.53}Ga_{0.47}As$ and $In_{0.52}Al_{0.48}As$. The solid curve, dashed curve, and dashed-dotted curve show the calculated in-plane, cross-plane, and averaged bulk thermal conductivity, respectively. $\Delta = 1$ Å in the calculations. The inset shows the ratio between the calculated in-plane and the averaged bulk thermal conductivities. (From S. Mei and I. Knezevic, *Journal of Applied Physics*, 118, 175101, 2015. With the permission of AIP Publishing.)

TABLE 33.1 Thermal Conductivity as a Function of Temperature for Materials in a QCL Structure

Materials	Thermal Conductivity (W/mK)
Au	$337–600 \times 10^{-4} T$
Si_3N_4	$30–1.4 \times 10^{-2} T$
In solder	$93.9–6.96 \times 10^{-2} T + 9.86 \times 10^{-5} T^2$

common assumption in simplified models; the weak temperature sensitivity means that ITBR dominates cross-plane thermal transport. These results show that it is important to carefully calculate the thermal conductivity tensor in QCL thermal simulation, and we use this thermal conductivity model in the device-level simulation.

33.3.1.4 Other Materials

The active core is not the only region we need to model in a device-level thermal simulation. Figure 33.7 shows a typical schematic (not to scale) of a QCL device in thermal simulation with a substrate-side mounting configuration [79]. The active core (in this case, consisting of 36 stages and 1.6-μm thick) with width W_{act} is embedded between two cladding layers (4.5-μm-thick GaAs). The waveguide is supported by a substrate (GaAs) with thickness D_{sub}. An insulation layer (Si_3N_4) with thickness D_{ins} is deposited around the waveguide and then etched away from the top to make the contact. Finally, a contact layer (Au) with thickness D_{cont} and a thin layer of solder (D_{sold}) are deposited on top. There is no heat generation in the regions other than the active core. Further, these layers are typically thick enough to be treated as bulk materials. Bulk substrate (GaAs or InP) thermal conductivities are readily obtained for III-V materials from experiment, as well as from relatively simple theoretical models [17,48,78,80] (Table 33.1).

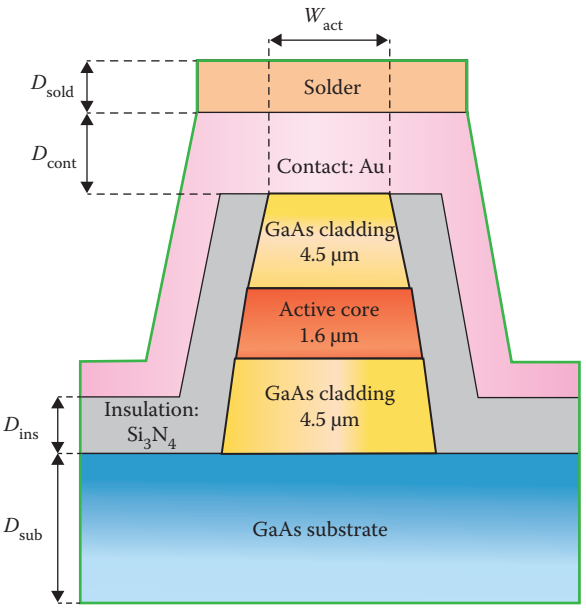

FIGURE 33.7 Schematic of a typical GaAs-based mid-IR QCL structure with a substrate (not to scale).

33.3.2 Device-Level Electrothermal Simulation

33.3.2.1 Device Schematic

The length of a QCL device is much greater than its width, therefore we can assume the length is infinite and carry out a 2D thermal simulation. The schematic of the simulation domain (not to scale) is shown in Figure 33.7. The boundary of the simulation region is delineated, and certain boundary conditions (heat sink at fixed temperature, convective boundary condition, or adiabatic boundary condition) can be applied (independently) to each boundary. Typically, the bottom boundary of the device is connected to a heat sink while other boundaries have the convective boundary condition at the environment temperature (single-device case) or the adiabatic boundary condition (QCL array case). Typical values for the layers thickness are $W_{act} = 15$ μm, $D_{sub} = 50$ μm, $D_{ins} = 0.3$ μm, $D_{cont} = 3$ μm, and $D_{sold} = 1.5$ μm.

We use the finite element method to solve for the temperature distribution. The whole device is divided into different regions according to their materials properties. Each stage of the active region is treated as a single unit with the heat generation rate tabulated in the device table in order to capture the nonuniform behavior among stages. The active core is very small but is also the only region with heat generation, small thermal conductivity, and spatial nonuniformity. To capture the behavior of the active region while saving computational time, we use a nonuniform mesh in the finite element solver to emphasize the active core region. Figure 33.8 shows a mesh generated in the simulation.

33.3.2.2 Simulation Algorithm

It is known that among all the stages in the active core, the temperature T_i and the electric field F_i (i represents the stage index) are not constant [56,78], but we have no *a priori* knowledge of how they depend on the stage index. However, we know that the charge–current continuity equation must hold, and in the steady state $\nabla \cdot \mathbf{J} = 0$; this implies that the current density J must be uniform, as the current flow is essentially in one dimension, along z. This insight is key to bridging the single-stage and device-level simulations.

From Section 33.2.3, we can obtain the heat generation rate Q inside the active core by running the single-stage EMC simulation. Each single-stage EMC is carried out at a specific electric field F and

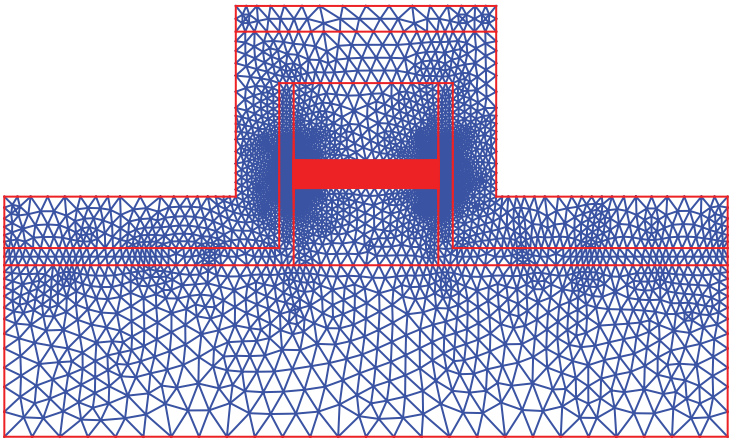

FIGURE 33.8 A typical nonuniform finite-element mesh of the simulated GaAs-based mid-IR QCL structure.

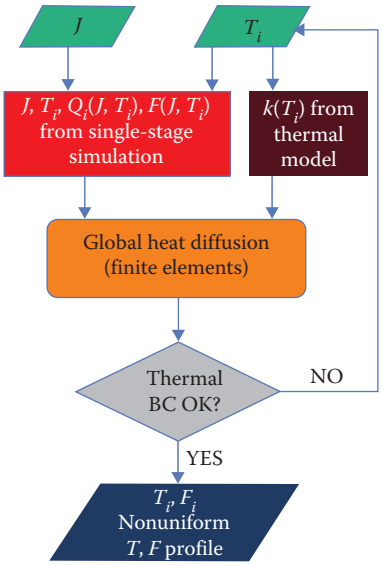

FIGURE 33.9 Flowchart of the device-level thermal simulation. We start by assuming a certain current density J and temperature profile T_i across the whole device. Based on the tabulated information from the single-stage simulation and assumed (J, T_i), we get stage-by-stage profiles for the electric field F_i and the heat generation rate Q_i. An accurate temperature-dependent thermal conductivity model, which includes the boundary resistances of layers, and the temperature profile guess are used as input to the heat diffusion equation, which is then iteratively solved (with updated temperature profile in each step) until the thermal boundary conditions are satisfied.

temperature T and outputs both the current density $J(F, T)$ and the heat generation rate $Q(F, T)$. By sweeping F and T in range of interest, we obtain a table connecting different field and temperature (F, T) to appropriate current density and heat generation rate (J, Q) $[(F, T) \rightarrow (J, Q)]$. However, based on the discussion above, the input in the thermal simulation needs to be the constant parameter J. Therefore, we "flip" the recorded $(F, T) \rightarrow (J, Q)$ table to a so-called device table $(J, T) \rightarrow (Q, F)$, suitable for coupled simulation [49].

Figure 33.9 depicts the flowchart of the device-level electrothermal simulation [49]. Before the simulation, we obtain the device table $[(J, T) \rightarrow (Q, F)]$, as discussed above. We also have to calculate

the thermal conductivities ($\kappa_{\parallel}$ and $\kappa_{\perp}$) of the active region as a function of temperature and tabulate them, based on the model described in Section 33.3.1. We also need the bulk thermal conductivity of other materials in the device (cladding layer, substrate, insulation, contact, and solder) as a function of temperature. These material properties are standard and already well characterized.

Each device-level thermal simulation is carried out in a certain environment (i.e., for a given set of boundary conditions) and with a certain current density J. At the beginning of the simulation, an initial temperature profile is assigned. With the input from the device table and the thermal conductivity data in each region, we use a finite element method to iteratively solve the heat diffusion equation until convergence. At the end of the simulation, we obtain a thermal map of the whole device. Further, from the temperature T_i in each stage and the injected current density J, we obtain the nonuniform electric field distribution F_i. With the electric field in each stage and given the stage thickness, we can accurately calculate the voltage drop across the device and obtain the current–voltage characteristic. By changing the mounting configuration (W_{act}, D_{sub}, D_{ins}, D_{cont}, D_{sold}) or the boundary conditions, the temperature profile can be changed.

33.4 Device-Level Electrothermal Simulation: An Example

In this section, we present detailed simulation results of a 9-μm GaAs/$Al_{0.45}Ga_{0.55}As$ mid-IR QCL [79] based on a conventional three-well active region design. The chosen structure has 36 repetitions of the single stage; each stage has 16 layers. Starting from the injection barrier, the layer thicknesses in one stage (in Å) are **46**/19/**11**/54/**11**/48/**28**/34/**17**/30/**18**/28/**20**/30/**26**/30. Here, the barriers ($Al_{0.45}Ga_{0.55}As$) are in bold while the wells (GaAs) are in normal font; the underlined layers are doped to a sheet density of $n_{Si} = 3.8 \times 10^{11}$ cm^{-2}. The results at 77 K are shown here.

33.4.1 Electronic Simulation Results

33.4.1.1 Band Structure

Figure 33.10 shows the electronic states of the chosen structure under the designed operating field of 48 kV/cm, calculated from the coupled **k**·**p**–Poisson solver (see Section 33.2.3.1). The active region states of the central stage are represented by dark solid curves 1, 2, and 3 are the ground state and the lower

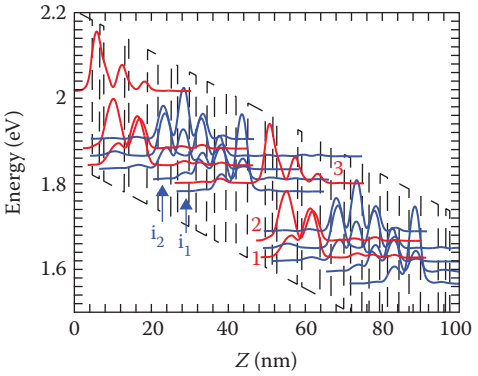

FIGURE 33.10 Energy levels and wavefunction moduli squared of Γ-valley subbands in two adjacent stages of the simulated GaAs/AlGaAs-based structure. The dark solid curves denote the active region states (1, 2, and 3 represent the ground state and the lower and upper lasing levels, respectively). The lighter solid curves represent injector states, with i_1 and i_2 denoting the lowest two. (From Y. B. Shi and I. Knezevic, *Journal of Applied Physics*, 116, 123105, 2014. With the permission of AIP Publishing.)

and upper lasing levels, respectively. Injector states are labeled i_1 and i_2. Other lightly colored solid curves denote the states that form the miniband. (When the electron density in the QCL is high, the electronic bands have to be calculated self-consistently with EMC.)

33.4.1.2 *J-F* Curve

The current density J versus field F curve, one of the key QCL characteristics at a given temperature, is intuitive to obtain in EMC. After calculating the electronic band structure at a certain field F, the wavefunctions, energy levels, and effective masses of each subband and each stage are fed into the EMC solver. In the EMC simulation, we include all the scattering mechanisms described in Section 33.2.3. Since we employ PBCs, the current density J can be extracted from how many electrons cross the stage boundaries in a certain amount of time in the steady state. The net flow n_{net} of electrons is calculated by subtracting the flow between the central stage and the previous stage (n_{backward}) from the flow between the central stage and the next stage (n_{forward}) in each time step. The current density is then calculated as follows:

$$J = \frac{en_{\text{net}}}{A_{\text{eff}}\delta t} = \frac{e(n_{\text{forward}} - n_{\text{backward}})}{A_{\text{eff}}\delta t}, \tag{33.7}$$

where δt is the time interval during which the flow is recorded. A_{eff} is the effective in-plane area of the simulated device. Since doping is the main source of electrons, the area is calculated as follows:

$$A_{\text{eff}} = \frac{N_{\text{ele}}}{N_{\text{s}}}, \tag{33.8}$$

where N_{ele} is the number of simulated electrons and N_{s} is the sheet doping density (in cm^{-2}) in the fabricated device. In the current simulation, $N_{\text{ele}} = 50,000$ and $N_{\text{s}} = 3.8 \times 10^{11}\ \text{cm}^{-2}$.

Because of the stochastic nature of EMC, we need to average the current density over multiple time steps. In practice, one can record the net cumulative number of electrons per unit area that leave a stage over time and obtain a linear fit to this quantity in the steady state; the slope yields the steady-state current density.

From each individual simulation, we extract the current density at a given electric field and temperature. To obtain the *J-F* curve at that temperature, we sweep the electric field. To demonstrate the importance of including nonequilibrium phonons effects, we carry out the simulation with thermal phonons alone and with both thermal and excess nonequilibrium phonons. Figure 33.11 is the *J-F* curve for the simulated structure with (filled squares) and without (empty squares) nonequilibrium phonons at 77 K. It can be seen that the current density at a given field considerably increases when nonequilibrium phonons are included and the trend holds up to 60 kV/cm. This difference is prominent at low temperatures (<200 K) and goes away at RT [26].

33.4.1.3 Modal Gain (G_{m}) and Threshold

We calculate the modal gain as follows [22]:

$$G_{\text{m}} = \frac{4\pi e^2 \langle z_{32}\rangle^2 \Gamma_{\text{w}} \Delta n}{2\varepsilon_0 \underline{n} \gamma_{32} L_p \lambda}, \tag{33.9}$$

where ε_0 is the permittivity of free space. Some constants are obtained from experiment: waveguide confinement factor $\Gamma_{\text{w}} = 0.31$, stage length $L_p = 45$ nm, optical-mode refractive index $\underline{n} = 3.21$, and full width at half maximum $\gamma_{32}(T_{\text{L}}) \approx 8.68\ \text{meV} + 0.045\ \text{meV/K} \times T_{\text{L}}$ [26,79]. The dipole matrix element between the upper and lower lasing levels ($\langle z_{32}\rangle = 1.7$ nm) and the emission wavelength ($\lambda = 9\ \mu$m) are

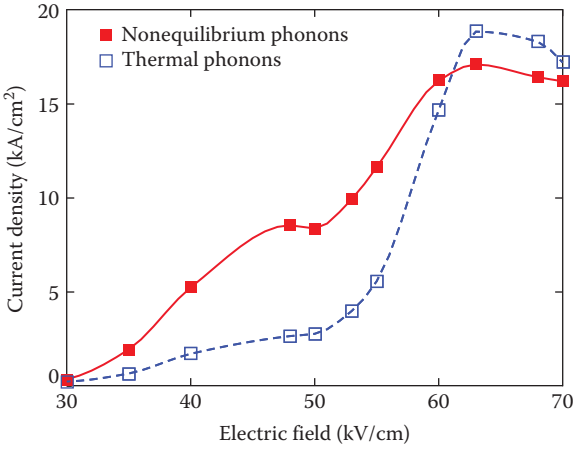

FIGURE 33.11 The current density versus electric field ($J - F$) curve of the simulated device with (filled squares) and without (empty squares) the nonequilibrium phonon effect at 77 K. The inclusion of nonequilibrium phonons considerably increases the current density at a given field up to 60 kV/cm.

also estimated in experiment [79], but we calculate these two terms directly. The dipole matrix element is calculated as follows:

$$\langle z_{32} \rangle = \int_0^d z\varphi_3^*(z)\varphi_2(z)dz\,. \tag{33.10}$$

The value is slightly different at different fields, as the band structure changes. At 48 kV/cm, the calculated matrix element is $\langle z_{32} \rangle = 1.997$ nm. Similarly, the wavelength of emitted photon also changes at different fields. One can calculate the value from the energy difference between the upper and lower lasing levels. The calculated wavelength at 48 kV/cm is 8.964 μm. $\Delta n = n_{\text{upper}} - n_{\text{lower}}$ is the population inversion obtained from EMC. Again, due to the randomness of EMC, the population inversion needs to be averaged over a period after the steady state has been reached.

Figure 33.12 shows the modal gain of the device with nonequilibrium (filled squares) and thermal (empty squares) phonons as a function of (1) electric field and (2) current density at 77 K. Horizontal dotted line indicates the total estimated loss in the device, which is used to help find the threshold current density, J_{th}. Lasing threshold is achieved when the modal gain G_{m} equals the total loss α_{tot}. We consider two sources of loss, mirror (α_{m}) and waveguide (α_{w}), so the total loss is $\alpha_{\text{tot}} = \alpha_{\text{m}} + \alpha_{\text{w}}$. The intercepts between the total loss line and the G_{m} versus F (Figure 33.12a) and G_{m} versus J (Figure 33.12b) curves give the threshold field F_{th} and threshold current density J_{th}, respectively. Like the current density, the modal gain of the device is also considerably higher when nonequilibrium phonons are considered, which leads to a lower F_{th} and a lower J_{th}. The reason for the increased current density and modal gain with nonequilibrium phonons can be attributed to the enhanced injection selectivity and efficiency [26].

33.4.1.4 Heat Generation Rate

The way to obtain the heat generation rate Q is similar to how we get the current density J. We record the cumulative net energy emission as a function of time and fit a straight line to the region where the simulation has reached a steady state. The slope of the line is used in place of $\sum(\hbar\omega_{\text{ems}} - \hbar\omega_{\text{abs}})/t_{\text{sim}}$. Figure 33.13 shows the heat generation rate as a function of electric field at 77 K. The filled squares and the empty squares depict the situation with and without nonequilibrium phonons, respectively.

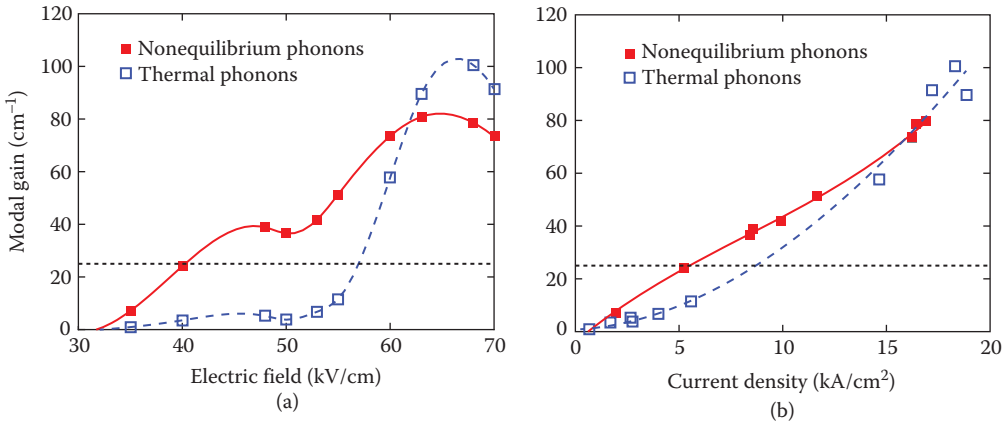

FIGURE 33.12 The modal gain (G_m) of the simulated device with nonequilibrium (filled squares) and thermal (empty squares) phonons as a function of (a) applied electric field and (b) current density J at 77 K. Horizontal dotted line shows the total estimated loss of the device.

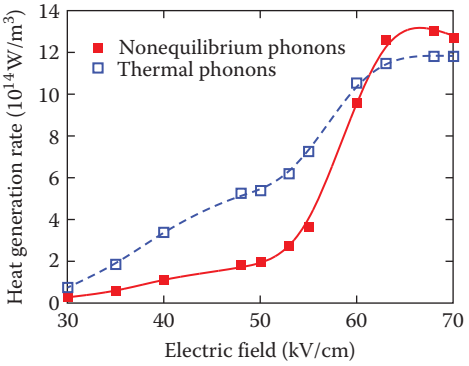

FIGURE 33.13 The heat generation rate of the simulated device as a function of electric field F at 77 K with (filled squares) and without (empty squares) nonequilibrium phonons.

33.4.2 Representative Electrothermal Simulation Results

This section serves to illustrate how the described simulation is implemented in practice and what type of information it provides at the single-stage and device levels.

First, the single-stage-coupled simulation has to be performed at different temperatures, as in Figure 33.14a. We noted the calculated J-F curves show a negative differential conductance region, which is typical for calculations, but generally not observed in experiment. Instead, a flat J-F dependence is typically recorded [30]. At every temperature and field, we also record the heat generation rate, as depicted in Figure 33.14b.

Second, the thermal model for the whole structure is developed. Considering that growth techniques improve over time, structures grown around the same time period should have similar properties. Since the device studied here was built in 2001 [79], we assume the active core should have similar effective rms roughness Δ to other lattice-matched GaAs/AlAs SLs built around the same time [69,70]. From our

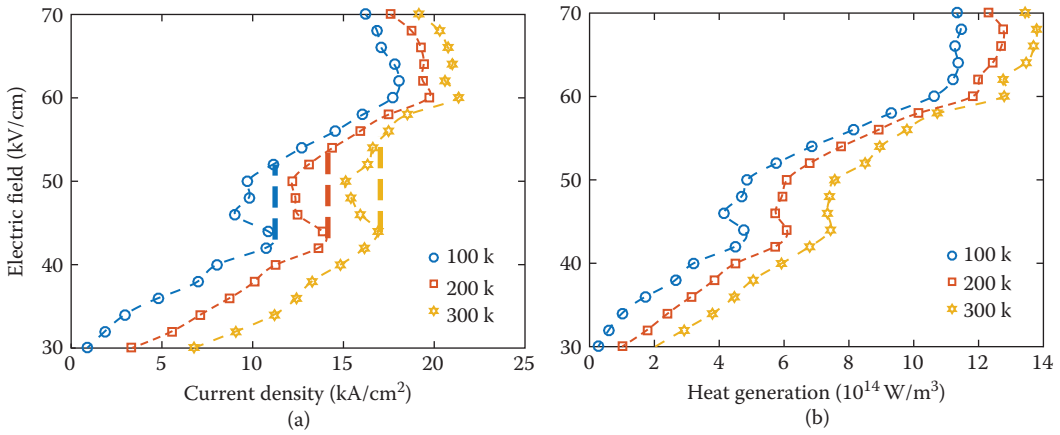

FIGURE 33.14 The field versus current density (a) and heat generation rate versus current density (b) characteristics for the simulated device at 100, 200, and 300 K, as obtained from single-stage simulation with nonequilibrium phonons.

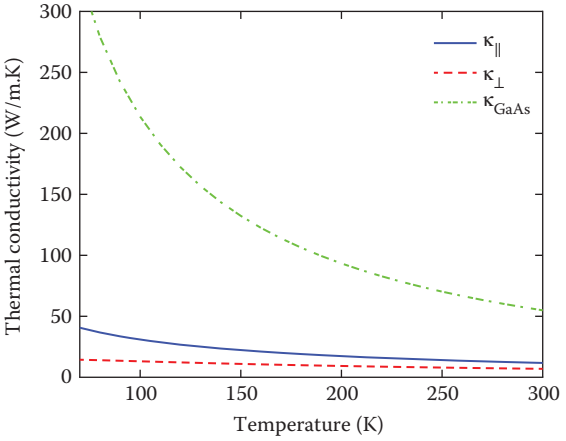

FIGURE 33.15 Calculated in-plane ($\kappa_{\parallel}$; solid line) and cross-plane ($\kappa_{\perp}$; dashed line) thermal conductivities of the active core along with the bulk thermal conductivity of the GaAs substrate (dash-dotted line). The effective rms roughness Δ is taken to be 5 Å.

previous simulation work on fitting the SL thermal conductivities [48], we choose an effective rms roughness $\Delta = 5$ Å in this calculation. Figure 33.15 shows the calculated thermal conductivities $\kappa_{\parallel}$ (solid line) and $\kappa_{\perp}$ (dashed line) along with the calculated bulk thermal conductivity (dash-dotted line) for the substrate GaAs.

The structure we considered operated in pulsed mode at 77 K. Depending on the duty cycle, the temperature distribution in the device can differ considerably. Figure 33.16 depicts a typical temperature profile across the device, while Figure 33.17 depicts the profile across the active core alone at duty cycles of 100% (essentially CW lasing, if the device achieved it) and 0.01% (as in experiment [79]). Clearly, CW operation would result in dramatic heating of the active region. Finally, Figure 33.18 shows the J–V curve of the entire simulated device at 77 K with duty cycles of 0.01%, 100%, and as observed in experiment [79].

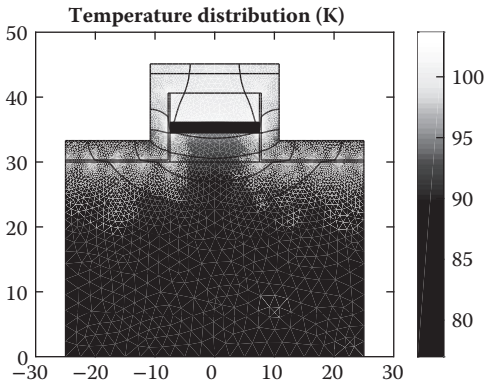

FIGURE 33.16 A typical temperature profile across the structure. At the bottom of the device is a heat sink held at 77 K, while adiabatic boundary conditions are applied elsewhere. The current density is 6 kA/cm^2 and the duty cycle is 100%.

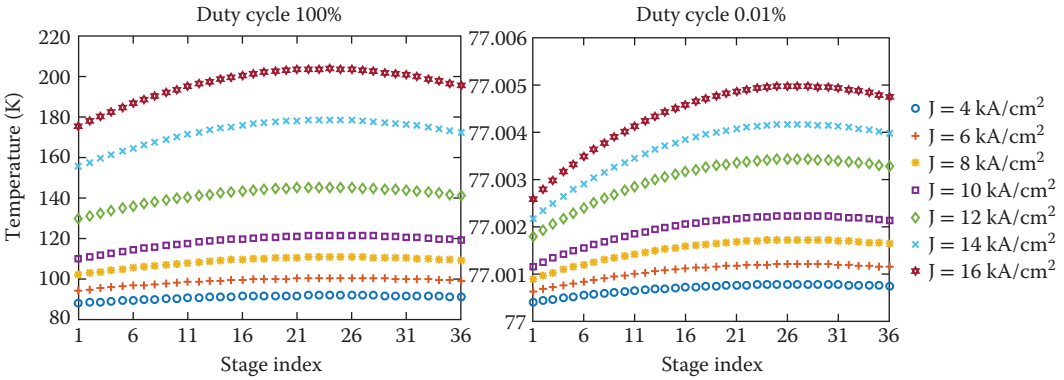

FIGURE 33.17 Temperature profile inside the active region at 100% duty cycle (left) and 0.01% duty cycle (right) for the QCL from [79].

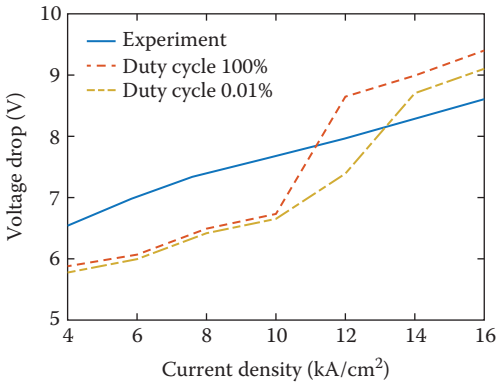

FIGURE 33.18 The current density versus voltage drop for the simulated device in experiment (solid curve) and as calculated at 100% (dashed curve) and 0.01% (dot-dashed curve) duty cycles. The bottom of the device is placed on a heat sink held at 77 K while adiabatic boundary conditions are assumed on the rest of the boundaries (see Section 33.3.2).

33.5 Conclusion

We overviewed electronic and thermal transport simulation of QCLs, as well as recent efforts in device-level electrothermal modeling of these structures, which is appropriate for transport below threshold, where the effects of the optical field are negligible. We specifically focused on mid-IR QCLs in which electronic transport is largely incoherent and can be captured by the EMC technique. The future of QCL modeling, especially for near-RT CW operation, will likely include improvements on several fronts: (1) further development of computationally efficient yet rigorous quantum transport techniques for electronic transport to fully account for coherent transport features that are important in short-wavelength mid-IR devices; (2) a better understanding and better numerical models for describing the role of electron–electron interaction, impurities, and interface roughness on device characteristics; and (3) holistic modeling approaches in which electrons, phonons, and photons are simultaneously and self-consistently captured within a single simulation. The goal of QCL simulation should be nothing less than excellent predictive value of device operation across a range of temperatures and biasing conditions, along with unprecedented insight into the fine details of exciting nonequilibrium physics that underscores the operation of these devices.

Acknowledgment

The authors gratefully acknowledge support by the U.S. Department of Energy, Basic Energy Sciences, Division of Materials Sciences and Engineering, Physical Behavior of Materials Program, Award No. DE-SC0008712. The work was performed using the resources of the UW-Madison Center for High Throughput Computing (CHTC).

References

1. J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, and A. Y. Cho. Quantum cascade laser. *Science*, 264(5158):553–556, 1994.
2. R. A. Suris and R. F. Kazarinov. Possibility of the amplification of electromagnetic waves in a semiconductor with a superlattice. *Soviet Physics: Semiconductors*, 5(4):707–709, 1971.
3. K. Y. Cheng. Molecular beam epitaxy technology of III-V compound semiconductors for optoelectronic applications. *Proceedings of the IEEE*, 85(11):1694–1714, 1997.
4. K. H. Goetz, D. Bimberg, H. Jürgensen, J. Selders, A. V. Solomonov, G. F. Glinskii, and M. Razeghi. Optical and crystallographic properties and impurity incorporation of $Ga_xIn_{1-x}As$ ($0.44 < x < 0.49$) grown by liquid phase epitaxy, vapor phase epitaxy, and metal organic chemical vapor deposition. *Journal of Applied Physics*, 54(8):4543–4552, 1983.
5. Q. Zhang, F. Q. Liu, W. Zhang, Q. Lu, L. Wang, L. Li, and Z. Wang. Thermal induced facet destructive feature of quantum cascade lasers. *Applied Physics Letters*, 96(14):141117, 2010.
6. D. Botez, J. C. Shin, J. D. Kirch, C. C. Chang, L. J. Mawst, and T. Earles. Multidimensional conduction-band engineering for maximizing the continuous-wave (CW) wallplug efficiencies of mid-infrared quantum cascade lasers. *IEEE Journal of Selected Topics in Quantum Electronics*, 19(4):1200312–1200312, 2013.
7. P. Q. Liu, A. J. Hoffman, M. D. Escarra, K. J. Franz, J. B. Khurgin, Y. Dikmelik, X. Wang, J.-Y. Fan, and C. F. Gmachl. Highly power-efficient quantum cascade lasers. *Nature Photonics*, 4:95–98, 2010.
8. A. J. Hoffman, Y. Yao, and C. F. Gmachl. Mid-infrared quantum cascade laser. *Nature Photonics*, 6(1749–4885):432–439, 2012.
9. Y. Bai, S. R. Darvish, S. Slivken, W. Zhang, A. Evans, J. Nguyen, and M. Razeghi. Room temperature continuous wave operation of quantum cascade lasers with watt-level optical power. *Applied Physics Letters*, 92(10):101105, 2008.

10. A. Lyakh, R. Maulini, A. Tsekoun, R. Go, C. Pflügl, L. Diehl, Q. J. Wang, F. Capasso, and C. Kumar N. Patel. 3 W continuous-wave room temperature single-facet emission from quantum cascade lasers based on nonresonant extraction design approach. *Applied Physics Letters*, 95(14):141113, 2009.
11. N. Bandyopadhyay, Y. Bai, B. Gokden, A. Myzaferi, S. Tsao, S. Slivken, and M. Razeghi. Watt level performance of quantum cascade lasers in room temperature continuous wave operation at $\lambda \sim 3.76$ μm. *Applied Physics Letters*, 97(13):131117, 2010.
12. N. Bandyopadhyay, S. Slivken, Y. Bai, and M. Razeghi. High power, continuous wave, room temperature operation of $\lambda \sim 3.4$ μm and $\lambda \sim 3.55$ μm InP-based quantum cascade lasers. *Applied Physics Letters*, 100(21):212104, 2012.
13. Y. Bai, N. Bandyopadhyay, S. Tsao, S. Slivken, and M. Razeghi. Room temperature quantum cascade lasers with 27% wall plug efficiency. *Applied Physics Letters*, 98(18):181102, 2011.
14. Y. Yu, W. Xiaojun, F. Jen-Yu, and C. F. Gmachl. High performance continuum-to-continuum quantum cascade lasers with a broad gain bandwidth of over 400 cm^{-1}. *Applied Physics Letters*, 97(8):081115, 2010.
15. M. Razeghi, Q. Y. Lu, N. Bandyopadhyay, W. Zhou, D. Heydari, Y. Bai, and S. Slivken. Quantum cascade lasers: From tool to product. *Optics Express*, 23(7):8462–8475, 2015.
16. M. Wienold, M. P. Semtsiv, I. Bayrakli, W. T. Masselink, M. Ziegler, K. Kennedy, and R. Hogg. Optical and thermal characteristics of narrow-ridge quantum-cascade lasers. *Journal of Applied Physics*, 103(8):083113, 2008.
17. H. K. Lee and J. S. Yu. Thermal analysis of short wavelength InGaAs/InAlAs quantum cascade lasers. *Solid-State Electronics*, 54(8):769–776, 2010.
18. J. D. Kirch, J. C. Shin, C. C. Chang, L. J. Mawst, D. Botez, and T. Earles. Tapered active-region quantum cascade lasers ($\lambda = 4.8$ μm) for virtual suppression of carrier-leakage currents. *Electronics Letters*, 48(4):234–235, 2012.
19. D. Botez, C. C. Chang, and L. J. Mawst. Temperature sensitivity of the electro-optical characteristics for mid-infrared ($\lambda = 3$–16 μm)-emitting quantum cascade lasers. *Journal of Physical D: Applied Physics*, 49:043001, 2016.
20. D. Indjin, P. Harrison, R. W. Kelsall, and Z. Ikonić. Self-consistent scattering theory of transport and output characteristics of quantum cascade lasers. *Journal of Applied Physics*, 91(11):9019–9026, 2002.
21. D. Indjin, P. Harrison, R. W. Kelsall, and Z. Ikonic. Influence of leakage current on temperature performance of GaAs/AlGaAs quantum cascade lasers. *Applied Physics Letters*, 81(3):400–402, 2002.
22. A. Mircetic, D. Indjin, Z. Ikonic, P. Harrison, V. Milanovic, and R. W. Kelsall. Towards automated design of quantum cascade lasers. *Journal of Applied Physics*, 97(8):084506, 2005.
23. R. Claudia Iotti and F. Rossi. Nature of charge transport in quantum-cascade lasers. *Physical Review Letters*, 87:146603, 2001.
24. H. Callebaut, S. Kumar, B. S. Williams, Q. Hu, and J. L. Reno. Importance of electron-impurity scattering for electron transport in terahertz quantum-cascade lasers. *Applied Physics Letters*, 84(5):645–647, 2004.
25. X. Gao, D. Botez, and I. Knezevic. X-valley leakage in GaAs-based midinfrared quantum cascade lasers: A Monte Carlo study. *Journal of Applied Physics*, 101(6):063101, 2007.
26. Y. B. Shi and I. Knezevic. Nonequilibrium phonon effects in midinfrared quantum cascade lasers. *Journal of Applied Physics*, 116(12):123105, 2014.
27. H. Willenberg, G. H. Döhler, and J. Faist. Intersubband gain in a Bloch oscillator and quantum cascade laser. *Physical Review B*, 67:085315, 2003.
28. S. Kumar and Q. Hu. Coherence of resonant-tunneling transport in terahertz quantum-cascade lasers. *Physical Review B*, 80:245316, 2009.
29. C. Weber, A. Wacker, and A. Knorr. Density-matrix theory of the optical dynamics and transport in quantum cascade structures: The role of coherence. *Physical Review B*, 79:165322, 2009.

30. E. Dupont, S. Fathololoumi, and H. C. Liu. Simplified density-matrix model applied to three-well terahertz quantum cascade lasers. *Physical Review B*, 81:205311, 2010.
31. R. Terazzi and J. Faist. A density matrix model of transport and radiation in quantum cascade lasers. *New Journal of Physics*, 12(3):033045, 2010.
32. O. Jonasson, F. Karimi, and I. Knezevic. Partially coherent electron transport in terahertz quantum cascade lasers based on a Markovian master equation for the density matrix. *Journal of Computational Electronics*, 15:1192, 2016.
33. O. Jonasson, S. Mei, F. Karimi, J. Kirch, D. Botez, L. Mawst, and I. Knezevic. Quantum transport simulation of high-power 4.6-μm quantum cascade lasers. *Photonics*, 3(2):38, 2016.
34. S.-C. Lee and A. Wacker. Nonequilibrium Green's function theory for transport and gain properties of quantum cascade structures. *Physical Review B*, 66:245314, 2002.
35. M. Bugajski, P. Gutowski, P. Karbownik, A. Kolek, G. Hałdaś, K. Pierściński, D. Pierścińska, J. Kubacka-Traczyk, I. Sankowska, A. Trajnerowicz, K. Kosiel, A. Szerling, J. Grzonka, K. Kurzydłowski, T. Slight, and W. Meredith. Mid-IR quantum cascade lasers: Device technology and non-equilibrium Green's function modeling of electro-optical characteristics. *Physica Status Solidi (b)*, 251(6):1144–1157, 2014.
36. A. Kolek, G. Haldas, M. Bugajski, K. Pierscinski, and P. Gutowski. Impact of injector doping on threshold current of mid-infrared quantum cascade laser–non-equilibrium greens function analysis. *IEEE Journal of Selected Topics in Quantum Electronics*, 21(1):124–133, 2015.
37. O. Jonasson and I. Knezevic. Dissipative transport in superlattices within the Wigner function formalism. *Journal of Computational Electronics*, 14:879–887, 2015.
38. A. Matyas, P. Lugli, and C. Jirauschek. Photon-induced carrier transport in high efficiency midinfrared quantum cascade lasers. *Journal of Applied Physics*, 110(1):013108, 2011.
39. M. Lindskog, J. M. Wolf, V. Trinite, V. Liverini, J. Faist, G. Maisons, M. Carras, R. Aidam, R. Ostendorf, and A. Wacker. Comparative analysis of quantum cascade laser modeling based on density matrices and non-equilibrium Green's functions. *Applied Physics Letters*, 105(10):103106, 2014.
40. R. C. Iotti and F. Rossi. Microscopic theory of semiconductor-based optoelectronic devices. *Reports on Progress in Physics*, 68(11):2533, 2005.
41. C. Jirauschek and P. Lugli. Monte-Carlo-based spectral gain analysis for terahertz quantum cascade lasers. *Journal of Applied Physics*, 105(12):123102, 2009.
42. C. Jirauschek and T. Kubis. Modeling techniques for quantum cascade lasers. *Applied Physics Reviews*, 1(1):011307, 2014.
43. C. Jirauschek. Monte Carlo study of carrier-light coupling in terahertz quantum cascade lasers. *Applied Physics Letters*, 96(1), 2010.
44. A. Wacker, M. Lindskog, and D. O. Winge. Nonequilibrium Green's function model for simulation of quantum cascade laser devices under operating conditions. *IEEE Journal of Selected Topics in Quantum Electronics*, 19(5):1–11, 2013.
45. C. A. Evans, D. Indjin, Z. Ikonic, P. Harrison, M. S. Vitiello, V. Spagnolo, and G. Scamarcio. Thermal modeling of terahertz quantum-cascade lasers: Comparison of optical waveguides. *IEEE Journal of Quantum Electronics*, 44(7):680–685, 2008.
46. H. K. Lee and J. S. Yu. Thermal effects in quantum cascade lasers at $\lambda \sim 4.6$ μm under pulsed and continuous-wave modes. *Applied Physics B*, 106(3):619–627, 2012.
47. Y. B. Shi, Z. Aksamija, and I. Knezevic. Self-consistent thermal simulation of GaAs/$Al_{0.45}Ga_{0.55}As$ quantum cascade lasers. *Journal of Computational Electronics*, 11(1):144–151, 2012.
48. S. Mei and I. Knezevic. Thermal conductivity of III-V semiconductor superlattices. *Journal of Applied Physics*, 118(17):175101, 2015.
49. Y. B. Shi, S. Mei, O. Jonasson, and I. Knezevic. Modeling quantum cascade lasers: Coupled electron and phonon transport far from equilibrium and across disparate spatial scales. *Fortschritte der Physik – Progress of Physics*, 1600084, 2016. doi:10.1002/prop.201600084

50. S.-C. Lee, F. Banit, M. Woerner, and A. Wacker. Quantum mechanical wavepacket transport in quantum cascade laser structures. *Physical Review B*, 73:245320, 2006.
51. K. Faist, D. Hofstetter, M. Beck, T. Aellen, M. Rochat, and S. Blaser. Bound-to-continuum and two-phonon resonance, quantum-cascade lasers for high duty cycle, high-temperature operation. *IEEE Journal of Quantum Electronics*, 38(6):533–546, 2002.
52. M. Yamanishi, T. Edamura, K. Fujita, N. Akikusa, and H. Kan. Theory of the intrinsic linewidth of quantum-cascade lasers: Hidden reason for the narrow linewidth and line-broadening by thermal photons. *IEEE Journal of Quantum Electronics*, 44(1):12–29, 2008.
53. X. Gao. Monte Carlo Simulation of Electron Dynamics in Quantum Cascade Lasers. PhD thesis, University of Wisconsin–Madison, 2008.
54. Y. Shi. Electrothermal Simulation of Quantum Cascade Lasers. PhD thesis, University of Wisconsin–Madison, 2015.
55. H. Callebaut and Q. Hu. Importance of coherence for electron transport in terahertz quantum cascade lasers. *Journal of Applied Physics*, 98(10):104505, 2005.
56. A. Wacker. Semiconductor superlattices: A model system for nonlinear transport. *Physics Reports*, 357(1):1–111, 2002.
57. X. Gao, D. Botez, and I. Knezevic. X-valley leakage in GaAsAlGaAs quantum cascade lasers. *Applied Physics Letters*, 89(19):191119, 2006.
58. S. L. Chuang. *Physics of Optoelectronic Devices*. New York, NY: Wiley, 1995.
59. X. Gao, D. Botez, and I. Knezevic. Confined phonon scattering in multivalley Monte Carlo simulation of quantum cascade lasers. *Journal of Computational Electronics*, 7(3):209–212, 2008.
60. P. Lugli, P. Bordone, L. Reggiani, M. Rieger, P. Kocevar, and S. M. Goodnick. Monte Carlo studies of nonequilibrium phonon effects in polar semiconductors and quantum wells. I. Laser photoexcitation. *Physical Review B*, 39:7852–7865, 1989.
61. B. K. Ridley. The electron-phonon interaction in quasi-two-dimensional semiconductor quantum-well structures. *Journal of Physics C: Solid State Physics*, 15(28):5899, 1982.
62. J. T. Lü and J. C. Cao. Monte Carlo simulation of hot phonon effects in resonant-phonon-assisted terahertz quantum-cascade lasers. *Applied Physics Letters*, 88(6):061119, 2006.
63. S. Usher and G. P. Srivastava. Theoretical study of the anharmonic decay of nonequilibrium LO phonons in semiconductor structures. *Physical Review B*, 50:14179–14186, 1994.
64. E. Pop, S. Sinha, and K. E. Goodson. Heat generation and transport in nanometer-scale transistors. *Proceedings of the IEEE*, 94(8):1587–1601, 2006.
65. D. G. Cahill, P. V. Braun, G. Chen, D. R. Clarke, S. Fan, K. E. Goodson, P. Keblinski, W. P. King, G. D. Mahan, A. Majumdar, H. J. Maris, S. R. Phillpot, E. Pop, and L. Shi. Nanoscale thermal transport. II. 2003–2012. *Applied Physics Reviews*, 1(1):011305, 2014.
66. T. Yao. Thermal properties of AlAs/GaAs superlattices. *Applied Physics Letters*, 51(22):1798–1800, 1987.
67. G. Chen, C. L. Tien, X. Wu, and J. S. Smith. Thermal diffusivity measurement of GaAs/AlGaAs thin-film structures. *Journal of Heat Transfer*, 116:325–331, 1994.
68. X. Y. Yu, G. Chen, A. Verma, and J. S. Smith. Temperature dependence of thermophysical properties of GaAs/AlAs periodic structure. *Applied Physics Letters*, 67(24):3554–3556, 1995.
69. W. S. Capinski and H. J. Maris. Thermal conductivity of GaAs/AlAs superlattices. *Physica B*, 219–220:699–701, 1996.
70. W. S. Capinski, H. J. Maris, T. Ruf, M. Cardona, K. Ploog, and D. S. Katzer. Thermal conductivity measurements of GaAs/AlAs superlattices using a picosecond optical pump-and-probe technique. *Physical Review B*, 59:8105–8113, 1999.
71. G. Chen. Size and interface effects on thermal conductivity of superlattices and periodic thin-film structures. *Journal of Heat Transfer*, 119(2):220, 1997.
72. G. Chen. Thermal conductivity and ballistic-phonon transport in the cross-plane direction of superlattices. *Physical Review B*, 57:14958–14973, 1998.

73. S. T. Huxtable, A. R. Abramson, C. L. Tien, A. Majumdar, C. LaBounty, X. Fan, G. Zeng, J. E. Bowers, A. Shakouri, and E. T. Croke. Thermal conductivity of Si/SiGe and SiGe/SiGe superlattices. *Applied Physics Letters*, 80(10):1737–1739, 2002.
74. Y. Wang, H. Huang, and X. Ruan. Decomposition of coherent and incoherent phonon conduction in superlattices and random multilayers. *Physical Review B*, 90:165406, 2014.
75. Y. Wang, C. Gu, and X. Ruan. Optimization of the random multilayer structure to break the random-alloy limit of thermal conductivity. *Applied Physics Letter*, 106:073104, 2015.
76. Z. Aksamija and I. Knezevic. Thermal conductivity of Si1−*x*G*ex*/Si1−*y*G*ey*superlattices: Competition between interfacial and internal scattering. *Physical Review B*, 88:155318, 2013.
77. E. T. Swartz and R. O. Pohl. Thermal boundary resistance. *Reviews of Modern Physics*, 61:605–668, 1989.
78. A. Lops, V. Spagnolo, and G. Scamarcio. Thermal modeling of GaInAs/AlInAs quantum cascade lasers. *Journal of Applied Physics*, 100(4):043109, 2006.
79. H. Page, C. Becker, A. Robertson, G. Glastre, V. Ortiz, and C. Sirtori. 300 k operation of a GaAs-based quantum-cascade laser at $\lambda \approx 9$ μm. *Applied Physics Letters*, 78(22):3529, 2001.
80. V. Spagnolo, A. Lops, G. Scamarcio, M. S. Vitiello, and C. Di Franco. Improved thermal management of mid-IR quantum cascade lasers. *Journal of Applied Physics*, 103(4):043103, 2008.

34

Vertical-Cavity Surface-Emitting Lasers

Tomasz Czyszanowski

Leszek Frasunkiewicz

and

Maciej Dems

34.1 Introduction

The history of vertical-cavity surface-emitting lasers (VCSELs) can be traced back almost as far as that of semiconductor lasers themselves. In 1965, Melngailis demonstrated coherent emission parallel to the direction of the current in an n^+pp^+ InSb diode laser.[1] The polished surface on the emission side of InSb and the gold current contact on its opposite side provided sufficient optical feedback to enhance laser emission. In this very early work, a design was proposed for the integration in an array that would provide coherent emission over a large area, with a small beam angle. Unfortunately, this idea was forgotten and not rediscovered until 1979 by Soda et al.,[2] who demonstrated a surface-emitting InP-based diode laser with two metallic mirrors. Further development was initiated by Ogura et al.,[3] who introduced highly reflective distributed Bragg reflectors (DBRs) in 1983.

Unlike typical edge-emitting lasers, in VCSELs the optical cavity is formed between the mirrors above and below the active region (Figure 34.1). The laser light resonates in the vertical direction. Inside the laser structure, the light passes the active region in the vertical direction—i.e., gain is provided over a short distance only and the amplification per photon round-trip is small. Therefore, the mirrors must be highly reflective (over 99% for the emitting mirror and almost 100% for the opposite one) so that the photons make many round-trips before they are emitted. High reflectivity is provided by DBRs composed of stacks of two alternating layers with high refractive index contrast. With quarter-wavelength layer thicknesses, the reflected waves from all DBR interfaces add up constructively, allowing for a total DBR reflectance of over 99%. Such layers are composed of 50 or more semiconductor layers. The current is injected vertically, similar to edge-emitting devices. However, in VCSELs, optical resonance occurs in the vertical direction.

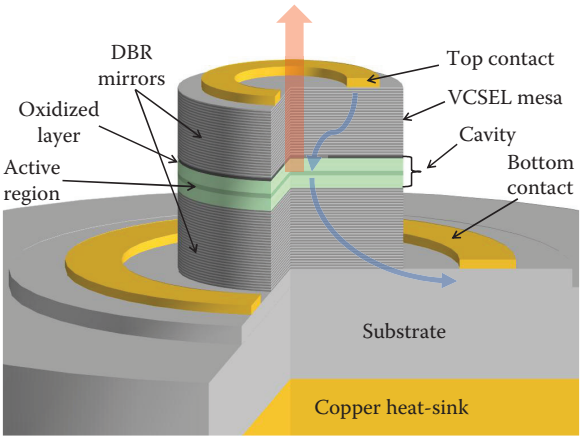

FIGURE 34.1 Intersection of a VCSEL structure showing key construction elements. The arrows inside VCSEL structure illustrate the path of the current flow and vertical arrow the emitted beam.

Optical resonance and current injection taking place in the same direction is the main weakness of these devices. Theoretically, preferential injection of the carriers to the active region may be achieved by placing two electrodes on opposite sides of the laser. However, in such configuration, the stimulated light resonates between the mirrors and no emission occurs, since the metal contacts are highly absorptive. To enable emission, a ring contact is typically used on the top DBR, through which the light is emitted. However, current flows through the shortest path, so maximal stimulated emission overlaps with the top contact, leading to high absorption in the metal and deterioration of the emission beam. To overcome this problem, the carriers must be forced to flow through the central part of the active region. This can be achieved either by selective proton implantation or by selective oxidation of the lateral regions in the VCSEL. Both of these approaches turn the semiconductor layers into electrical insulators. With selective processes, the central part of the VCSEL is conductive and radiative recombination takes place within the central part of the active region. Both methods also allow for cylindrical aperture formation, supporting cylindrical beam emission that cannot be achieved with typical edge-emitting lasers. Such symmetry of the aperture, together with the use of ring contacts, favors the choice of cylindrical geometry for numerical approaches, which are used to simulate the phenomena that take place in VCSELs during laser operation.

34.2 Physical Phenomena

VCSELs are very complex systems due to their multilayered structure (comprising in excess of a hundred layers). Often with nonplanar or buried-type architectures, they feature many heterojunctions, graded layers, quantum wells (QWs), quantum dots (QDs) or quantum wires, superlattices, oxide and oxidized layers, mesa structures, photonic microstructures, and so on. To analyze the operation of VCSELs precisely, therefore, one needs to consider four main classes of phenomena in the device: electrical, thermal, optical, and recombination, which together create a nonlinear network of mutual interrelations. Each class of phenomena is responsible for the following:

1. Optical: Cavity modes, their modal gain and loss; emission wavelengths and optical field distributions within the laser cavity; formation of the output beam
2. Electrical: Current spreading between the top and bottom contacts through the centrally located active region; injection of carriers of both polarities into the active region and their subsequent radiative or nonradiative recombination after radial out-diffusion in the active region; overbarrier carrier leakage

3. Thermal: Generation of heat flux (nonradiative recombination, reabsorption of spontaneous radiation, as well as volume and barrier Joule heating); heat spreading within heat sinks
4. Recombination: Material gain as a result of stimulated recombination (to determine the active region optical gain spectrum); nonradiative monomolecular and Auger recombinations

The main effects induced by nonlinear interaction between the models are thermal focusing and emitted power rollover. The first is related to the dependence of the refractive index on temperature. The heat induced predominantly in the central part of the active region dissipates in all directions, but mostly toward the bottom heat sink. This, in turn, induces different temperatures in the laser cavity. Increasing the temperature of semiconductor materials triggers an increase in their refractive index. The most significant increase of the refractive index is in the central part of the laser, which is heated most intensely. This causes the waveguiding effect, whereby the modes squeeze radially into the central part of the laser. On the one hand, stronger focusing of the modes in the active region increases their interaction with carriers that are injected into the central part of the active region, reducing the current necessary for laser operation (threshold current). On the other hand, if there are more optical modes confined and excited within the central part, this contributes to a broadening of the emission spectrum, since each mode relates to a slightly different emission wavelength. The quality of the output beam also deteriorates, since interference between different modes occurs at various emission angles.

The second effect—emission power rollover—limits the emitted power of the laser, mostly due to thermal effects. The active regions of VCSELs are designed in such a way that the wavelength corresponding to maximal material gain overlaps with the wavelength of the optical resonance induced by the VCSEL cavity. A change in the temperature of the device, due to heat dissipation or ambient temperature, may modify the bandgap and consequently change the wavelength corresponding to the maximum of the material gain. The optical modes do not follow the maximum of the material gain, as is the case in edge-emitting lasers, but follow the optical resonance induced by the cavity. The resonance wavelength also changes with temperature, since the refractive indices of the cavity change. The changes in maximal material gain and of cavity resonance, however, are of different proportions. Typically, the change in the gain spectrum is more sensitive to temperature changes than in the case of optical resonance. Hence, material gain can be insufficient at resonant wavelength to sustain laser operation for higher operational temperatures and the laser ceases to emit stimulated radiation. This effect is manifested as a reduction in emitted power as the injected current increases.

To simulate such complex behaviors, a comprehensive model including electrical, thermal, recombination, and optical submodels is needed. Three of these submodels can be solved in a very similar manner to the optoelectronic devices considered previously in this book. However, special attention must be given to the optical model, since electromagnetic field resonance takes place in the vertical direction, which is orthogonal to the resonance that occurs in typical edge-emitting lasers. In the remainder of this chapter, we have therefore focused on optical simulation.

34.3 Light Confinement in VCSELs

VCSELs inherently emit in single longitudinal mode due to their very short cavities. Controlling the modes in the transversal direction is more complicated. In typical VCSELs made from arsenide-based materials, well-established wet oxidation technology can be used,[4] which produces the waveguide effect through a selectively oxidized Al-rich layer that also plays the role of an electric insulator for improved current funneling to the active region. Another less material dependent technique is proton implantation,[5] which defines highly resistive regions. The waveguide effect is produced by thermal focusing induced by thermal increase of the refractive index in regions where nonradiative recombination and Joule heating occur. Other methods of transverse mode confinement require more sophisticated and expensive technologies. Possible solutions include micro-optical structures, such as surface relief,[6] antiresonant patterning,[7] tunnel junction patterning,[8] photonic crystals[9] and highcontrast gratings.[10] Each of these methods requires

a separate analytical approach, since different phenomena are responsible for mode confinement: the waveguide effect, total interior reflections, Bragg reflections or Fano resonance. Since not all those phenomena can be observed in all VCSEL constructions with equal intensity, numerical models should be properly chosen to provide a balance between generalization and efficiency.

34.4 Fundamental Equations

Interaction between a VCSEL and the optical field is governed by Maxwell's equations. The boundary conditions correspond to an optical field that decays to 0 in infinity. These boundary conditions mean that Maxwell's equations turn into an eigenvalue problem. Its solutions are discrete eigenvalues and the corresponding eigenvectors. Eigenvalues of Maxwell's equations are vacuum wavevectors (k_0), corresponding angular frequencies (ω), or wavelengths in a vacuum (λ). Eigenvectors are distributions of electric (**E**) and magnetic (**H**) fields of electromagnetic waves corresponding to particular eigenvalues.

The three-dimensional electric **E** and magnetic **H** field vectors satisfy the following partial differential equations,[11] which correspond to Faraday's and Amper's laws, respectively:

$$\nabla \times \mathbf{E} = -\mu\mu_0 \frac{\partial \mathbf{H}}{\partial t}, \tag{34.1}$$

$$\nabla \times \mathbf{H} = \varepsilon\varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}, \tag{34.2}$$

where μ and μ_0 are the magnetic permittivity of the material and the vacuum, respectively, and ε and ε_0 are analogous dielectric constants.

Let us consider a monochromatic wave only, of which the angular frequency is ω. The time dependency of **E** and **H** is given by the following relations:

$$\mathbf{E} = \mathbf{E} \exp(i\omega t), \tag{34.3}$$

$$\mathbf{H} = \mathbf{H} \exp(i\omega t). \tag{34.4}$$

Embedment of Equations 34.3 and 34.4 to Equations 34.1 and 34.2 results in a time-independent set of equations:

$$\nabla \times \mathbf{E} = -i\omega\mu\mu_0 \mathbf{H}, \tag{34.5}$$

$$\nabla \times \mathbf{H} = i\omega\varepsilon\varepsilon_0 \mathbf{E}. \tag{34.6}$$

We assume that μ is uniform and equal to 1, which is always true for the dielectrics and nonferromagnetic metals typically used in VCSELs. Combining Equations 34.5 and 34.6, we can reduce the set of six equations for all vector components to a set of three equations for the electric field only:

$$\nabla \times \nabla \times \mathbf{E} = k_0^2 n^2 \mathbf{E}, \tag{34.7}$$

where k_0 is the wavenumber of the mode in a vacuum and

$$\sqrt{\varepsilon} = n = n_{\text{re}} + i n_{\text{im}} \tag{34.8}$$

is the complex refractive index. Its imaginary part accounts for the loss and the gain of the medium in which the wave propagates. The relation between the imaginary part of n and losses α is given by

$$n_{\text{im}} = -\frac{\alpha}{2 k_0}. \tag{34.9}$$

Equation 34.9 can be rewritten in an equivalent form:

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E} = \omega^2\mu\varepsilon\mathbf{E},$$
$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E} = k_0^2 n^2\mathbf{E}. \tag{34.10}$$

VCSELs are typically processed to be circular in the epitaxial plane to support symmetric beam emission. This geometry allows the complexity of Equation 34.7 to be reduced by expressing it in a cylindrical coordinate system where the angular dependence of **E** can be given in simple analytical terms. However, in recent years, numerous confinement approaches based on photonic microstructures have been proposed. Many of these structures cannot be analyzed using the cylindrical coordinate system and therefore the three-dimensional Cartesian system must be used. We will therefore now show how Equations 34.5, 34.6, and 34.10 can be solved using three approaches of incremental complexity. The first approach is based on a one-dimensional analysis of the optical field along the optical axis. The second assumes that the VCSEL has cylindrical geometry and considers both the scalar optical field in two dimensions along the optical axis and the radius of the VCSEL. The third considers the vectorial optical field in the three-dimensional Cartesian coordinate system.

34.5 VCSEL Structure

Figure 34.2 shows the VCSEL structure with construction details listed in Table 34.1. This structure was used in References[12,13] as a benchmark for comparative analysis of optical numerical models, however it must be stressed here, the structure is significantly simplified in comparison to real-life VCSELs. The structure consists of one wavelength-long GaAs cavity with a 5-nm active GaAs layer at the antinode of the standing wave and a selectively oxidized AlAs layer ~15 nm in thickness. The cavity is sandwiched between 24 and 29.5 pairs of GaAs/AlAs DBRs. In comparison to a real-life VCSEL, there is no vertical carrier confinement that would capture and support optical recombination of the carriers. There is also no doping scheme that would allow efficient injection of the current to the active region.

In what follows, we will be using two coordinate systems: Cartesian and cylindrical (Figure 34.1). These share a z-axis along the optical axis of the device. The Cartesian plane $x - y$ and cylindrical plane $r - \varphi$ are parallel to the epitaxial layers of the VCSEL, and the x-axis relates to $\varphi = 0$.

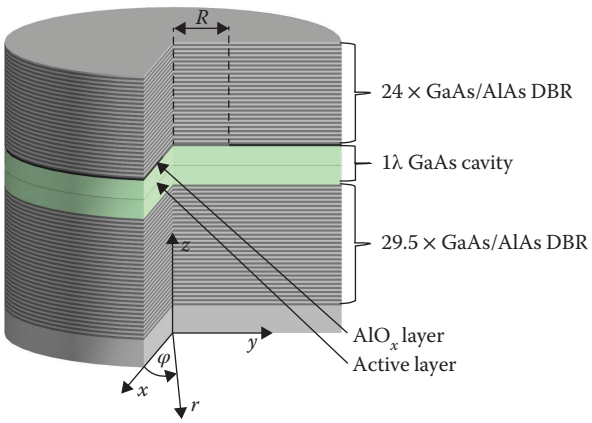

FIGURE 34.2 Intersection of the VCSEL structure with assigned radius (R) of the oxide aperture and definition of Cartesian (x, y, z) and cylindrical (r, φ, z) coordinate systems. DBR = distributed Bragg reflectors.

TABLE 34.1 Refractive Indices and Thicknesses of VCSEL Layers

Refractive Index	Thickness (nm)	Repetition
1.0	∞	–
3.53	69.49	24.5
3.08	79.63	
3.08	63.71	–
2.95 for $r < R$ 1.6 for $r > R$	15.93	–
3.53	136.49	–
3.53 (active)	5.00	–
3.53	136.49	–
3.08	79.63	29.5
3.53	69.49	
3.53	∞	–

Note: The two outer layers are assumed to be infinite.

Vectors **E** and **H** are given in the Cartesian system as

$$\mathbf{E} = \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix}, \tag{34.11}$$

and in the cylindrical system as

$$\mathbf{E} = \begin{bmatrix} E_r \\ E_\phi \\ E_z \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} H_r \\ H_\varphi \\ H_z \end{bmatrix}. \tag{34.12}$$

34.6 One-Dimensional Scalar Approach

The simplest approach is to consider the distribution of light within the VCSEL along the z-axis only, assuming an electromagnetic field in the form of a plane wave within the device and that the layers are uniform in the lateral direction. This extremely simplified approach allows only the longitudinal modes to be determined. However, it is of tremendous importance in VCSEL wafers designing, as it enables one to determine the size of the cavity and the number of necessary DBRs, as well as the positions of active or highly doped layers.

The assumption of uniform layers corresponds to the Gauss's law relating to electrical fields:

$$\nabla \cdot \varepsilon \mathbf{E} = 0. \tag{34.13}$$

Since ε is constant within any particular layer of the VCSEL, Gauss's law can be reduced to the following form:

$$\nabla \cdot \mathbf{E} = 0, \tag{34.14}$$

which simplifies Equation 34.10:

$$\nabla^2 \mathbf{E} = k_0^2 n^2 \mathbf{E}. \tag{34.15}$$

The operator ∇^2 can be expressed in either coordinate system. However, for the sake of clarity in what follows, we shall express it in the cylindrical system:

$$\nabla^2 \equiv \begin{bmatrix} \partial_r^2 + \frac{1}{r}\partial_r + \frac{1}{r^2}\partial_\varphi^2 + \partial_z^2 & 0 & 0 \\ 0 & \partial_r^2 + \frac{1}{r}\partial_r + \frac{1}{r^2}\partial_\varphi^2 + \partial_z^2 & 0 \\ 0 & 0 & \partial_r^2 + \frac{1}{r}\partial_r + \frac{1}{r^2}\partial_\varphi^2 + \partial_z^2 \end{bmatrix}. \tag{34.16}$$

Substituting Equation 34.16 to Equation 34.15, one obtains three independent equations for each field component of **E**:

$$\left(\partial_r^2 + \frac{1}{r}\partial_r + \frac{1}{r^2}\partial_\varphi^2 + \partial_z^2\right) E_u = k_0^2 n^2 E_u, \tag{34.17}$$

where $u = r, \varphi, z$. An additional assumption with regard to the plane wave imposes $E_r = \text{const}$ and $E_\varphi = \text{const}$, which simplifies Equation 34.17 to a one-dimensional expression:

$$\partial_z^2 E_u = k_0^2 n^2 E_u. \tag{34.18}$$

Each of the equations can be solved separately; although by imposing the assumption of layer uniformity, we lose information on the relations between the components of the electric field. Hence, it is useful to assume that when linearly polarized light propagates along the z-axis then

$$E_z = 0$$

and

$E_\varphi = E_r \cot\varphi$ for x polarization or
$E_\varphi = E_r \tan\varphi$ for y polarization.

To solve Equation 34.18, we used the transfer matrix method (TMM),[14] which assumes that within the jth layer the solution of Equation 34.18 can be expressed by the function (Figure 34.3):

$$E_r^j = A_j \exp\left(ik_j z\right) + B_j \exp\left(-ik_j z\right), \tag{34.19}$$

where $k_j = k_0 n_j$ and E_r is continuous and smooth at the interfaces between the consecutive layers of the VCSEL (Figure 34.3). This leads to the following relations:

$$A_j \exp\left(ik_j d_j\right) + B_j \exp\left(-ik_j d_j\right) = A_{j+1} \exp\left(ik_{j+1} 0\right) + B_{j+1} \exp\left(-ik_{j+1} 0\right), \tag{34.20}$$

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(A_j \exp\left(ik_j z\right) + B_j \exp\left(-ik_j z\right)\right)\Big|_{z=d_j} = \frac{\mathrm{d}}{\mathrm{d}z}\left(A_{j+1} \exp\left(ik_{j+1} z\right) + B_{j+1} \exp\left(-ik_{j+1} z\right)\right)\Big|_{z=0}, \tag{34.21}$$

at neighboring layers expressed in local coordinates within layers; k_0 is unknown and corresponds to the vacuum wavevector of the mode.

As a consequence, we can write the relation that transforms E_r between the layers as follows:

$$\begin{aligned} A_j \tfrac{k_{j+1}+k_j}{2k_{j+1}} \exp\left(ik_j d_j\right) + \tfrac{k_{j+1}-k_j}{2k_{j+1}} B_j \exp\left(-ik_j d_j\right) &= A_{j+1} \\ A_j \tfrac{k_{j+1}-k_j}{2k_{j+1}} \exp\left(ik_j d_j\right) + \tfrac{k_{j+1}+k_j}{2k_{j+1}} B_j \exp\left(-ik_j d_j\right) &= B_{j+1} \end{aligned}, \tag{34.22}$$

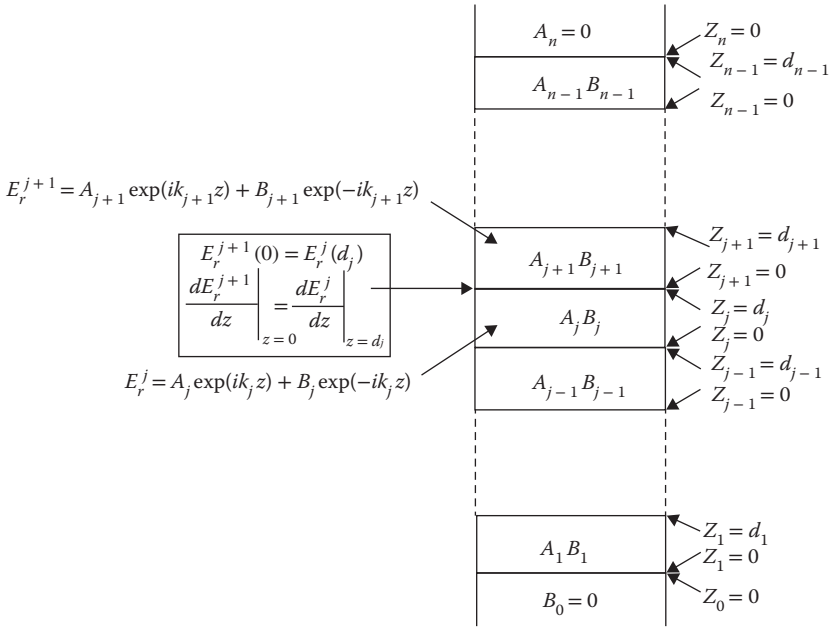

FIGURE 34.3 Schematic view of structure illustrating parameters and variables used in TMM algorithm.

and this can be given in a matrix form:

$$\begin{bmatrix} A_{j+1} \\ B_{j+1} \end{bmatrix} = \begin{bmatrix} \frac{k_{j+1}+k_j}{2k_{j+1}} \exp\left(ik_jd_j\right) & \frac{k_{j+1}-k_j}{2k_{j+1}} \exp\left(-ik_jd_j\right) \\ \frac{k_{j+1}-k_j}{2k_{j+1}} \exp\left(ik_jd_j\right) & \frac{k_{j+1}+k_j}{2k_{j+1}} \exp\left(-ik_jd_j\right) \end{bmatrix} \begin{bmatrix} A_j \\ B_j \end{bmatrix} = \mathbf{M_j} \begin{bmatrix} A_j \\ B_j \end{bmatrix}. \tag{34.23}$$

Matrix $\mathbf{M}_j$ connects the amplitudes of the electric field from layers j and $j+1$. All layers of the VCSEL can be combined using the recursive relation:

$$\begin{bmatrix} A_n \\ B_n \end{bmatrix} = \prod_{j=0}^{n} \mathbf{M}_j \begin{bmatrix} A_0 \\ B_0 \end{bmatrix} = \mathbf{M} \begin{bmatrix} A_0 \\ B_0 \end{bmatrix}. \tag{34.24}$$

Boundary conditions impose that the electric field decays to infinity, so the exponential elements responsible for infinite values of E in $+\infty$ and $-\infty$ must be eliminated:

$$\begin{bmatrix} 0 \\ B_n \end{bmatrix} = \mathbf{M} \begin{bmatrix} A_0 \\ 0 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} A_0 \\ 0 \end{bmatrix}. \tag{34.25}$$

This relation implies that

$$m_{11}\left(k_0\right) = 0, \tag{34.26}$$

which is the condition in the numerical algorithm responsible for finding the longitudinal modes of the VCSEL. k_0, which fulfills the condition (Equation 34.26), corresponds to the longitudinal mode of the VCSEL and is interpreted as the wavenumber of the longitudinal mode in the vacuum. Simple relations give the emitted wavelength as well as the frequency of the mode. The imaginary part of k_0 corresponds to

the total losses or gain of the mode. If $\text{im}(k_0) \geq 0$, the mode reaches the threshold and can be observed in the emitted spectrum.

The vertical mode distribution for the structure detailed in Table 34.1, calculated using TMM, is given in Figure 34.4. Mode intensity is defined as follows:

$$I = E_u E_u^*, \tag{34.27}$$

where E^* is a complex conjugate of the electric field of an electromagnetic wave. Figure 34.4 shows typical oscillatory decay of optical field intensity within the top and bottom DBRs. In the simplified example considered here, in which the cavity is uniform, optical intensity is maximal and the amplitude is constant in the cavity. Typically, VCSELs are designed to ensure the coincidence of optical field intensity and the active region to maximize stimulated emission.

Using this simple approach, one can design the dominant wavelength of the VCSEL wafer by tuning the length of the cavity. Figure 34.5 illustrates the dependence of the resonant wavelength and modal gain as functions of detuning the cavity length from its optimal length. The cavity length and DBR stopband were originally designed for 980 nm. Detuning the cavity length and leaving the parameters of the DBRs unchanged shift the emitted wavelength. However, modal gain remains close to its maximal value in the limited range of detuning. Greater detuning induces significant shift of the resonant wavelength, which is more weakly reflected by the DBRs, and modal gain decreases.

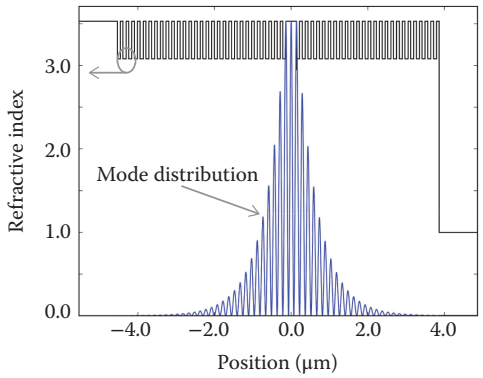

FIGURE 34.4 Distribution of the longitudinal mode and distribution of the refractive index along the optical axis of the VCSEL (Table 34.1).

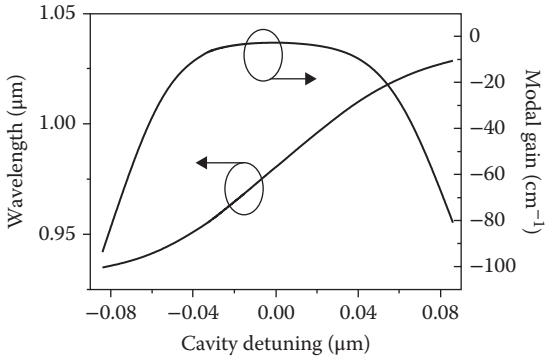

FIGURE 34.5 Modification of the resonant wavelength and modal gain induced by modification of the cavity length.

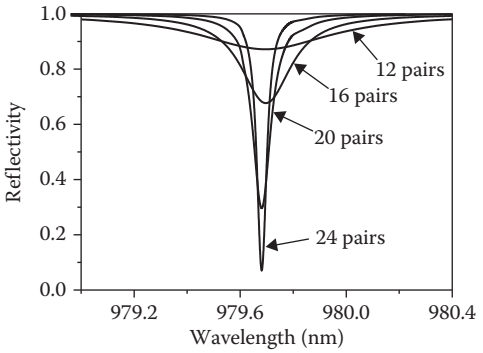

FIGURE 34.6 Power reflectance of the VCSEL wafer as a function of the incident wavelength for different numbers of DBR pairs.

Using the same method, but changing the boundary conditions to allow wave propagation toward the structure, one can calculate the reflectance or transmittance of the VCSEL wafer. In the case of a structure with an unpumped active region (in our calculations, absorption of 3000 cm^{-1} in the active region was assumed), there is a characteristic dip in the reflectance spectrum (Figure 34.6). If the dip is narrow and tends toward 0 reflectance, this indicates that the structure provides strong resonance for incident light that is absorbed by the active region. Such calculations can be compared directly with experimental measurements taken from the reflectance spectrum of the VCSEL wafer.

34.7 Two-Dimensional Scalar Approach

A more comprehensive approach, allowing lateral mode determination but still requiring moderate usage of computational memory and time, assumes that the electric field can be expressed independently using three coordinates: r, φ, and z:

$$E_u = E_u^{\text{rad}}(r)E_u^{\text{azim}}(\varphi)E_u^{ax}(z). \tag{34.28}$$

If the geometry of the device is cylindrical, one can assume the periodic dependence of

$$E_u^{\text{azim}}(\varphi) \sim \sin(m\varphi) \text{ or } E_u^{\text{azim}}(\varphi) \sim \cos(m\varphi), \tag{34.29}$$

which in practice reduces the problem to two dimensions (r, z).

Another assumption that makes use of the one-dimensional solution of (Equation 34.18) axial dependence can be expressed as

$$E_u^{ax} = A\exp\left(ik_{ax}z\right) + B\exp\left(-ik_{ax}z\right), \tag{34.30}$$

where k_{ax} is calculated for each cylinder (Figure 34.7) separately if the lateral distributions of the refractive indices in each cylinder are uniform in the radial direction. Equation 34.17 can then be transformed into

$$\left(\partial_r^2 + \frac{1}{r}\partial_r + k_{ax}^2 - k_0^2 n^2 - \frac{m^2}{r^2}\right)E_u = 0, \tag{34.31}$$

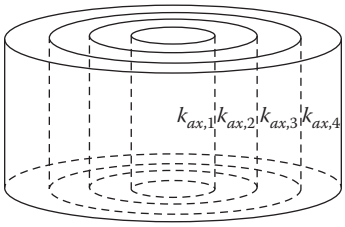

FIGURE 34.7 Schematic arrangement of cylinders. Wavenumbers $k_{ax,i}$ were calculated using one-dimensional TMM.

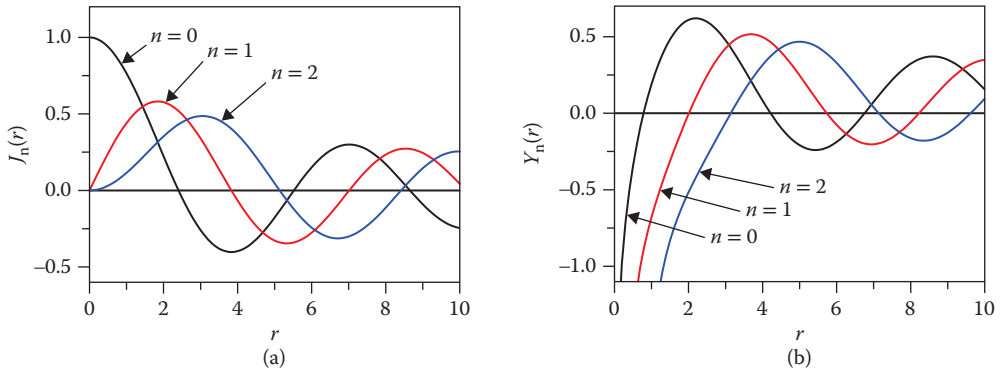

FIGURE 34.8 Bessel functions of the first (a) and second (b) kinds for $n = 0, 1, 2$.

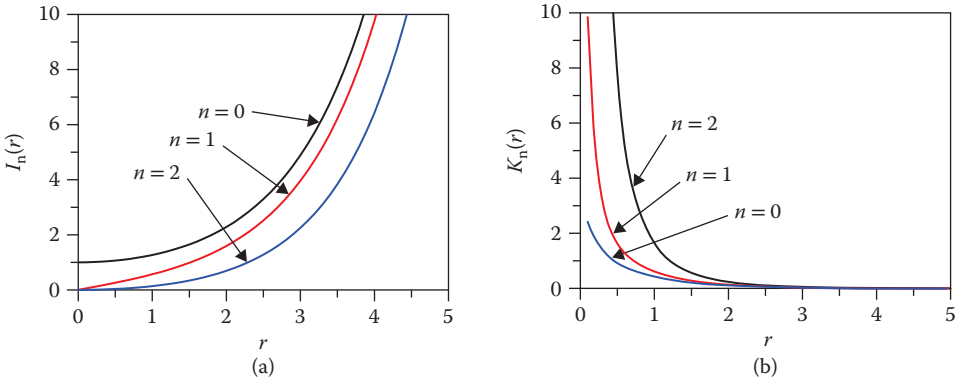

FIGURE 34.9 Modified Bessel function of the first (a) and second (b) kinds for $n = 0, 1, 2$.

which is solved separately for different m. The solutions of these Bessel differential equations are combinations of Bessel functions of the first $J_m(ar)$ and second kind $Y_m(ar)$ (Figure 34.8), where

$$a = k_{ax}^2 - k_0^2 n^2. \tag{34.32}$$

When $a < 0$, it is convenient to use modified Bessel functions of the first $I_m(-ar)$ and second kind $K_m(-ar)$ (Figure 34.9).

The following sums give the solution for an arbitrary layer:

$$E_{u,j}^{\text{rad}} = A_j J_m\left(a_j r\right) + B_j Y_m\left(a_j r\right) \tag{34.33}$$

or

$$E_{u,j}^{\text{rad}} = A_j I_m\left(-a_j r\right) + B_j K_m\left(-a_j r\right). \tag{34.34}$$

Implementing conditions for continuity of E_u^{rad} and continuity of their derivatives at the interfaces between cylinders, one can construct a transfer matrix between the electric fields in separate cylinders. In a similar manner as in Equation 34.24, one can construct a relation combining the fields from the first and last cylinders. The boundary conditions correspond to a finite value for the field at the center of the structure and to a decaying field for $r \to \infty$. The first condition eliminates $Y_m(ar)$ from the solution in the first cylinder and $I_m(-ar)$ in the last, which implies a condition similar to Equation 34.26. The set of k_0's that fulfill this condition correspond to lateral modes. Algorithms based on these approaches are very efficient and can be performed on personal computers.

The intensities of the lateral modes and the corresponding emission wavelengths of the VCSEL (Table 34.2), calculated using the algorithm detailed in this chapter, are presented in Figure 34.10. The set of modes illustrated were calculated assuming $E_u^{\text{azim}}(\varphi) \sim \cos(m\varphi)$. A corresponding set can be calculated for $E_u^{\text{azim}}(\varphi) \sim \sin(m\varphi)$, for which the emitted wavelengths will be exactly the same as for $\sim \cos(m\varphi)$. Superposition of these solutions and their complex conjugation leads to the elimination of angular dependence and only the oval shapes of the modes can be observed. However, in real-life structures in which perfect cylindrical symmetry can be broken by the geometry of the optical aperture or by the crystallographic structure of the layers, separate modes corresponding to both solutions can be observed.

TABLE 34.2 Resonant Wavelengths of Successive Lateral Modes Calculated Using Two-Dimensional Scalar Model

Mode Number	Wavelength (nm)
LP_{01}	979.3025
LP_{11}	978.8136
LP_{21}	978.175
LP_{02}	977.9555
LP_{31}	977.3976
LP_{12}	976.9519
LP_{41}	976.4901
LP_{22}	975.8181
LP_{03}	975.6216
LP_{51}	975.46
LP_{32}	974.5793
LP_{61}	974.3159
LP_{13}	974.2325
LP_{42}	973.3101
LP_{71}	973.0707
LP_{81}	971.7489

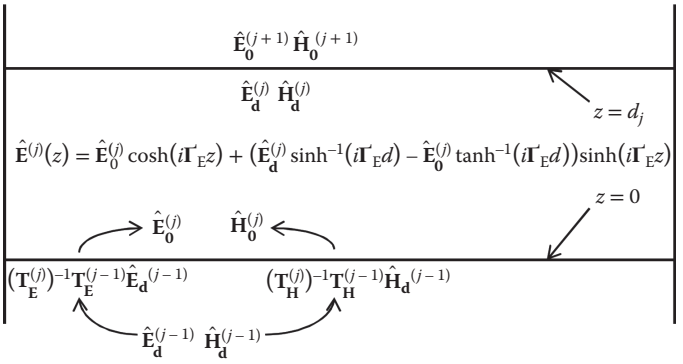

FIGURE 34.10 Definition of symbols for a single, uniform layer. $\hat{\mathbf{E}}_0$ and $\hat{\mathbf{H}}_0$ are virtual fields for the bottom interface of the layer, $\hat{\mathbf{E}}_d$ and $\hat{\mathbf{H}}_d$ are virtual fields for the upper interface, and d_j— is thickness of the layer.

34.8 Three-Dimensional Vectorial Approach

The three-dimensional vectorial approach relies on relations (Equations 34.5 and 34.6). To consider these in general geometries, we use the Cartesian coordinate system and assume that the medium in which field propagates is nonisotropic. The curl operator $\nabla\times$ in the Cartesian coordinate system is defined as follows:

$$\nabla\times \equiv \begin{bmatrix} 0 & -\partial_z & \partial_y \\ \partial_z & 0 & -\partial_x \\ -\partial_y & \partial_x & 0 \end{bmatrix}. \tag{34.35}$$

We can eliminate E_z and H_z:

$$\begin{aligned} H_z &= \begin{bmatrix} -\frac{i}{\omega\mu_z\mu_0}\partial_y & \frac{i}{\omega\mu_z\mu_0}\partial_x \end{bmatrix} \begin{bmatrix} E_x \\ E_y \end{bmatrix} \\ E_z &= \begin{bmatrix} -\frac{i}{\omega\varepsilon_z\varepsilon_0}\partial_x & \frac{i}{\omega\varepsilon_z\varepsilon_0}\partial_y \end{bmatrix} \begin{bmatrix} H_y \\ H_x \end{bmatrix}, \end{aligned} \tag{34.36}$$

which leads to the relations:

$$\begin{aligned} \partial_z\bar{\mathbf{E}} &= \begin{bmatrix} -\partial_x\frac{i}{\omega\varepsilon_z\varepsilon_0}\partial_x - \omega\mu_y\mu_0 & \partial_x\frac{i}{\omega\varepsilon_z\varepsilon_0}\partial_y \\ -\partial_y\frac{i}{\omega\varepsilon_z\varepsilon_0}\partial_x & \partial_y\frac{i}{\omega\varepsilon_z\varepsilon_0}\partial_y - \omega\mu_x\mu_0 \end{bmatrix} \bar{\mathbf{H}} = \mathbf{R_E}\bar{\mathbf{H}} \\ \partial_z\bar{\mathbf{H}} &= \begin{bmatrix} -\partial_y\frac{i}{\omega\mu_z\mu_0}\partial_y - \omega\varepsilon_x\varepsilon_0 & \partial_y\frac{i}{\omega\mu_z\mu_0}\partial_x \\ -\partial_x\frac{i}{\omega\mu_z\mu_0}\partial_y & \partial_x\frac{i}{\omega\mu_z\mu_0}\partial_x - \omega\varepsilon_y\varepsilon_0 \end{bmatrix} \bar{\mathbf{E}} = \mathbf{R_H}\bar{\mathbf{E}} \end{aligned}, \tag{34.37}$$

where

$$\bar{\mathbf{H}} = \begin{bmatrix} H_x \\ H_y \end{bmatrix} \text{ and } \bar{\mathbf{E}} = \begin{bmatrix} E_x \\ E_y \end{bmatrix}. \tag{34.38}$$

Equations 34.37 can be transformed into

$$\begin{aligned} \partial_z^2\bar{\mathbf{E}} &= \mathbf{R_H}\mathbf{R_E}\bar{\mathbf{E}} \\ \partial_z^2\bar{\mathbf{H}} &= \mathbf{R_E}\mathbf{R_H}\bar{\mathbf{H}} \end{aligned}. \tag{34.39}$$

Both equations can be solved in the base, which simplifies the matrices $\mathbf{R_H R_E}$ and $\mathbf{R_E R_H}$ into diagonal forms. The following analysis leads to the solution of Equation 34.39 and enables the characteristic value of the problem to be found. The solution corresponds to the wavenumber of the mode, as well as to the characteristic vectors, which determine the distribution of the electromagnetic field within the structure.

The equation for the electric field rewritten in the new virtual base (after diagonalization) is

$$\partial_z^2 \hat{\mathbf{E}} = \mathbf{R_H R_E} \hat{\mathbf{E}}, \tag{34.40}$$

where $\hat{\mathbf{E}}$ stand for electric field in the new, virtual base and can be defined as

$$\hat{\mathbf{E}} = \mathbf{T_E}^{-1} \bar{\mathbf{E}}, \tag{34.41}$$

where the matrix $\mathbf{T_E}$ diagonalizes $\mathbf{R_H R_E}$:

$$\mathbf{T_E}^{-1} \mathbf{R_H R_E T_E} = \mathbf{\Gamma_E}^2. \tag{34.42}$$

Here $\mathbf{\Gamma_E}^2$ is a diagonal matrix, which allows a solution of Equation 34.39 in the well-known form of a standing wave:

$$\hat{\mathbf{E}}(z) = \mathbf{A} \cosh\left(i\mathbf{\Gamma_E} z\right) + \mathbf{B} \sinh\left(i\mathbf{\Gamma_E} z\right). \tag{34.43}$$

For further derivation, we can use values of the electric field equal to $\hat{\mathbf{E}}_0$ and $\hat{\mathbf{E}}_d$ on the extreme borders of the uniform region along the z direction. The thickness of the considered uniform layer equals d (Figure 34.10).

We can therefore relate the solution to the values of the field at the opposite borders of the uniform region:

$$\hat{\mathbf{E}}(z) = \hat{\mathbf{E}}_\mathbf{0} \cosh\left(i\mathbf{\Gamma_E} z\right) + \left(\hat{\mathbf{E}}_\mathbf{j} \sinh^{-1}\left(i\mathbf{\Gamma_E} d\right) - \hat{\mathbf{E}}_\mathbf{0} \tanh^{-1}\left(i\mathbf{\Gamma_E} d\right)\right) \sinh\left(i\mathbf{\Gamma_E} z\right). \tag{34.44}$$

The solution for the magnetic field can be found in a similar manner.

Using the solution for $\hat{\mathbf{E}}$ and $\hat{\mathbf{H}}$ and substituting to Equation 34.37, one can find the relation connecting electric and magnetic fields within the layer:

$$\begin{bmatrix} \hat{\mathbf{H}}_0 \\ -\hat{\mathbf{H}}_d \end{bmatrix} = \begin{bmatrix} \mathbf{y_1} & \mathbf{y_2} \\ \mathbf{y_2} & \mathbf{y_1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{E}}_0 \\ \hat{\mathbf{E}}_d \end{bmatrix}, \tag{34.45}$$

where

$$\mathbf{y_1} = \left(\mathbf{T_E}^{-1} \mathbf{R_H T_H}\right)^{-1} \mathbf{\Gamma_E} \tanh^{-1}\left(i\mathbf{\Gamma_E} d\right), \tag{34.46}$$

$$\mathbf{y}_2 = -\left(\mathbf{T_E}^{-1} \mathbf{R_H T_H}\right)^{-1} \mathbf{\Gamma_E} \sinh^{-1}\left(i\mathbf{\Gamma_E} d\right). \tag{34.47}$$

The relation between the fields from the ith and $(i-1)$th layers is given in the formulas:

$$\hat{\mathbf{E}}_0^{(i)} = \left(\mathbf{T_E}^{(i)}\right)^{-1} \mathbf{T_E}^{(i-1)} \hat{\mathbf{E}}_d^{(i-1)} = \mathbf{t_E}^{(i)} \hat{\mathbf{E}}_d^{(i-1)}, \tag{34.48}$$

$$\hat{\mathbf{H}}_0^{(i)} = \left(\mathbf{T_H}^{(i)}\right)^{-1} \mathbf{T_H}^{(i-1)} \hat{\mathbf{H}}_d^{(i-1)} = \mathbf{t_H}^{(i)} \hat{\mathbf{H}}_d^{(i-1)}. \tag{34.49}$$

Using Equations 34.45, 34.48, and 34.49, the relationship between the electric and magnetic fields, for the interface separating layers $i+1$ and i, takes an iterative formula:

$$\hat{\mathbf{H}}_d^{(i)} = \mathbf{Y}^{(i)}\hat{\mathbf{E}}_d^{(i)}, \tag{34.50}$$

$$\mathbf{Y}^{(i)} = -\left(\mathbf{y}_2^{(i)}\left(\mathbf{t}_{\mathbf{H}}^{(i)}\mathbf{Y}^{(i-1)}\left(\mathbf{t}_{\mathbf{E}}^{(i)}\right)^{-1} - \mathbf{y}_1^{(i)}\right)^{-1}\mathbf{y}_2^{(i)} + \mathbf{y}_1^{(i)}\right). \tag{34.51}$$

Propagating by this procedure from the bottom limit of the simulated structure to the upper limit, the characteristic equation can be determined. However, from the numerical point of view, it is more efficient and precise to perform the procedure starting from the bottom limit to the center, and then to start again from the upper limit down to the center. The center does not correspond to the geometrical center but to the plane at which the electromagnetic field reaches or it is close to its maximum, which is termed as the "matching interface." One can write the relation (Equation 34.50) for the top-down and bottom-up algorithms as:

$$\hat{\mathbf{H}}_d^{(m)} = \mathbf{Y}_{\text{up}}^{(m)}\hat{\mathbf{E}}_d^{(m)}, \tag{34.52}$$

$$\hat{\mathbf{H}}_d^{(l)} = \mathbf{Y}_{\text{down}}^{(l)}\hat{\mathbf{E}}_d^{(l)}. \tag{34.53}$$

m counts layers from the bottom and l from the top.

Since the magnetic field is continuous in the matching interface in the real base, one can write the relation

$$\bar{\mathbf{H}}_d^{(m)} = \bar{\mathbf{H}}_d^{(l)}. \tag{34.54}$$

Transforming back Equation 34.54 to the virtual base:

$$\mathbf{T}_{\mathbf{H}}^{(m)}\hat{\mathbf{H}}_d^{(m)} = \mathbf{T}_{\mathbf{H}}^{(m)}\hat{\mathbf{H}}_d^{(l)}, \tag{34.55}$$

using Equations 34.52 and 34.53:

$$\mathbf{T}_{\mathbf{H}}^{(m)}\mathbf{Y}_{\text{up}}^{(m)}\hat{\mathbf{E}}_d^{(m)} = \mathbf{T}_{\mathbf{H}}^{(m)}\mathbf{Y}_{\text{down}}^{(l)}\hat{\mathbf{E}}_d^{(l)}, \tag{34.56}$$

and finally using relation (Equation 34.41) to lead us to an equation that includes the fields in the real base:

$$\left(\mathbf{T}_{\mathbf{H}}^{(m)}\mathbf{Y}_{\text{up}}^{(m)}\left(\mathbf{T}_{\mathbf{E}}^{(m)}\right)^{-1} - \mathbf{T}_{\mathbf{H}}^{(m)}\mathbf{Y}_{\text{down}}^{(l)}\left(\mathbf{T}_{\mathbf{E}}^{(l)}\right)^{-1}\right)\bar{\mathbf{E}} = \mathbf{Y}\bar{\mathbf{E}} = 0. \tag{34.57}$$

Equation 34.57 is also the characteristic value equation. The solution of the equation determines the characteristic values, which correspond to the complex wavenumber. As before, the imaginary part of the characteristic value relates to the modal gain of the propagating wave. Equation 34.57 has a nontrivial solution only if matrix $\mathbf{Y}$ is singular, which imposes the condition that one of the eigenvalues of $\mathbf{Y}$ is 0. This condition is useful in numerical searches for solutions to Equation 34.57.

The intensity distributions of lateral modes calculated using the vectorial approach are very close to those depicted in Figure 34.11. The main difference between the solutions lies in the different number of modes (Table 34.3). In the vectorial solution, there are no modes of shorter wavelengths than LP_{22}. This may be associated with diffraction losses, which are more severe for higher order modes but are not included in the scalar model.

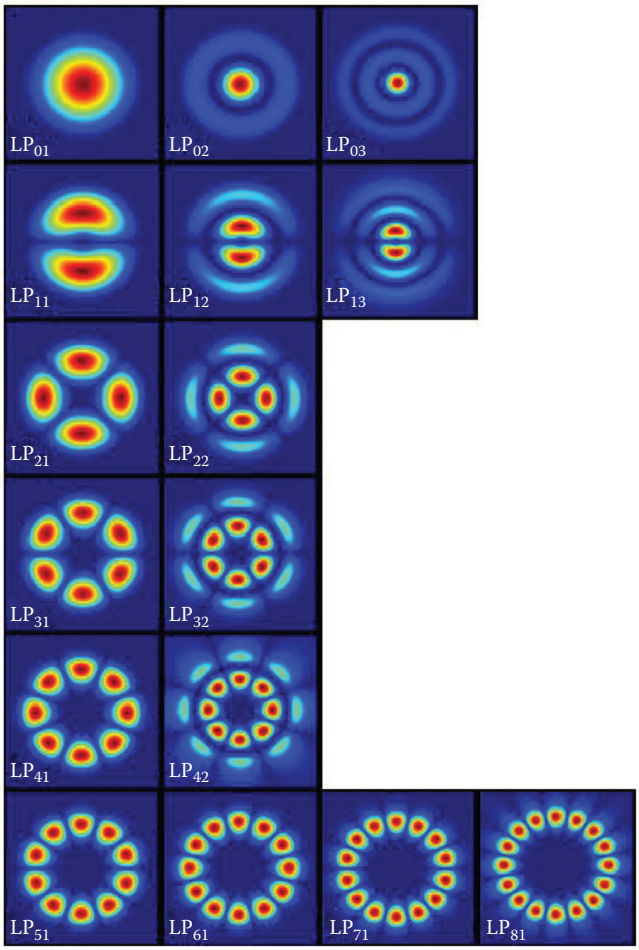

FIGURE 34.11 Distributions of lateral modes in the VCSEL structure detailed in Table 34.1, where $R = 4$ μm, calculated using the scalar two-dimensional approach.

TABLE 34.3 Resonant Wavelengths of Successive Lateral Modes Calculated Using the Three-Dimensional Vectorial Model

Mode Number	Wavelength (nm)
LP_{01}	979.3014
LP_{11}	978.8108
LP_{21}	978.1691
LP_{02}	977.9616
LP_{31}	977.3912
LP_{12}	976.9431
LP_{41}	976.4811
LP_{22}	975.8031

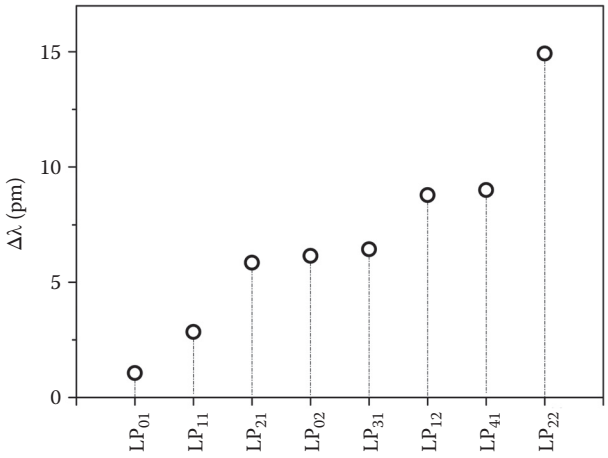

FIGURE 34.12 Difference in resonant wavelengths of successive lateral modes calculated using scalar and vectorial models.

Figure 34.12 illustrates the difference ($\Delta\lambda$) with respect to the wavelength of modes calculated using scalar and vectorial models. The vectorial model provides solutions with shorter wavelengths. However, the difference is on a scale of picometers. The monotonic increase in $\Delta\lambda$ shows that the plane wave assumption used in the scalar model is more accurate in the case of lower order modes.

34.9 Impact of Oxide Aperture Diameter and Position

Finally, we shall analyze the influence of oxide aperture on the emitted wavelength of the VCSEL considered in this study. Although the structure is a very simplified design, the general tendencies presented in Figures 34.13 and 34.14 are applicable to real-life oxide-confined VCSEL structures.

Reducing the aperture (Figure 34.13) size leads to a characteristic blue shift, since a narrower aperture introduces stronger interaction between the mode and the oxide layer, which has a low refractive index. The wider the aperture diameter, the closer the values for the wavelengths determined using each approach.

The location of the oxide layer has a significant effect on the VCSEL parameters (Figure 34.14). If it is positioned at the antinode of the standing wave, the interaction between the optical field and carriers within the active region is intensified, since the modes are strongly confined within the aperture. This configuration reduces the threshold current and broadens the spectrum of emission, as numerous lateral modes are excited. If the oxide layer is in the node position, the number of modes is reduced, which narrows the

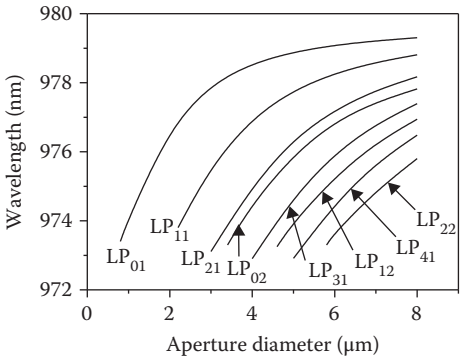

FIGURE 34.13 Dependence of the lateral mode wavelengths on aperture diameter.

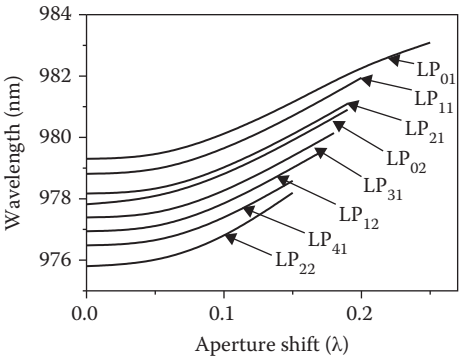

FIGURE 34.14 Dependence of lateral mode wavelengths on aperture shift from the antinode position relative to the wavelength of the standing wave.

emitted spectrum. However, interaction between the mode and the oxide layer is weaker and the threshold current increases. Shifting the oxide layer from the antinode position leads to an increase in the wavelength. This can be explained by weaker interaction between the field and the oxidized layer, which leads to an increase in the effective index of the mode and eventually causes redshift.

34.10 Conclusion

Scalar one- and two-dimensional optical approaches represent very efficient tools for VCSEL designers. Their calculation algorithms, requiring short calculation times and minor memory storage, enable one to perform rapid calculations using even PCs. However, the assumptions that make them so efficient introduce simplifications, which may produce inaccurate results in more sophisticated structures. Most significantly, the basic scalar assumption of plane optical wave propagation is violated in the case of aperture sizes comparable to the radiation wavelength and for higher order modes.

Vectorial methods are much more complicated, since they are based on the solution of six coupled equations instead of only one (in the simplified scalar approach). They give more exact results at the price of significantly longer computational time and a considerably more complicated algorithm. Vectorial methods may not be the optimal approach to modeling a standard VCSEL. However, analysis of VCSELs with in-built refractive index steps in lateral directions or subwavelength patterning certainly requires the use of vectorial methods.

References

1. Melngailis, I. Longitudinal injection plasma laser of InSb. *Appl Phys Lett*. 6, 59–60 (1965).
2. Soda, H. et al. GaInAsP/InP surface emitting injection lasers. *Jap J Appl Phys*. 18, 2329–2330 (1979).
3. Ogura, M. et al. GaAs/$Al_xGa_{1-x}As$ multilayer reflector for surface emitting laser diode. *Jap J Appl Phys*. 22, L112–L114 (1983).
4. MacDougal, M. H. et al. Ultralow threshold current vertical-cavity surface-emitting lasers with AlAs oxide-GaAs distributed Bragg reflectors. *IEEE Photon Technol Lett*. 7, 229–231 (1995).
5. Tai, K. et al. Use of implant isolation for fabrication of vertical cavity surface emitting lasers. *Electron Lett*. 25, 1644–1645 (1989).
6. Gadallah, A.-S. et al. High-output-power single-higher-order transverse mode VCSEL with shallow surface relief. *IEEE Photon Technol Lett*. 23, 1040–1042 (2011).

7. Zhou, D. et al. High-power single-mode antiresonant reflecting optical waveguide-type vertical-cavity surface-emitting lasers. *IEEE J Quantum Electron*. 38, 1599–1606 (2002).
8. Kapon, E. et al. Long-wavelength VCSELs: Power-efficient answer. *Nat Photon*. 3, 27–29 (2009).
9. Danner, A. J. et al. Progress in photonic crystal vertical cavity lasers. *IEICE Trans Electron*. E88, 944–950 (2005).
10. Chang-Hasnain, C. J. et al. High-contrast gratings for integrated optoelectronics. *Adv Opt Photon*. 4, 379–440 (2012).
11. Saleh, B. E. A., Teich, M. C. *Fundamentals of Photonics*. New York, NY: John Wiley & Sons (1991).
12. Bienstman, P. R. et al. Comparison of optical VCSEL models on the simulation of oxide-confined devices. *IEEE J Quant Electron*. 37, 1618–1631 (2001).
13. Kuszelewicz, R. *COST 268 Modeling Exercise*. Paris, France: CNET (1998).
14. Bergmann, M. J. et al. Optical-field calculations for lossy multiple-layer $Al_xGa_{1-x}N/In_xGa_{1-x}N$ laser diodes. *J Appl Phys*. 84, 1196–1203 (1998).

VII

Photodetectors and Modulators

35

Photodetector Fundamentals

Prasanta Basu

35.1 Introduction

Photodetectors (PDs) convert light energy into electrical energy, which usually manifests as a photocurrent. They are a kind of transducer. The electrical signal is useful in electronic imaging, providing a means for viewing images much faster than the earlier photographic imaging. PDs also find important applications in present-day fiber-optic communication systems. In this case, PDs receive the stream of weak optical pulses at the far end of the optical fiber originally transmitted by the transmitter. The role of the detector is to convert these optical pulses into electrical pulses, which are then processed for use in a telephone, a computer, or other terminals at the receiving end (Bhattacharyya 1996; Deen and Basu 2012; Kumar and Deen 2014).

Other important applications of PDs include automatic door opener, use in remote control elements in TV, VCR, air conditioners, etc., in CCDs in a video camera, and in astronomy, research, and defense.

This chapter gives an introduction to the working principle of photodetection by semiconductors, the material properties, different types, and their structures, important parameters pertaining to PDs, and materials and makes a list of the application areas.

35.2 Basic Working Principle

The conversion of optical energy into electrical energy by a PD involves three basic processes. First, a bunch of incident photons is absorbed by the semiconducting material to create excess electron–hole pairs (EHPs). The carriers are then transported across the absorption or transit region. In the third stage, the transported carriers are collected to generate photocurrent. In this section, we consider the basic principle, structure, and the process of absorption (Bhattacharyya 1996; Deen and Basu 2012).

35.2.1 Conversion of Light into Current

The process of creation of excess EHPs and their collection is illustrated in Figure 35.1. In Figure 35.1a, a photon of energy $\hbar\omega \geq E_g$, incident on the semiconductor, is absorbed by an electron in the valence band (VB), which then moves up in the conduction band (CB), leaving a hole in the VB. An excess EHP, in excess of the thermal equilibrium values, is therefore created. To collect these carriers, a reverse-biased p-n junction is employed. As shown in Figure 35.1b, the electric field in the depletion layer drives the excess electrons and excess holes in the opposite directions to be collected by the terminals and thus to produce current. The reverse saturation current in the absence of illumination is called the dark current. The photogenerated current is proportional to the flux of incident photons or the power of incident light (Kaiser 2013; Senior and Jamro 2009).

As shown in Figure 35.1b, absorption may also take place outside the depletion layer. The excess EHPs generated therein move by the slower process of diffusion and are collected by the terminals. The diffusion process limits the speed of operation and distorts the electrical pulses.

35.2.2 Absorption in Semiconductors

The process of absorption is now discussed in some detail in Figure 35.2. In Figure 35.2a, a photon of energy $\hbar\omega \geq E_g$ is shown incident on the semiconductor. The photon lifts an electron from the VB to a higher lying state in the CB. The energetic electron then relaxes to the bottom of the CB via phonon emission. A photon thus creates an excess EHP. The pair ultimately recombines to give rise to emission (luminescence). The same processes are illustrated in the E-k diagrams of a direct gap semiconductor in Figure 35.2b and of an indirect gap semiconductor in Figure 35.2c. The photon momentum $\hbar\mathbf{k}_{photon}$ is usually negligible and the momentum ($\mathbf{k}$) conservation can be written as $\mathbf{k}_e = \mathbf{k}_h + \mathbf{k}_{photon} \approx \mathbf{k}_h$, where $\mathbf{k}_e(\mathbf{k}_h)$ denotes the wave vector for electron (hole). Accordingly, the absorption in a direct gap semiconductor is indicated by a vertical

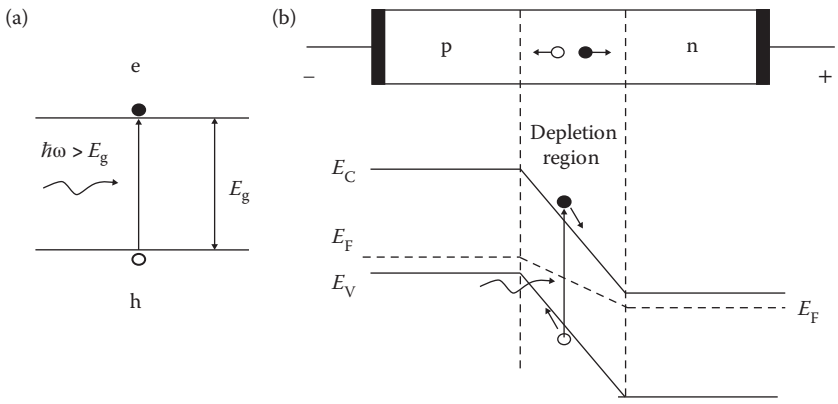

FIGURE 35.1 Conversion of optical energy into electrical energy: (a) absorption of photons and generation of excess EHPs; (b) collection of carriers by using a reverse-biased p-n junction.

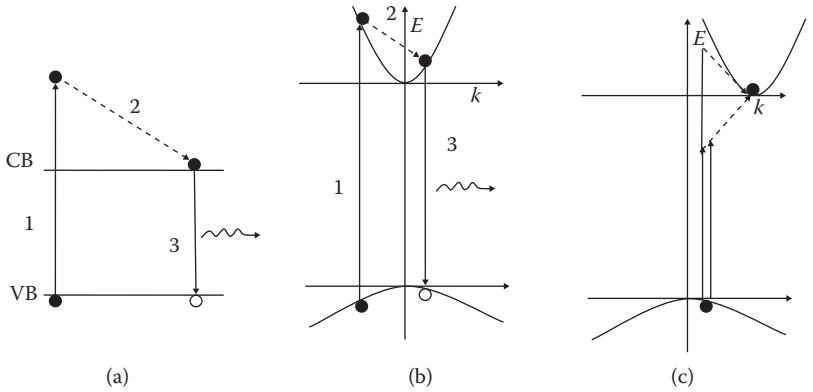

FIGURE 35.2 (a) Absorption (1), relaxation (2), and recombination (3) processes in semiconductors. CB (VB) denotes conduction (valence) band edges; (b) absorption (1), relaxation (2), and recombination (3) processes shown in the E-k diagram of a direct gap semiconductor; (c) the same processes in the E-k diagram of an indirect gap semiconductor. Dashed lines correspond to phonon scattering processes.

line as shown in Figure 35.2b. However, for indirect gap semiconductor, an extra momentum is needed to place the electron in the CB, which is provided by, among other agencies, phonons. In Figure 35.2c, the indirect absorption is indicated by both phonon absorption and emission. The indirect absorption has a very low probability and orders of magnitude being lower than in direct gap semiconductors (Basu 1997).

The variation of light intensity inside a semiconductor, $I(z)$, is governed by the following expression known as the Lambert's law:

$$I(z) = I_0 \exp(-\alpha z), \tag{35.1}$$

where I_0 is the intensity at the surface ($z = 0$), and α is the absorption coefficient of the material, which is a function of photon energy or wavelength.

The absorption coefficient for direct gap semiconductors is much higher than for indirect gap semiconductors.

35.2.2.1 Absorption in Direct Gap Semiconductors

As mentioned already, the absorption process in a direct gap semiconductor is indicated by a vertical line in the E-k diagram. The absorption coefficient in direct gap semiconductors is calculated by using the first-order time-dependent perturbation theory and the Fermi golden rule. The expression for absorption coefficient is of the form (Basu 1997)

$$\alpha(\hbar\omega) = A\left(\hbar\omega - E_g\right)^{1/2}\left(f_c - f_v\right), \hbar\omega \geq E_g, \tag{35.2}$$

where the prefactor A contains the fundamental constants, material parameters such as permittivity, etc., and $f_{c(v)}$ denotes the Fermi occupation probabilities of electrons in the CB (VB).

35.2.2.2 Absorption in Indirect Gap Semiconductors

Since a photon and a phonon are involved in the absorption process, the second-order perturbation theory is used to calculate the absorption coefficient in indirect gap semiconductors. The expression takes the

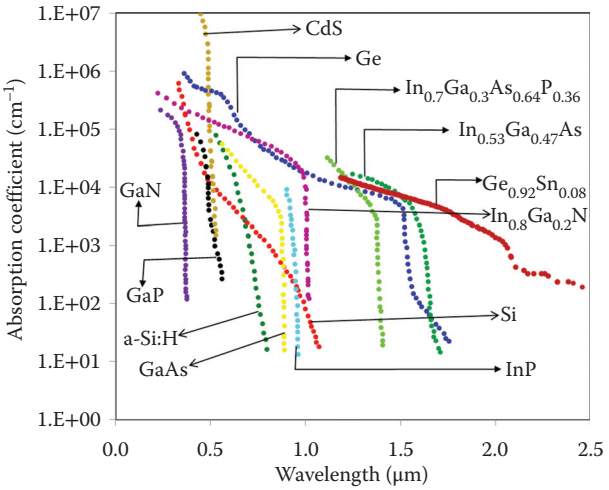

FIGURE 35.3 Absorption spectra of some common semiconductors.

following form at 0 K when a single type of phonon of energy $\hbar\omega_p$ is involved (Basu 1997):

$$\alpha(\hbar\omega) = A_a\left(\hbar\omega - E_g + \hbar\omega_p\right)^2 + A_e\left(\hbar\omega - E_g - \hbar\omega_p\right)^2, \tag{35.3}$$

where the first and second terms correspond, respectively, to the absorption followed by phonon absorption and that followed by phonon emission, and the prefactors A's are distinguished by the corresponding suffixes. There is, in general, a possibility of involvement of different types of phonons. Figure 35.3 shows typical absorption coefficients of some representative semiconductors.

35.2.2.3 Absorption in Quantum Nanostructures

Recent detectors make increasing use of quantum nanostructures such as quantum wells (QWs), quantum wires (QWRs), and quantum dots (QDs), rather than the bulk semiconductors. In QWs, for example, the thickness of the layer of interest is comparable to the de Broglie wavelength of the particle. A QW is realized by sandwiching a thin layer of GaAs between two layers of a higher gap material such as $Al_xGa_{1-x}As$. The typical band diagram of the QW is shown in Figure 35.4a. Due to very small thickness d (<de Broglie wavelength) of GaAs, the motion of electrons and holes is quantized, and assuming that the barrier heights are infinite, the quantized energy levels, or subbands as they are called, are expressed as follows:

$$E_n = \frac{\hbar^2}{2m_{\text{eff}}}\left(\frac{n\pi}{d}\right)^2, \quad n = 1, 2, 3, \ldots. \tag{35.4}$$

The effective mass, m_{eff}, is to be replaced by the electron effective mass, m_e, heavy hole effective mass, m_{hh}, and light hole effective mass, m_{lh}, in order to calculate the respective subband energies (Basu 1997).

The absorption processes in a QW are shown in Figure 35.4b. Here, transitions can occur in several ways. An electron from a valence subband (both lh and hh) can absorb a photon and can come to one of the subbands in the CB. This is a typical band-to-band absorption; however, the absorption threshold, that is, the photon energy marking the sharp rise in the absorption coefficient, increases from the threshold in bulk material that equals the bandgap (Equation 35.2). In addition, an electron (or a hole) in a lower subband (i) can absorb a photon and may move up to a higher subband (j) in the same band. This is a representative intersubband transition. The absorption processes in QWs are governed by some selection rules.

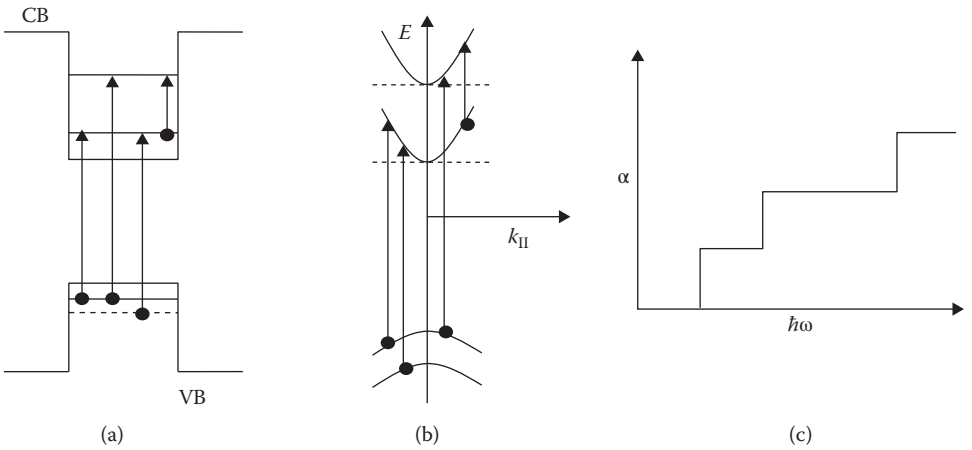

FIGURE 35.4 (a) Band diagram and quantized energy levels or subbands in a QW structure; (b) optical absorption processes in a QW; both band-to-band and intersubband processes are shown; (c) idealized absorption spectra for QWs. α and $\hbar\omega$ denote the absorption coefficient and the photon energy, respectively. k_{II} is the wave vector along the QW layer plane. CB (VB) denotes conduction (valence) band edges.

The idealized band-to-band absorption coefficient in the QW of a direct gap semiconductor takes the following form (Basu 1997):

$$\alpha(\hbar\omega) = K \sum_n H\left(\hbar\omega - E_{nn}\right), \tag{35.5}$$

where the prefactor K contains fundamental constants, effective (reduced) masses of electrons and holes, and the permittivity of the QW material. In Equation 35.5, H is a Heaviside step function, and E_{nn} is the energy difference between nth hole and nth electron subband.

Typical absorption spectra for QWs obtained from Equation 35.5 are shown in Figure 35.4c, which indicates that the spectra are staircase like.

The idea can be extended to the cases of QWRs and QDs, in which the motion of the particles is inhibited along two and three directions, respectively.

The absorption spectra get modified when excitonic effects are included. Excitons are complexes formed by the EHP bound by the mutual Coulomb interaction.

35.3 Important Characteristics

In this section, some of the important characteristics of PDs are discussed.

35.3.1 Wavelength Selectivity

Different materials are used for the detection of photons at different wavelengths. The wavelength range covered by a particular material is determined by the cutoff wavelength, above which the absorption by it is negligibly small. The cutoff wavelength is determined by the expression $\hbar\omega = E_g$, where E_g is the bandgap, from which it is simple to write

$$\lambda_c = \frac{hc}{E_g} = \frac{1.24}{E_g(\text{eV})}\mu\text{m}. \tag{35.6}$$

For a bandgap of 1 eV, the cutoff wavelength is 1.24 μm.

35.3.2 Quantum Efficiency

Quantum efficiency (QE), sometimes called the internal quantum efficiency (IQE), is the ratio of the generated EHPs and the number of incident photons. The higher the QE is, larger is the sensitivity of the PD and it is therefore an important figure of merit.

More practical measure of the sensitivity is the external QE (EQE), which is the ratio of EHPs collected at the terminals to the number of incident photons. EQE is less than IQE since a part of the incident photon flux is reflected at the surface and some EHPs are lost by recombination at the surface or by nonradiative processes. EQE is a function of wavelength.

Let I_{pc} be the photocurrent, P_i be the incident power, and $\hbar\omega$ be the photon energy. By definition,

$$\eta = \frac{\text{electron generation rate}}{\text{photon incidence rate}} = \frac{I_{\text{pc}}/e}{P_i/\hbar\omega} = \frac{I_{\text{pc}}}{P_i}.\frac{\hbar\omega}{e} = \frac{1.24}{\lambda\,(\mu\text{m})}\frac{I_{\text{pc}}}{P_i}. \tag{35.7}$$

Equation 35.7 may also be written as a product of three factors as follows:

$$\eta = \zeta(1 - e^{-\alpha d})(1 - R). \tag{35.8}$$

In Equation 35.8, ζ is the fraction of the photogenerated carriers that reach the outer circuit without recombination, the second term within parentheses denotes the fraction of light absorbed in the absorption region of width d, and the last term represents the amount of light transmitted into the semiconductor, R being the reflection coefficient.

35.3.3 Responsivity

Responsivity is defined as the ratio of the primary photocurrent without gain, I_{pc}, and the incident optical power and is expressed as

$$\mathfrak{R} = \frac{I_{\text{pc}}}{P_l} = \eta(\lambda)\left[\frac{e}{(hc/\lambda)}\right]. \tag{35.9}$$

In PDs, responsivity, or sensitivity as it is alternately called, may be increased by increasing the absorption region thickness and minimizing reflection from its surface. The latter can be accomplished by coating the surface with an antireflection (AR) coating of refractive index $n_{\text{AR}} = \sqrt{n_{\text{air}} n_{\text{SC}}}$, where the subscripts refer to AR material, air, and semiconductor, respectively.

35.3.4 Response Time

The speed of response of PDs is limited by three main factors, as discussed next.

35.3.4.1 Diffusion Time of Carriers

The time taken by carriers to diffuse a distance d is given by (Streetman and Banerjee 2016)

$$\tau_{\text{diff}} = \frac{d^2}{2D}, \tag{35.10}$$

where D is the diffusion constant. Diffusion is a slower process than drift. The carriers that are optically generated within a diffusion length from the depletion layer of a photodiode move to the depletion layer by this slow process. The photogenerated current pulse due to a narrow optical pulse shows "tails" due to the slow diffusion process.

35.3.4.2 Drift Time through Depletion Region

The drift velocity of both electrons and holes in a semiconductor attains a saturation value in the presence of very high fields (Streetman and Banerjee, 2016). Usually, a high electric field exists in the depletion layer of a reverse-biased junction. The longest transit time of carriers in the depletion layer of width W may therefore be written as follows:

$$\tau_{\text{drift}} = \frac{W}{v_{\text{sat}}}, \tag{35.11}$$

where v_{sat} is the saturation velocity of the carrier, which is different for electrons and holes. The maximum speed of operation in a PD is limited by transit time.

35.3.4.3 Junction Capacitance Effect

The electrical equivalent circuit of a PD using a p-i-n diode as an example is shown in Figure 35.5. It includes a photogenerated current source (I_{ph}), the depletion layer capacitance (C_{J}), a series resistance (R_{S}) due to bulk resistance and ohmic contacts, a parasitic capacitance (C_{P}), and a load resistance (R_{L}). The electrical 3-dB bandwidth may be expressed as

$$f_{3\text{dB}} = \frac{1}{2\pi(C_{\text{J}} + C_{\text{P}})(R_{\text{L}} + R_{\text{S}})}. \tag{35.12}$$

35.3.5 Noise Equivalent Power and Detectivity

The term *noise* refers to unwanted electrical signal that masks the signal to be detected. Several noise sources, such as thermal, shot, dark current, and flicker noise, are present in PDs and associated circuits and make it difficult to detect weak signals. The minimum detectable signal corresponds to an rms output signal equal to that generated by noise. The signal-to-noise ratio (SNR) is then unity. The noise equivalent power (NEP), a measure of the minimum detectable signal, is defined as the power of sinusoidally modulated monochromatic optical signal that generates the same rms output power in the ideal noise-free detector as the noise signal produced in a real detector.

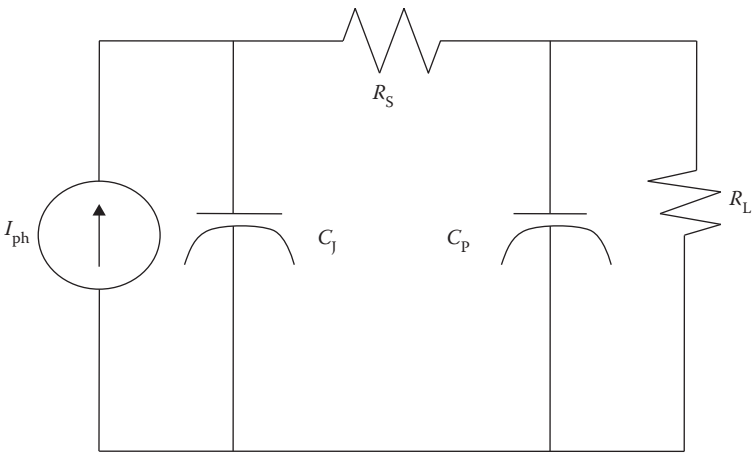

FIGURE 35.5 Electrical equivalent circuit of a photodiode. I_{ph} is the photogenerated current, and C_{J} and C_{p} are, respectively, the junction and parasitic capacitances. R_{S} and R_{L} are, respectively, the series and load resistances.

A unit NEP^* is defined as

$$NEP^* = \frac{NEP}{(A\Delta f)^{1/2}}, \tag{35.13}$$

where $A$is the area of the detector and Δf the bandwidth. The commonly used parameter is, however, the specific detectivity D^*, which is the reciprocal of NEP^*.

35.4 Classification of PDs

PDs can be classified in different ways. It may be an intrinsic or an extrinsic type. In the former, the absorption is band to band and the excess EHPs created produce a photocurrent. In the extrinsic types, the photon energy is less than the bandgap energy, and an impurity or deep-level defect in the forbidden gap is involved. An electron from this deep level is lifted by photon absorption process to the CB. Alternately, an electron in the VB may come to the deep level by photon absorption leaving behind a hole in the VB. In both the cases, the excess carriers created produce photocurrent. The absorption may also take place between two confined states in the CB or VB in a QW, QWR, or QD. Such processes give rise to the intersubband transitions.

PDs can be of two types: with or without internal gain. Photoconductors, avalanche PDs (APDs), and phototransistors show internal gain, that is, the primary photocurrent is multiplied. On the other hand, p-n or rather p-i-n PDs do not have any gain.

In another classification scheme, PDs may be classified as having either vertically illuminated or horizontally illuminated structures. In the former type, the absorbing layer thickness is low to reduce the transit time and to be useful for high-speed operation, though sacrificing the QE. However, a high packing density of arrays of PDs can be achieved, also facilitating coupling of the PDs with fibers. The low QE may be increased by enclosing the active layer within a resonant cavity structure, which is realized by using multilayer dielectrics (Unlu et al. 2001).

The simplest example of a horizontally illuminated PD is an edge-illuminated detector. However, due to diffraction, light spreads outside the high-field region and slow diffusion process leads to tails in the impulse response. The problem may be solved by using a waveguide structure, confining the light into the high-field region covering the optical waveguide. The speed of the devices is limited by the capacitance due to a thin intrinsic absorbing layer. A traveling wave detector may solve the limitations due to capacitance.

Another way of classifying the PDs is based on the type of absorbing layer. Various possibilities exist that include (1) bulk semiconductor, (2) QWs/multiple QWs (MQWs), (3) QWRs, (4) QDs, (5) strained quantum nanostructures, and (6) quantum nanostructures of different band alignments (see Basu 1997) such as type I, type II, type III, etc., as the active absorbing region.

35.5 Different PD Configurations

In this section, we shall discuss the structure and working principles of photoconductors, p-i-n and APDs, and PDs involving intersubband transitions. PDs relying on extrinsic processes will not be addressed. Detailed discussions about different types are given in many text books and reviews (see, for example, Bhattacharyya 1996; Bowers and Burrus 1987; Bowers and Wey 1995; Deen and Basu 2012; Kaiser 2013; Senior and Jamro 2009)

35.5.1 Photoconductive Detectors

The simplest form of PD is a photoconductor, in which the conductivity of the semiconductor is changed by absorption of photons. The changed photoconductivity produces photocurrent proportional to the incident optical power. In this section, the structure, principle, and noise analysis of such photoconductive detectors are presented.

35.5.1.1 Structure and Analysis

Consider a slab of semiconductor of length L, width W, and thickness d as shown in Figure 35.6. A voltage V is applied across the length and the slab is illuminated by light at the top. The current I flowing through the photoconductor may be expressed as

$$I = eWd(n\mu_e + p\mu_h)\left(\frac{V}{L}\right) = eWd\left(\frac{V}{L}\right)[(n_0\mu_e + p_0\mu_h) + (\Delta n\mu_e + \Delta p\mu_h)] = I_D + I_{ph}, \tag{35.14}$$

where I_D is the dark current, I_{ph} is the photocurrent, n_0 and p_0 are the equilibrium electron and hole concentrations, respectively, Δn (Δp) is the excess electron (hole) concentration, and μ_e (μ_h) is the electron (hole) mobility.

The average generation rate r_g within the photoconductor is expressed as follows:

$$r_g = \frac{\eta I_0 WL}{\hbar\omega WLd} = \frac{\eta I_0}{\hbar\omega d}, \tag{35.15}$$

where η is the efficiency, I_0 the light intensity at the surface, and $\hbar\omega$ the photon energy. Under steady state, the generation rate equals the recombination rate so that $r_g = r_r = \Delta n/\tau = \Delta p/\tau$, where τ is the recombination lifetime. Thus, one obtains

$$\Delta n = \Delta p = r_g\tau. \tag{35.16}$$

Substituting Equation 35.16 in Equation 35.14, the photocurrent is obtained as follows:

$$I_{ph} = \frac{Wd}{L} r_g \tau e(\mu_e + \mu_h)V. \tag{35.17}$$

The photoconductive gain, defined as the ratio of the rate of flow of electrons from the device to the rate of generation of EHPs within the device, takes the following form:

$$G = \frac{I_{ph}}{e}\frac{1}{r_g WLd} = \frac{\tau(\mu_e + \mu_h)V}{L^2}. \tag{35.18}$$

Equation 35.17 has been used to arrive at the last equality in Equation 35.18. The variation in gain with voltage V may be understood if we assume that $\mu_e >> \mu_h$, which is usually the case. The transit time of electrons through the device is $\tau_{tre} = L^2/\mu_e V$ and from Equation 35.18 $G = \tau/\tau_{tre}$. At low bias, $\tau < \tau_{tre}$ and the gain is <1. With larger V, electrons move faster, $\tau > \tau_{tre}$, and the gain becomes larger than unity.

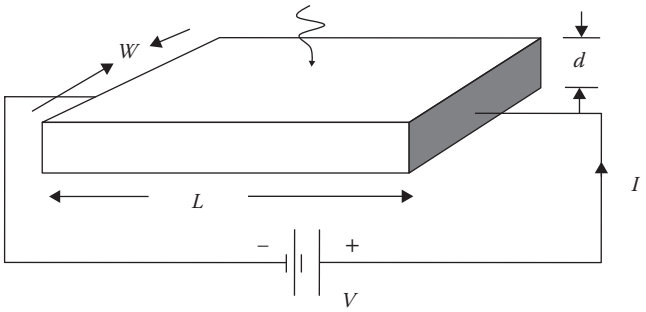

FIGURE 35.6 Schematic of a photoconductive PD. L, W, and d are the length, width, and depth of the photoconductor, respectively. V is the applied voltage and I is the current.

At still larger bias, the velocities of both electrons and holes are saturated. The gain is again reduced below unity. By following these arguments, it is easily concluded that higher values of gain may be achieved by increasing V, or by decreasing L, or by increasing the recombination lifetime.

35.5.1.2 Noise in Photoconductors

There are three main sources of noise in a photoconductive detector. The first is thermal or Johnson noise that arises due to random motion of carriers in the device. The mean squared noise current due to a conductor having finite resistance R is given as follows:

$$\left\langle i_{\text{th}}^2 \right\rangle = \frac{4k_{\text{B}}TB}{R}, \tag{35.19}$$

where B is the bandwidth of the device. In developing the noise equivalent circuit, the noisy photoconductor is replaced by a noiseless photoconductor and an ideal noise current or voltage generator.

Fluctuations in generation and recombination rates give rise to fluctuations in current and hence generate noise. The mean squared generation-recombination (GR) current is expressed as follows:

$$\left\langle i_{\text{GR}}^2 \right\rangle = \frac{4eGI_0B}{1+\omega^2\tau^2}, \tag{35.20}$$

where I_0 is the steady photogenerated current. GR noise is also termed as shot noise. Since shot noise decreases as ω^{-2}, the thermal noise is the dominating noise source at high frequencies.

The third kind of noise, the flicker noise, has its origin in the presence of surface and interface defects and traps in the bulk material, and the mean squared current is expressed as:

$$\left\langle i_f^2 \right\rangle \propto \frac{1}{f}. \tag{35.21}$$

Equation 35.21 indicates that flicker noise is of importance at very low frequencies, less than 1 KHz.

In the noise equivalent circuit, the noise current sources appear in parallel. Assuming that the noise sources are uncorrelated and also that flicker noise is negligible, one may write for the SNR as

$$\frac{S}{N} = \frac{i_{\text{ph}}^2}{\left\langle i_{\text{th}}^2 \right\rangle + \left\langle i_{\text{GR}}^2 \right\rangle}. \tag{35.22}$$

Let us assume that the optical signal incident on the detector is sinusoidally modulated by a modulating signal of angular frequency ω_{m}, and its power variation has the form $P_1 \exp(j\omega_{\text{m}}t)$. The rms optical power is $P_1/\sqrt{2}$. The rms photocurrent is expressed as

$$i_{\text{ph}} = \frac{eP_1\eta}{\sqrt{2}\hbar\omega}\frac{\tau}{\tau_{\text{tr}}}\frac{1}{\left(1+\omega_{\text{m}}^2\tau^2\right)^{1/2}}. \tag{35.23}$$

In arriving at Equation 35.23, the expression for responsivity, Equation 35.9, has been used to relate photocurrent with power and the gain is replaced by the ratio τ/τ_{tr}. The SNR as given by Equation 35.22 may now be written by using Equations 35.21 through 35.23 as

$$\frac{S}{N} = \frac{\eta P_1^2}{8B\hbar\omega P_0}\frac{1}{1+\frac{k_BT}{eG}\left(1+\omega_{\text{m}}^2\tau^2\right)\frac{1}{RI_0}}. \tag{35.24}$$

35.5.2 p-n Photodiode

The simplest PD without internal gain is a reverse-biased p-n junction, shown schematically in Figure 35.7 along with the electric field profile. Both depletion and diffusion regions are shown. Photons are absorbed in the absorption region ($\sim 1/\alpha(\lambda)$). In case of weak absorption, the absorption region may cover the whole device, a part of which may contain the depletion region. Note that EHPs are generated in both the depletion and diffusion regions. While the electric field in the depletion region sweeps the photogenerated carriers quickly, the EHPs created in the diffusion region move slowly and thus limit the response of the PD. To improve the response, it is to be ensured that all photons are absorbed in the depletion region. Therefore, its width is made as large as possible by using lower doping.

35.5.3 p-i-n Photodiode

An efficient way to overcome the limitations of a p-n PD is to use a p-i-n configuration. Near the absorption edge, the weak absorption coefficient makes the absorption region quite long. The depletion region must match this, and to achieve this, the n region is low-doped to make it intrinsic (i). To make a low-resistance contact, a heavily doped n layer (n^+) is added. The resulting structure, a p-i-n structure, is shown in Figure 35.8 including the electric field profile, depletion and absorption regions. As shown, in this structure, the EHPs created are subject to the depletion layer field.

The speed of response of the p-i-n photodiode depends on (1) the transit time of carriers across the depletion layer as given by Equation 35.11 and (2) the RC time constant, where C is the junction capacitance.

The junction capacitance may be used either by using a mesa structure or by using a planar geometry with selective diffusion in the contact region (Bandyopadhyay and Deen 2001). The capacitance may also be reduced by increasing the width of the i layer. This increases the absorption and hence the QE, but increases the transit time, thereby decreasing the speed of response. To have high values of both QE and

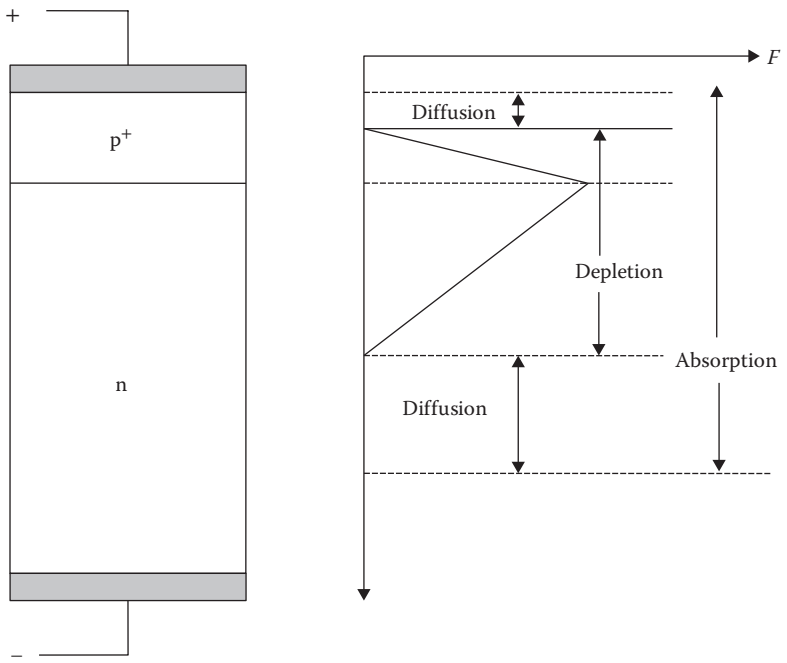

FIGURE 35.7 p-n PD structure (left) and electric field profile across the depth of the device (right). The depletion, diffusion, and absorption regions are shown.

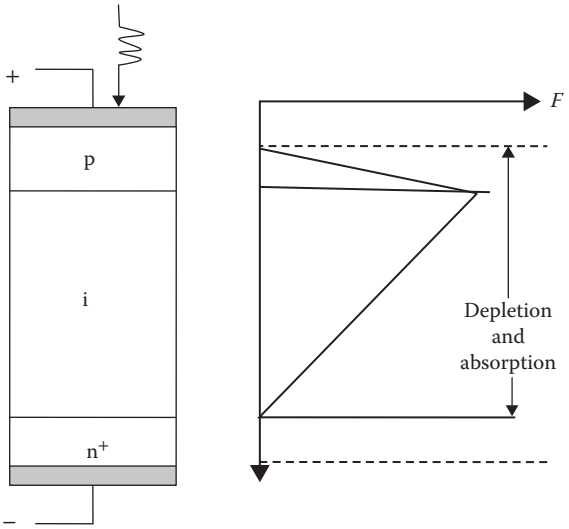

FIGURE 35.8 The p-i-n photodiode structure (left) and absorption and depletion regions and electric field (F) profile in the device (right).

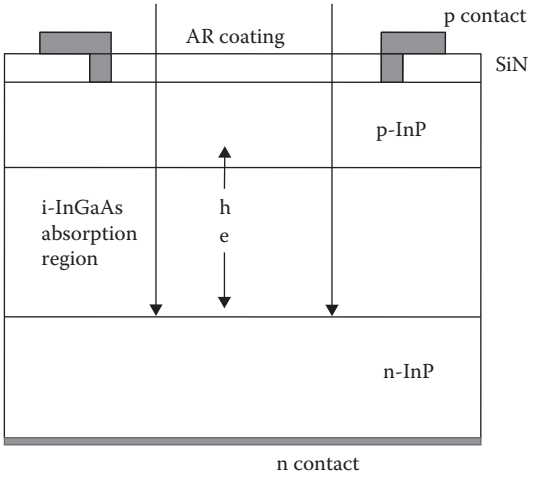

FIGURE 35.9 A cross-sectional view of a vertically illuminated p-i-n photodiode using p-InP/i-InGaAs/n-InP and SiN as antireflection (AR) coating.

speed (or bandwidth), the usual methods are (1) to employ an edge-illuminated structure or (2) to use a resonant-cavity-enhanced (RCE) structure.

A typical structure for a vertically illuminated p-i-n PD is shown in Figure 35.9. The n and p layers may be InP and the i layer is an undoped (or unintentionally doped) InGaAs layer. Since InP is transparent to 1.3–1.55 μm, the incident light is absorbed in the InGaAs layer, in which the electric field appears. Thus, there is no recombination or diffusion of the carriers. This leads to a fast response. The area of the absorbing layer must be optimized for giving high-speed operation (Bandyopadhyay and Deen 2001; Beling and Campbell 2009).

Edge-illuminated and RCE structures are presented in subsections 35.6.1 and 35.6.2, respectively.

35.5.4 Avalanche Photodiodes

APDs possess internal gain and as such are used to detect very weak optical signals. The working principle of an APD is discussed first and then some of the structures for improved performance are presented. The noise performance of APDs is discussed next.

35.5.4.1 Principles

The internal gain arises due to impact ionization in the presence of a high electric field in the depletion layer. A primary electron gets enough kinetic energy to break a covalent bond, thereby creating one secondary EHP. The primary and secondary carriers are accelerated by the depletion field and may again create tertiary EHPs. The multiplication process gives rise to an increased photocurrent. A primary hole can also initiate the ionization process. By this process, one initial electron or a hole may generate M extra EHPs, where M is the multiplication gain of the PD. In general, the ionization coefficients for electrons (α) and for holes (β) are different. The structure of an APD is shown in Figure 35.10, in which the electric field profile is also included.

In order to cause impact ionization, the carriers must possess a minimum energy, called the ionization threshold, which must exceed the bandgap energy. The ionization thresholds for electrons and holes are different. This leads to different values of ionization coefficients α and β, for electrons and holes, respectively. The ionization coefficient is the reciprocal of the average distance traversed by a carrier in the direction of the electric field to create an EHP. In its motion, a carrier may lose or gain energy via carrier–phonon interaction, which becomes more dominant with rise in temperature, thereby reducing the ionization probability. The breakdown voltage in a p-n junction therefore increases with temperature.

Defects in the device can lead to microplasma effect, in which breakdown at a lower voltage than in the remaining part of the junction occurs in a small area. This needs careful device processing and fabrication.

One of the important performance parameters for a PD used in optical communication is the maximum bit rate of operation. The detector must convert faithfully the incident optical pulses into electrical pulses. This is governed by how quickly the detector current responds to a pulse of photons. At low bias, carriers may be generated in the undepleted region and the response of an APD to an optical pulse may produce

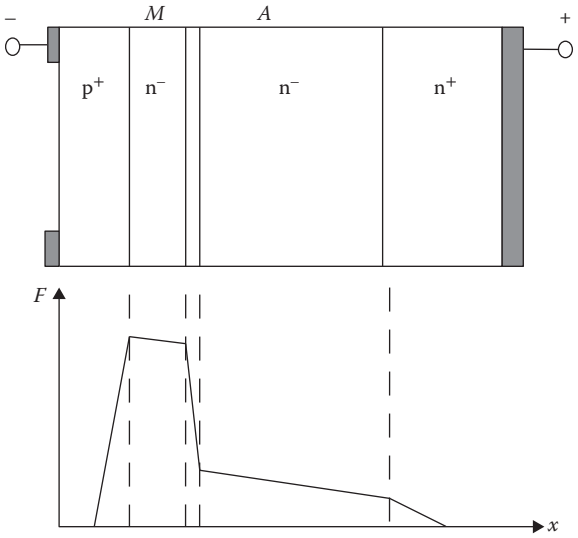

FIGURE 35.10 Structure of an APD (top) and the electric field (F) profile within it (bottom). M and A denote the multiplication and absorption regions, respectively.

a diffusion tail in the current pulse as in a p-i-n diode. When the bias is high enough to fully deplete the junction, the following three factors contribute to the response time:

1. Transit time of the carriers across the absorption layer of width W, given by Equation 35.11
2. The time taken by the carriers to create EHPs by avalanche multiplication
3. The RC time constant, where R is the device and load resistance and C is the junction capacitance

For low gain, the response time is determined by the transit time and the RC time constant. These two limit the value of bandwidth. When the gain is large, avalanche multiplication time becomes dominant and the bandwidth is reduced. APDs are therefore characterized by a constant gain bandwidth product.

35.5.4.2 Separate Absorption Grading and Multiplication and Separate Absorption, Grading, Charge Sheet, and Multiplication APDs

PDs for optical communication at a wavelength of 1.55 μm have $In_{0.53}Ga_{0.47}As$ as the absorbing layer. The usual p-n junction shows a high dark current due to the low bandgap of InGaAs. Furthermore, at high reverse field in the depletion region, the breakdown may be due to tunneling as opposed to the avalanche process. Therefore, the p-n junction is made using InP. The design involves separate absorption and multiplication (SAM). In an SAM structure, a sufficiently high electric field exists to cause avalanche multiplication in InP, and at the same time the field is low enough to prevent tunneling in the lower bandgap InGaAs layer in which optical absorption takes place. However, the photogenerated holes in the InGaAs layer while flowing toward InP layers find a band offset ΔE_v at the InGaAs/InP interface. Holes are thus trapped at the heterointerface, giving rise to a slower response. Incorporation of one or more layers of intermediate gap InGaAsP in between the absorption and multiplication regions eliminates this problem. The structure is called a separate absorption grading and multiplication (SAGM) structure.

The presence of quaternary and ternary layers in SAGM structures leads to another problem. The ionization rate of holes is higher than that of electrons in InP, but is lower in both the InGaAs and InGaAsP layers. This may give rise to unwanted multiplication in the ternary and quaternary layers. As the ionization ratio k approaches unity there, higher noise and lower speed are expected. The doping and thickness of the multiplication layer should be accurately controlled in order to have the increased gain–bandwidth product. A planar SAGM APD that avoids premature edge breakdown employs a partial charge sheet layer in between the multiplication layer and the grading region. The structure is known as a separate absorption, grading, charge sheet, and multiplication (SAGCM) structure. The structure consists of a lightly (or unintentionally) doped wide gap multiplication region where the field is high and an adjacent doped charge layer or a field control region. The presence of the charge (C) layer offers an extra degree of freedom such that the electric field profiles in the active region and in the periphery of the device can be controlled independently.

A cross-sectional view of an SAGCM APD using InP multiplication, InP charge, InGaAsP grading, and InGaAs absorption layer is shown in Figure 35.11. APDs used in telecommunications are reviewed by Campbell (2007).

35.5.4.3 Noise in APDs

Three important noise sources are operational in p-i-n PDs and in APDs. They are (1) thermal noise, (2) dark current noise, and (3) shot noise. In addition, APDs are characterized by an excess noise due to the random nature of the avalanche multiplication process. The statistical variation of multiplication is responsible for multiplication excess noise. The excess noise factor F depends on the ratio $k = \alpha/\beta$ and is expressed as

$$F = kM + (1-k)\left(2 - \frac{1}{M}\right) = M\left\{1 - \left[(1-k)\left(\frac{M-1}{M}\right)^2\right]\right\}. \tag{35.25}$$

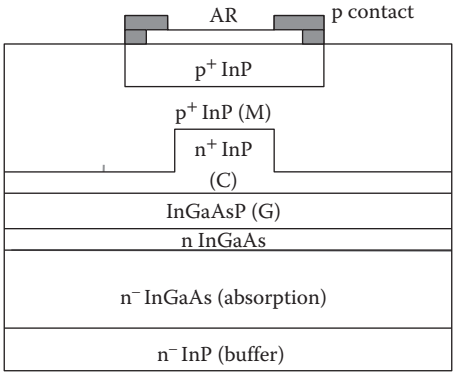

FIGURE 35.11 Cross-sectional view of an SAGCM APD. Antireflection (AR), multiplication (M), charge sheet (C), and graded (G) regions are shown.

The excess noise factor therefore reduces with large value of k. The total dark current in an APD is given by

$$I_d = I_s + I_{dm}M, \tag{35.26}$$

where I_s is the unmultiplied surface current, and I_{dm} is the unmultiplied dark current. Let I_{ph0} be the primary DC photocurrent, I_{bk} the photo-induced background current, and B the bandwidth. The mean square shot noise current is then

$$\langle I_M \rangle^2 = 2e\left[I_s + \left(I_{dm} + I_{bk} + I_{ph0}\right) M^2 F\right] B. \tag{35.27}$$

The SNR of APD is written as

$$\text{SNR} = \frac{M^2 I_{ph0}^2}{\langle I_M \rangle^2 + \left(4k_B TB/R\right)} = \frac{M^2 I_{ph0}^2}{2e\left[I_s + \left(I_{dm} + I_{bk} + I_{ph0}\right) M^2 F\right] B + \left(4k_B TB/R\right)}. \tag{35.28}$$

In the above expression, the thermal noise contribution $\left(4k_B TB/R\right)$ has been included. Note that by putting $M = 1$ and $F = 1$, we may obtain the SNR for p-i-n PD.

It is easy to assess the relative contributions of the shot noise (first term in the denominator of Equation 35.28) and of the thermal noise term (second term in the denominator), by dividing the numerator and denominator by M^2. The shot noise increases with M as does F. However, the thermal noise decreases with M as M^{-2}. It appears therefore that there exists an optimum value of M for which SNR can be a maximum. The variation of M is effected by change of reverse bias and may be expressed by the following empirical expression:

$$M \approx \frac{1}{\left\{1 - \left(\frac{V_a - I_{ph0}R}{V_B}\right)^n\right\}}, \tag{35.29}$$

where V_a is the bias voltage, R the effective resistance, V_B the breakdown voltage, and n a constant that depends on the design of the device.

35.5.5 Schottky Barrier PD

In a Schottky barrier (SB) PD, the rectifying junction is formed between a metal and a semiconductor, rather than by a p-n junction. A schematic structure of an SB PD is shown in Figure 35.12, which indicates that the active or absorbing layer is quite narrow. This leads to short transit time and hence higher bandwidth, but to a lower QE.

35.5.6 Metal-Semiconductor-Metal PD

A metal-semiconductor-metal (MSM) PD is shown schematically in Figure 35.13. The schematic layer structure consists of a thin undoped layer grown on a semi-insulating substrate. Metal electrodes are then deposited on the active layer as interdigitated fingers. Each set of fingers forms a Schottky contact with the semiconductor and is connected to a large pad for connection to the external circuit. Both contacts in the MSM PD are on the same side of the substrate. Since the active layer thickness is small, the QE is less than in a p-n PD. Moreover, the reflections from the metal and semiconductor surfaces lead to smaller responsivity. However, the small transit time and very low capacitance greatly enhance the bandwidth.

A contact between a metal and a semiconductor may be either ohmic or rectifying. To obtain a rectifying contact, a certain combination of metal and a given semiconductor is needed. In such a junction, potential barrier develops due to the difference between the work functions of the metal and semiconductor. Another difference between a p-n junction and an SB diode is that both electrons and holes contribute to the current in a p-n junction, whereas SB diodes are majority carrier devices. The MSM PD consists of two junctions,

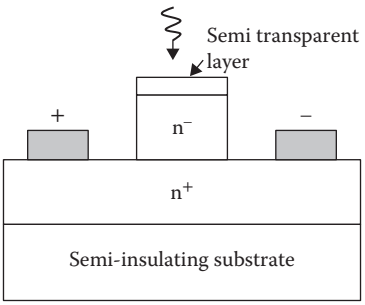

FIGURE 35.12 Schematic diagram of an SB PD. Shaded regions represent metallic contacts.

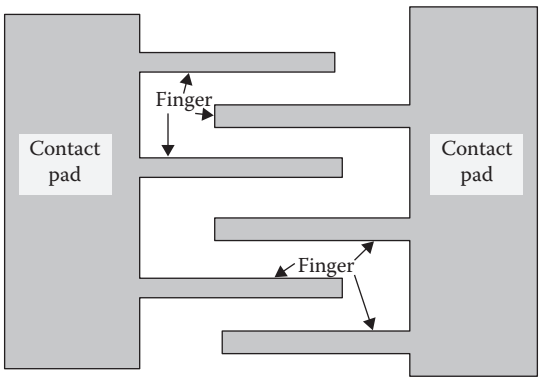

FIGURE 35.13 Top view of an MSM PD showing fingers and contact pads for connection to external circuits.

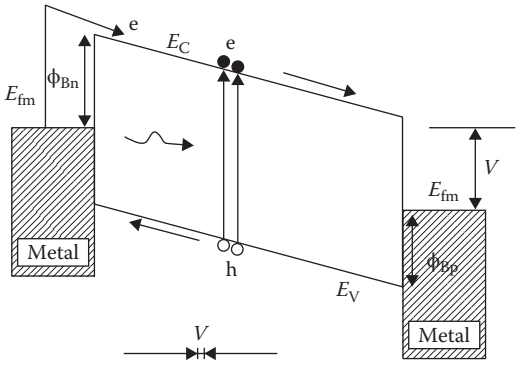

FIGURE 35.14 Band diagram of an MSM PD showing directions of flow of electrons and holes. E_c and E_v are conduction and valence band edges, respectively, E_{fm} is the Fermi level of the metal, ϕ_{Bn} and ϕ_{Bp} are the barrier heights at the n and p sides, respectively, and V is the applied bias between two metallic contacts. The PD is represented by two rectifying contacts shown at the bottom.

both of which may be rectifying or one junction ohmic and the other rectifying. The band diagram of an MSM PD is shown in Figure 35.14, in which the potential barrier ϕ_b is due to the work function difference. The two electrodes are connected serially, back-to-back, and the device contains two diodes as shown in Figure 35.14.

The dark current in the PD is due to the flow of thermally generated EHPs and by carriers from the metal surmounting the potential barrier ϕ_b. Under illumination, excess EHPs are generated and flow of electrons and holes is in opposite directions under the influence of the external bias V, as shown in Figure 35.14.

As stated already in connection with SB PDs, MSM PDs make a compromise between QE and bandwidth. The use of multiple fingers increases the QE. However, to increase the bandwidth, the effective absorption layer thickness needs to be reduced. Introduction of a highly doped layer at a certain depth may restrict the electric field within a certain absorption region and may provide a solution to the bandwidth problem.

Current improved technology may produce very thin fingers with narrow spacings between fingers. This reduces the transit time, but the speed of response is governed by the RC time constant. The following simple analytical expression for the capacitance may be used to calculate the time constant (Averine et al. 2001):

$$C = 0.226NL\varepsilon_0\left(\varepsilon_S + 1\right)\left(6.5\theta^2 + 1.08\theta + 2.37\right). \tag{35.30}$$

In Equation 35.30, L is the length of a finger, ε_0 is the permittivity of free space, ε_S is the relative permittivity of the semiconductor, and $\theta = (D + t)/t$ is the finger-to-period ratio, where D is the width of the metal finger and t is the gap spacing between the metal fingers.

The EQE is given by

$$\eta = (1 - R)\frac{t}{t + D}\left[1 - \exp\left(-\alpha d\right)\right], \tag{35.31}$$

where R is the reflection coefficient, α the absorption coefficient, and d the thickness of the absorption layer of the semiconductor. The QE may be increased by replacing the metal electrodes by transparent conducting indium tin oxide (ITO), thereby reducing R. It can be increased by reducing the finger width at the cost of increased resistance however.

35.5.7 Phototransistors

Phototransistors may be a bipolar junction transistor (BJT) or metal-oxide-semiconductor field-effect transistor (MOSFET). The BJT may be either p-n-p or n-p-n type. It may be used in two-terminal (2T) configuration, with electrical contacts to the emitter and collector with the base floating, or in three-terminal (3T) configuration as in a normal BJT.

We first consider the 2T configuration. The structure is akin to a heterojunction bipolar transistor in which the emitter region is a higher bandgap material. The emitter is transparent to the incident light and thus acts as a window layer; the incident light falls directly on the base and is primarily absorbed in the base, base-collector depletion region, and the collector region. The heterojunction also increases the current gain. The base-collector junction is reverse biased so that the transistor acts as a reverse-biased PD.

Under no illumination the usual collector saturation current flows. However, when light is incident, the photogenerated electrons cross the depletion region in the base-collector junction and are collected at the collector terminal. The photogenerated holes travel in the opposite direction to the emitter terminal, thus reducing the emitter injection efficiency and the current gain.

The incident light is absorbed in the base and the base-collector depletion layer. The collector current for the (2T) configuration is given by (Chand et al. 1985):

$$I_c = I_{ph}(1 + h_{fe}), \tag{35.32}$$

where I_{ph} is the photogenerated current and h_{fe} the common emitter current gain of the BJT. A dark current also contributes to the collector current.

The currently used phototransistors are heterojunction bipolar transistors with larger gap and smaller gap semiconductors forming the emitter and base layers, respectively [see Basu et al. (2015) and Chang et al. (2016)] for work on GeSn-based heterojunction phototransistors and references therein for work on III-V compounds and alloy-based devices).

As stated already, a p-i-n diode is formed by the base, collector, and subcollector of a heterojunction phototransistor (HPT). Under optical illumination in a 2T-HPT, the photocurrent generated within the p-i-n diode is injected across the emitter-base junction and then amplified through transistor action. In this way, both the photocurrent in the p-i-n structure and the dc current gain determine the final collector photocurrent and the optical gain. When the incident optical power is low, the dc current gain is low, thereby making the optical gain low. This situation is improved in a 3T-HPT, in which a dc bias is applied to the base. The bias shifts the operating point to a higher collector current, making the dc current gain and optical gain higher.

A detailed analysis of the 3T-HPT has been presented by Frimel and Roenker (1997a,b) using a thermionic emission-diffusion model. The theory is too lengthy to be included here.

Useful references for 3T-HPTs for III–V compound-based materials as well as recently developed GeSn alloys may be found in Basu et al. (2015) and Chang et al. (2016).

In an MOS phototransistor, light incident at the source-gate region changes the surface potential under the gate oxide. This change causes a change in the current flowing through the channel.

35.5.8 QWIP and QDIP

It has been mentioned in Section 35.4 that the absorbing layer in a PD may contain quantum nanostructures. QWs and MQWs, in particular, are incorporated in the intrinsic (i) layer of a p-i-n PD, an APD, or in the base of a hetero phototransistor. The absorption in MQWs is from a valence subband to a conduction subband.

Quantum well infrared photodetectors (QWIPs) have QWs as the absorbing region, but the transition occurs from one subband to another in the same band, as shown in Figure 35.11. The intersubband

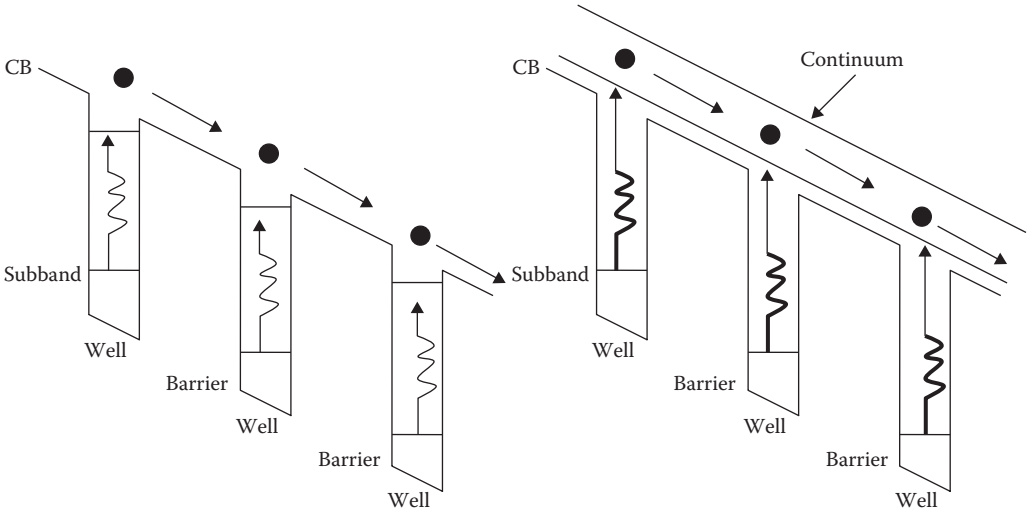

FIGURE 35.15 Intersubband absorption in QWIPs: (a) a bound-to-bound transition and (b) a bound-to-continuum absorption. The movement of electrons (solid balls) is shown in the figures.

transition energy is lower than the energy involved in band-to-band transitions in a QW. The QWIPs therefore work at the mid-infrared range (3–20 μm). Equation 35.4 may be used to calculate the energy difference between two subbands corresponding to the wavelength to be detected (see Equation 35.6). The width of the QW is then calculated from the value of the energy difference.

The types of transitions involved for CB are illustrated in Figure 35.15. The MQWs form the intrinsic layer of a p-i-n structure, in which the reverse bias causes the band bending. As shown in Figure 35.15a, the transition takes place from the lowest subband to the upper subband that lies close to the CB edge in the barrier layer. The electron lifted to the upper subband tunnels to the continuum states in the barrier and is drifted by the built-in electric field. In Figure 35.15b, an electron from the lone lowest subband absorbs the MIR photon and goes to the continuum states in the barrier and is then transported (see Schneider and Liu 2007).

Instead of QWs, multiple stacks of QDs can act as the absorbing layer, and the PD is called a quantum dot infrared photodetector (QDIP) (Barve et al. 2010).

35.5.9 Single-Photon Detectors

Single-photon emitters and detectors are needed for quantum information processing systems. Semiconductor-based single-photon detectors are APDs that operate with reverse bias exceeding the breakdown voltage. In this regime, the gain is enormously high. An incoming photon creates a charge, the number of which is multiplied (avalanche process) until it saturates at a current typically limited by an external circuit. In order that the single-photon APD should respond to a subsequent optical pulse, the saturated avalanche current must be terminated by lowering the bias voltage below the breakdown voltage (Eisaman et al. 2011).

35.6 Improved Devices

The PD structures such as p-i-n, APDs, etc., may show improved performances if some structural or materials modifications are introduced, some of which are discussed briefly in this section.

35.6.1 Waveguide PDs

In all the earlier discussions, vertically illuminated PD structures have been considered. For good PD performance, both the QE and speed of operation must be as high as possible. The absorbing layer in vertically illuminated devices must be very thin in order to reduce transit time; however, the thin layer cannot absorb the incident photons, thereby reducing the QE. To achieve high QE and high speed, edge-illuminating devices are used (Beling and Campbell 2014). The advantage of using waveguide structure has been pointed out in Section 35.4. The waveguide confines the otherwise spreading light into the high-field region. (Beling and Campbell 2014).

The waveguide photodetector (WPD) structure employs waveguides with absorbing layers either embedded in the waveguide or grown adjacent to the waveguide. The edge-coupled structure has been realized in p-i-n, APD, MSM, or uni traveling wave configuration.

The simplest type of edge-illuminated WPD structure is shown in Figure 35.16. The incident light falls directly into the edge of the absorbing layer. Here the direction of propagation of light is normal to the direction of transit of the photogenerated carriers. Note that the incident light may spread outside the absorbing region, making the optical confinement factor low and effective absorption less than in the bulk. However, the absorption length is quite large (~mm) so that responsivity and QE are enhanced. At the same time, the photogenerated carriers traverse a very short distance equal to the thickness of the absorbing layer making the transit time low. The increase in the capacitance due to the thin depletion layer may be arrested by reducing the area of the device. The state-of-the art technology for WPDs is discussed by Beling and Campbell (2014).

One drawback of the edge-coupled WPD is the misalignment of the optical field with the waveguide region. Use of lensed fiber or additional optics can increase the coupling efficiency. A useful way to improve coupling efficiency or to avoid the misalignment problem is to employ the evanescent waveguide structure as shown in Figure 35.17. In this configuration, the absorbing layer is placed adjacent to the waveguide core. The optical coupling is enhanced by inserting a transition layer between the waveguide and the detector. Since the waveguide mode has a smaller cross-section than that of a single-mode fiber, the input coupling efficiency is increased by using different approaches, for example, using a spot-size converter.

35.6.2 RCE Structure

The performance of a PD can be enhanced by placing the active device structure inside a Fabry–Pérot cavity. The cavity acts as a resonant structure providing wavelength selectivity and a large enhancement of the optical field, which in turn allows thinner PD layer. The device shows faster response and increased QE at the resonant wavelength. The wavelength selectivity and high-speed response make the devices the right candidates for wavelength division multiplexing applications (Unlu et al. 2001).

Photonic crystal structures have been employed to realize nanoscale PDs offering ultralow capacitance, thereby increasing the bandwidth and enabling ultralow power operation (Nozaki et al. 2013) (Figure 35.16).

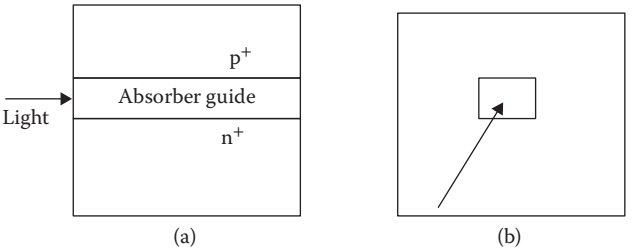

FIGURE 35.16 Schematic view of a WPD: (a) side view and (b) front view.

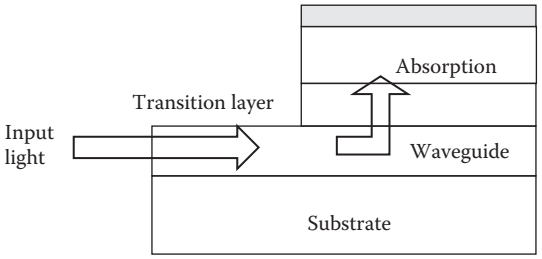

FIGURE 35.17 Schematic of an evanescently coupled WPD. Light from the waveguides couples into the absorption region via the transition layer.

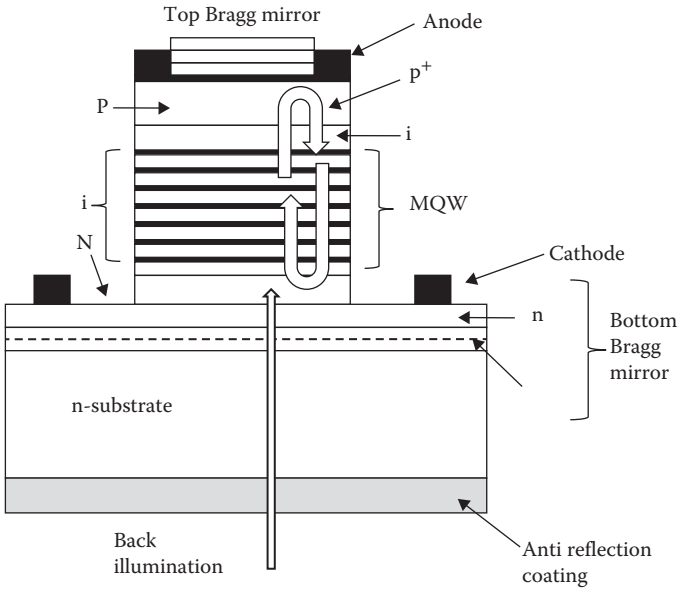

FIGURE 35.18 Schematic diagram of a RCE PD. MQW, multiple quantum well.

35.6.3 Incorporation of Heterojunctions and Quantum Nanostructures

The active materials for detection are usually grown on a suitable substrate leading to heterojunctions. However, heterojunctions and MQWs and stacks of QDs are also used in the active regions of PDs. The heterostructures extend or control the range of wavelength to be detected. For example, in Ge_xSi_{1-x}/Si heterostructures or MQWs, the absorption takes place in the lower gap alloy material, and the heterobarrier prevents leakage of carriers, thus reducing the dark current.

35.7 Materials

The wavelength range covered by a PD depends on the active material, the bandgap of which determines the maximum wavelength (cutoff wavelength). Table 35.1 includes a few important materials and the corresponding wavelength range.

TABLE 35.1 Photodetector Materials and the Wavelength Range Covered

Material	Bandgap (eV)	Wavelength Range (μm)	Substrate
GaAs	1.42	<0.87	GaAs
$In_{0.53}Ga_{0.47}As$	0.81	1.3–1.6	InP
Si	1.12	0.2–1.12	Si
Ge	0.66	<1.8	Si
$Si_{1-x}Ge_x$	1.12–0.66	<1.8	Si
$Si_{1-x}Ge_x$/Si QWs	1.12–0.66	2–12	Si
$Ge_{1-x}Sn_x$ ($x > 0.08$)	0.58– ~ 0	1.2–2.2	Si
GaAlAsSb/GaSb	0.60	0.75–1.9	GaSb
InGaAlAs/InGaAs/InP	–	0.5–2.1	InP
HgCdTe/CdTe	–	0.8–2.0	CdTe
$Ga_{1-x}Al_xN$	0.38–0.22	0.36–<0.2	Sapphire

Acknowledgments

The author is thankful to Dr. Rikmantra Basu of NIT Delhi, and Dr. Bratati Mukhopadhyay and Vedatrayee Chakraborty of Institute of Radio Physics and Electronics for their help in preparation of the manuscript.

References

Averine S. V., Chan Y. C., and Lam Y. L. (2001) Geometry optimization of interdigitated Schottky-barrier metal-semiconductor-metal photodiode structures. *Solid State Electron.*, vol. 45, 441–446.

Bandyopadhyay A. and Deen M. J. (2001) Photodetectors for optical fiber communications, Ch. 5 in *Photodetectors and Fiber Optics*, Nalwa H. S (ed) San Diego, CA: Academic Press

Barve A J., Lee S. J., Noh S. K., and Krishna S. (2010) Review of current progress in quantum dot infrared photodetectors. *Laser Photon. Rev.*, vol. 6, 738–750.

Basu P. K. (1997) *Theory of Optical Processes in Semiconductors: Bulk and Microstructures*. Oxford, UK: Oxford University Press.

Basu R., Chakraborty V., Mukhopadhyay B., and Basu P. K. (2015) Predicted performance of Ge/GeSn hetero-phototransistors on Si substrate at 1.55 μm. *Opt. Quant. Electron.*, vol. 47, no. 2., 387–399.

Beling A. and Campbell J. C. (2009) InP-based high-speed photodetectors. *J. Lightwave Technol*, vol. 27, no. 3, 343–355.

Beling A. and Campbell J. C. (2014) High-speed photodiodes. *IEEE J. Sel. Top. Quantum Electron*, vol. 20, no. 6, 3804507.

Bhattacharyya P. (1996) *Semiconductor Optoelectronic Devices*, 2nd edn. Upper Saddle Point, NJ: Prentice-Hall.

Bowers J. E. and Burrus, C. A., Jr. (1987) Ultrawide-band long-wavelength p-i-n photodetectors. *J. Lightwave Technol.*, vol. 5, no. 10, 1339–1350.

Bowers J. E. and Wey Y. G. (1995) High speed photodetectors, Ch. 17 in *Handbook of Optics, Fundamentals, Techniques and Design*, Vol. 1, Bass M (ed.). Pennsylvania, NY: McGraw-Hill.

Campbell J. C. (2007) Recent advances in telecommunications avalanche photodiodes. *J. Lightwave Technol*, vol. 25, no. 1, 109–121.

Chand N., Houston P. A., and Robson P. N. (1985) Gain of a heterojunction bipolar phototransistor. *IEEE Trans. Electron Dev*, vol. 32, no. 3, 622–627.

Chang G. E., Basu Rikmantra, Mukhopadhyay B. and Basu P. K. (2016) Design and modeling of GeSn-based heterojunction phototransistors for communication applications. *IEEE J. Sel. Topi. Quantum Electronics*, vol. 22, no. 6, 1–9.

Deen M. J. and Basu P. K. (2012) *Silicon Photonics: Fundamentals and Applications*, Chichester: John Wiley.

Eisaman M. D., Fan J., Migdall A., and Polyakov S. V. (2011) Single-photon sources and detectors. *Rev. Sci. Instruments*, vol. 82, 071101.

Frimel S. M. and Roenker K. P. (1997a) A thermionic-field-diffusion model for Npn bipolar heterojunction phototransistors. *J. Appl. Phys.*, vol. 82, 1427–1437.

Frimel S. M. and Roenker K. P. (1997b) Gummel-Poon model for Npn heterojunction bipolar phototransistor. *J. Appl. Phys.*, vol. 82, 3581–3592.

Kaiser G. (2013) *Optical Fiber Communication*, 5th edn. New Delhi; McGraw Hill Education (India) Pvt. Ltd.

Kumar S. and Deen M. J. (2014) *Fiber Optic Communications: Fundamentals and Applications*. Hoboken, NJ: Wiley.

Nozaki K., Matsuo S., Takeda K., Sato T., Kuramochi E., and Notomi M. (2013) InGaAs nanophotodetectors based on photonic crystal waveguide including ultracompact buried heterostructure. *Opt. Express*, vol. 21, no. 16, 19022–19028.

Schneider H. and Liu H. C. (2007) *Quantum Well Infrared Photodetectors: Physics and Applications*. Berlin: Spriner Verlag.

Senior M. and Jamro M. Y. (2009) *Optical Fiber Communications: Principles and Practice*. Harlow, England: Pearson Prentice-Hall.

Streetman B. G. and Banerjee S.K. *Solid State Electronic Devices*, 5th edn. Harlow, England: Pearson Education Ltd.

Unlu M. S., Ulu G., and Gokkavas M. (2001) Resonant cavity enhanced photodetectors, Ch. 2 in *Photodetectors and Fiber Optics*, Nalwa H. S (ed) San Diego, CA: Academic Press

36

P-N Junction Photodiodes

Weida Hu

36.1 Introduction

Semiconductor p-n junction photodiodes have been available for half a century. By applying an electric field across the p-n junction, the light-generated carriers create an electrical current flow, thus the p-n junction photodiodes can convert optical signals into electrical signals (Shi and Li, 2014). Sensing applications in infrared, visible, and ultraviolet spectral bands promote theoretical research of these photodiodes. Numerical simulations, containing structural details such as layer thicknesses, doping profiles, and trap concentrations, provide key insights into device design and reliability degradation mechanisms, which could effectively reduce costly and time-consuming testing when developing and characterizing a new semiconductor device or technology. However, simulations of p-n junction photodiodes involve many important physical models, which play a decisive role in the calculations of accurate results. This chapter covers some basic equations and models used in the simulations of p-n junction photodiodes. Another key point is the accurate knowledge of device material parameters. Because of the complexity of an actual device structure, different device and material parameters are needed for various designs and structures. A number of these characteristic input parameters required for accurate modeling when building a basic theoretical framework are explained in this chapter.

First, this chapter presents several basic equations and models in detail. Second, principles and characteristics of several common optical generation models are reviewed and compared. Then, a number of important characteristic input parameters are introduced. Several p-n junction photodiode simulation examples are listed in the end of this chapter.

Figure 36.1 presents the general process simulation of p-n junction. After building the device structure, the device model could be defined by adding mesh generation and doping for different region. Different physical models are added in the optical/electrical simulation step, and then the desired information could be achieved. Figure 36.2 examples the results of some of the important characteristics of p-n junction.

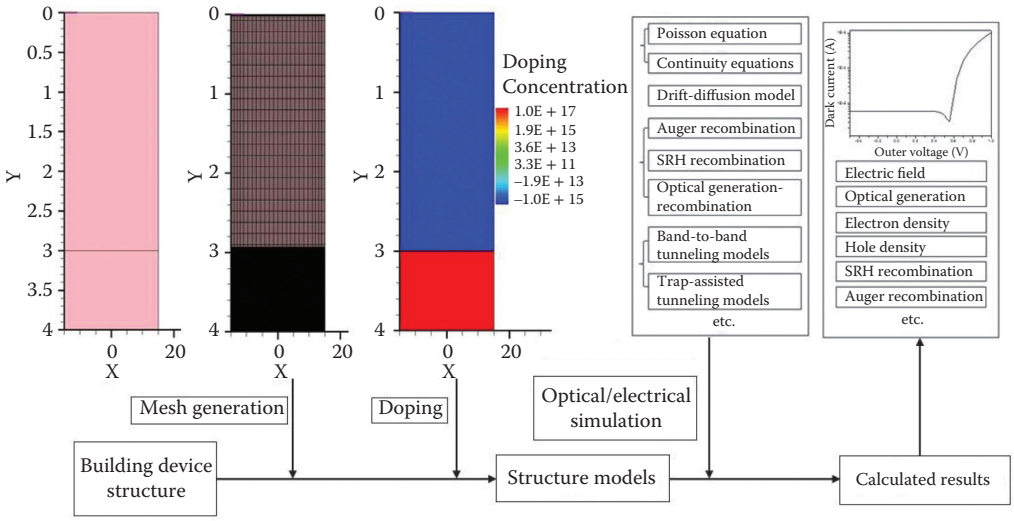

FIGURE 36.1 The flow chart of numerical simulations of p-n junction photodiodes.

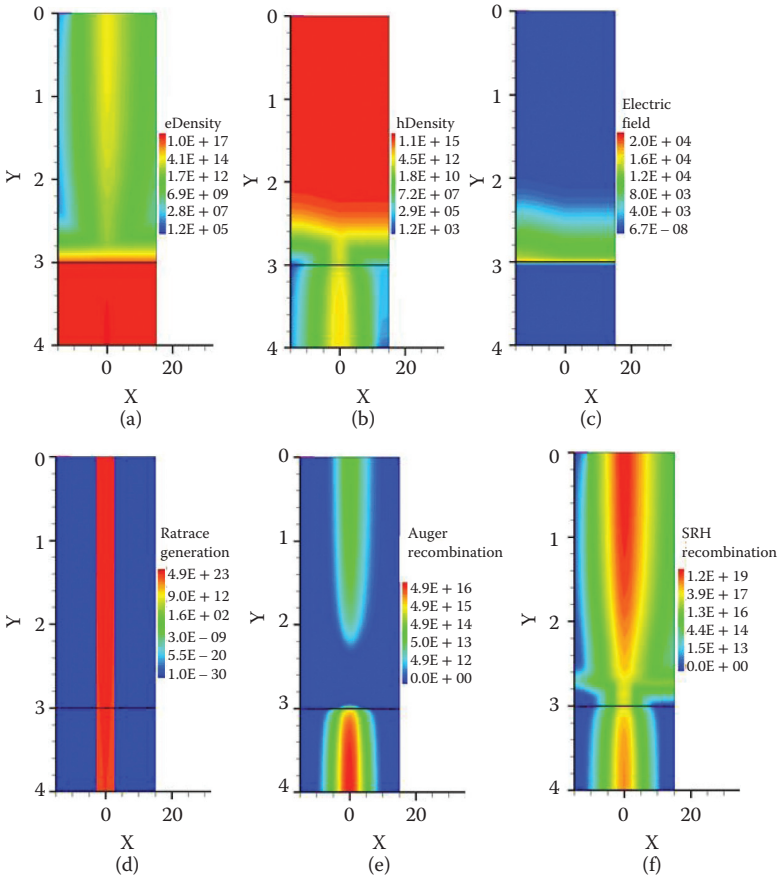

FIGURE 36.2 Important output/plots for a two-dimensional numerical simulation: (a) electron density (eDensity), (b) hole density (hDensity), (c) electric field, (d) raytrace generation, (e) Auger recombination, and (f) Shockley–Read–Hall (SRH) recombination.

36.2 Basic Equations and Models

Physical phenomena in semiconductor devices are very complicated and, for different devices and materials, models of different levels of complexity are needed in simulations. Different partial differential equations and models determine different microscopic physical mechanisms in a simulation, and can depend on different materials and device structures. This chapter introduces a number of widely used basic equations and models.

Depending on the p-n junction photodiodes under investigation and the level of modeling accuracy required, there are three governing equations for charge transport in semiconductor devices: *the Poisson equation* and *the electron and hole continuity equations*. The Poisson equation is as follows:

$$\nabla\varepsilon \cdot \nabla\psi = -q\left(p - n + N_{D^+} - N_{A^-}\right) \tag{36.1}$$

where ε is the electrical permittivity, ψ is the electrostatic potential, q is the elementary electronic charge, n and p are the electron and hole densities, N_{D^+} is the number of ionized donors, and N_{A^-} is the number of ionized acceptors. The electron and hole continuity equations are written as:

$$\frac{1}{q}\nabla\overrightarrow{J_n} + \left(G^{\text{opt}} - R\right) = 0 \tag{36.2}$$

$$\frac{1}{q}\nabla\overrightarrow{J_p} - \left(G^{\text{opt}} - R\right) = 0 \tag{36.3}$$

where R is the net electron–hole recombination rate, $\overrightarrow{J_n}$ the electron current density, and $\overrightarrow{J_p}$ is the hole current density.

Otherwise, *the drift-diffusion model* is widely used for the simulation of carrier transport in p-n junction photodiodes and is defined by the basic semiconductor equations, where current densities for electrons and holes are given by

$$\overrightarrow{J_n} = -nq\mu_n\nabla\phi_n \tag{36.4}$$

$$\overrightarrow{J_p} = -pq\mu_p\nabla\phi_p \tag{36.5}$$

where μ_n and μ_p are the electron and hole mobilities, ϕ_n and ϕ_p are the electron and hole quasi-Fermi potentials, respectively.

Additionally, *the thermodynamic model*, which is defined by a basic set of partial differential equations, and the lattice heat flow equations have been used in the simulations. The relations are generalized to include the temperature gradient as a driving term:

$$\overrightarrow{J_n} = -nq\mu_n\left(\nabla\phi_n + P_n\nabla T\right) \tag{36.6}$$

$$\overrightarrow{J_p} = -pq\mu_p\left(\nabla\phi_p + P_p\nabla T\right) \tag{36.7}$$

where P_n and P_p are the absolute thermoelectric powers.

36.3 Optical Generation Models

In the numerical simulation of the photoelectric detector, different structures and devices have different requirements for optical absorption models, including photoelectric incident position, power density, incident angle, reflection, refraction, and so on. Therefore, it is very important to select suitable optical

TABLE 36.1 Comparison of Some Common Optical Generation Models

	Optical Beam	Raytracing	Transfer Matrix	FDTD	Beam Propagation
Layered media	×	√	√	√	√
Anisotropy on horizontal direction	×	√	×	√	√
Geometrical optics	×	√	×	√	√
Diffraction	×	×	×	√	×
Reflection	×	√	√	√	√
Oblique incidence light	×	√	×	√	√

Note: FDTD = finite-difference time-domain.

generation models for different devices. The application conditions of some common optical generation models are listed in Table 36.1, and the characteristics of each model are introduced in this chapter.

Optical beam method is a simple optical absorption simulation method using Beer's law, and is only applicable to very simple optical process. Raytracing is capable of simulating a wide variety of optical effects such as reflection, refraction, and scattering, but its accuracy is severely limited, and cannot handle complex optical problems such as diffraction. Transfer matrix is very suitable for layered media, but requires media to be transversely isotropic. The FDTD method belongs in the general class of grid-based differential numerical modeling methods (finite difference methods). The time-dependent Maxwell's equations (in partial differential form) are discretized using central-difference approximations to the space and time partial derivatives, and could deal with more complex optical effects such as diffraction. Beam propagation method (BPM) is a quick and easy method of solving for fields in integrated optical devices, and a higher accuracy could be achieved. However, the BPM relies on the slowly varying envelope approximation, and is inaccurate for the modeling of discretely or fastly varying structures. What's more, BPM has some problems when dealing with scattering and diffraction problems.

36.3.1 Optical Beam Absorption

The optical beam absorption method supports the simulation of photogeneration using Beer's law. Multiple vertical photon beams can be defined to represent the incident light. The following equations describe useful relations for the photogeneration problem:

$$E_{\text{ph}} = \frac{hc}{\lambda} \tag{36.8}$$

$$J_0 = \frac{P_0}{E_{\text{ph}}} \tag{36.9}$$

where J_0 denotes the optical beam intensity (number of photons that cross an area of 1 cm^2 per 1 s) incident on the semiconductor device, P_0 is the incident wave power per area (W/cm^2), λ is the wavelength (cm), h is Planck's constant (J · s), c is the speed of light in a vacuum (cm/s), and E_{ph} is the photon energy that is approximately equal to $\frac{1.24}{\lambda\,(\mu\text{m})}$ in eV.

The optical beam absorption model computes the optical generation rate along the z-axis taking into account that the absorption coefficient varies along the propagation direction of the beam according to

$$G^{\text{opt}}(z,t) = J_0 F_t(t) F_{xy} \cdot \alpha(\lambda, z) \cdot \exp\left(-\left|\int_{z_0}^{z} \alpha(\lambda, z)\,\mathrm{d}z\right|\right) \tag{36.10}$$

where t is the time, $F_t(t)$ is the beam time behavior function: for a Gaussian pulse, it is equal to 1 for t in $[t_{min}, t_{max}]$ and shows a Gaussian distribution decay outside the interval with the standard deviation σ_t; F_{xy} is equal to 1 inside the illumination window and zero otherwise; z_0 is the coordinate of the semiconductor surface; and $\alpha(\lambda, z)$ is the nonuniform absorption coefficient along the z-axis.

The optical beam absorption method supports a wide range of window shapes as well as arbitrary beam time behavior functions. It is also not limited to beams propagating along the z-axis.

36.3.2 Raytracing

The optical generation method of raytracing supports the simulation of photogeneration in two dimensions (2Ds) and three dimensions (3Ds) for arbitrarily shaped structures. The calculation of refraction, transmission, and reflection follows geometric optics. A plane wave can be partitioned and each partition is represented by a 1D ray of light. Raytracing can approximate the behavior of a plane wave on a device by following such rays.

The optical generation along a ray, when the propagation is thought to be in the z-axis for a nonuniform absorption coefficient $\alpha(\lambda, x, y, z)$, can be written as a photon beam generation:

$$G^{opt}(z,t) = J\left(x,y,z_0\right)\alpha\left(\lambda,z\right)\cdot\exp\left(-\left|\int_{z_0}^{z}\alpha\left(\lambda,z\right)\mathrm{d}z\right|\right) \tag{36.11}$$

where $J(x, y, z_0)$ is the beam spatial variation of intensity over a window where rays enter the device, and z_0 is the position along the ray where absorption begins.

When a ray passes through a material boundary, the raytracer uses a recursive algorithm: It starts with a source ray and builds a binary tree that tracks the transmission and reflection of the ray. A reflection/transmission process occurs at interfaces with refractive index differences.

An incident ray impinges on the interface of two different refractive index (n_1 and n_2) regions, resulting in a reflected ray and a transmitted ray. The angles involved in raytracing are governed by Snell's law using

$$n_1 \sin\left(\theta_1\right) = n_2 \sin\left(\theta_2\right) \tag{36.12}$$

Absorption of photons occurs when there is an imaginary component (extinction coefficient), κ, to the complex refractive index. To convert the absorption coefficient to the necessary units, the following formula is used for power/intensity absorption:

$$\alpha\left(\lambda\right)\left[\mathrm{cm}^{-1}\right] = \frac{4\pi\kappa}{\lambda} \tag{36.13}$$

In the complex refractive index model, the refractive index is defined element wise. In each element, the intensity of the ray is reduced by an exponential factor defined by $\exp(-\alpha L)$ where L is the length of the ray in the element. Therefore, the photon absorption rate in each element is

$$G^{opt}\left(x,y,z,t\right) = I\left(x,y,z\right)\left(1-e^{(-\alpha L)}\right) \tag{36.14}$$

where $I(x, y, z)$ is the rate intensity of the ray in the element. After all of the photon absorptions in the elements have been computed, the values are interpolated onto the neighboring vertices and are divided by its sustaining volume to obtain the final absorption rate. The absorption of photons occurs in all materials with a positive extinction coefficient for raytracing. Depending on the quantum yield, a fraction of this value is added to the carrier continuity equation as a generation rate so that correct accounting of particles is maintained.

36.3.3 Transfer Matrix Method

The propagation of plane waves through layered media could be calculated by using a transfer matrix approach.

In the underlying model of the optical carrier generation rate, monochromatic plane waves with arbitrary angles of incidence and polarization states penetrating a number of planar, parallel layers are assumed. Each layer must be homogeneous, isotropic, and optically linear. In this case, the amplitudes of forward and backward running waves $A_j^{\pm}$ and $B_j^{\pm}$ in each layer shown in Figure 36.3 are calculated with the help of transfer matrices.

These matrices are functions of the complex wave impedances Z_j given by $Z_j = n_j \cdot \cos\theta_j$ in the case of E polarization (TE) and by $Z_j = n_j/\cos\theta_j$ in the case of H polarization (TM). Here, n_j denotes the complex index of refraction and θ_j is the complex counterpart of the angle of refraction ($n_0 \cdot \sin\theta_0 = n_j \cdot \sin\theta_j$).

The transfer matrix of the interface between layers j and $j+1$ is defined by

$$T_{j,j+1} = \frac{1}{2Z_j} \cdot \begin{bmatrix} Z_j + Z_{j+1} & Z_j - Z_{j+1} \\ Z_j - Z_{j+1} & Z_j + Z_{j+1} \end{bmatrix} \tag{36.15}$$

The propagation of the plane waves through layer j can be described by the transfer matrix:

$$T_j\left(d_j\right) = \begin{bmatrix} \exp\left(2\pi i n_j \cos\theta_j \frac{d_j}{\lambda}\right) & 0 \\ 0 & \exp(-2\pi i n_j \cos\theta_j \frac{d_j}{\lambda}) \end{bmatrix} \tag{36.16}$$

with the thickness d_j of layer j, and the wavelength λ of the incident light. The transfer matrices connect the amplitudes shown in Figure 36.3 as follows:

$$\begin{bmatrix} B_j^+ \\ A_j^+ \end{bmatrix} = T_{j,j+1} \begin{bmatrix} A_{j+1}^- \\ B_{j+1}^- \end{bmatrix} \tag{36.17}$$

$$\begin{bmatrix} A_j^- \\ B_j^- \end{bmatrix} = T_j\left(d_j\right) \begin{bmatrix} B_j^+ \\ A_j^+ \end{bmatrix} \tag{36.18}$$

It is assumed that there is no backward running wave behind the layered medium, and the intensity of the incident radiation is known. Therefore, the amplitudes $A_j^{\pm}$ and $B_j^{\pm}$ at each interface can be calculated with appropriate products of transfer matrices.

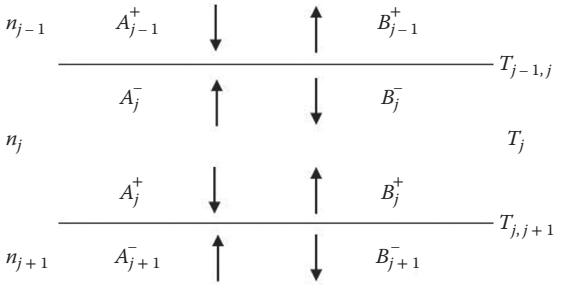

FIGURE 36.3 Wave amplitudes in a layered medium and transfer matrices connecting them.

For both cases of polarization, the intensity in layer j at a distance d from the upper interface $(j, j+1)$ is given by

$$I_{T(E,\mathrm{TM})}(d) = \frac{\Re(Z_j)}{\Re(Z_0)} \cdot \left\| T_j(d) \cdot \begin{pmatrix} A_j^- \\ B_j^- \end{pmatrix} \right\|^2 \tag{36.19}$$

with the proper wave impedances, $\Re(Z_j)$ means the real part of Z_j. If δ is the angle between the vector of the electric field and the plane of incidence, the intensities should be added according to

$$I(d) = aI_{\mathrm{TM}}(d) + I_{\mathrm{TE}}(d) \tag{36.20}$$

where $I_{\mathrm{TM}} = (1-a)I(d)$ and $I_{\mathrm{TE}} = aI(d)$ with $a = \cos^2\delta$.

36.3.4 FDTD Method

Light propagation through the sub-wavelength microstructure (such as nanoparticles, nanometal grating, photonic crystals, photo-trapping structures, and diffractive microlenses)-enhanced photodetectors is simulated by FDTD method based on a rigorous vector solution of Maxwell's equations. The computation area of the FDTD method is discretized into grid space where the electric field vector and the magnetic field vector are respectively arranged in the space and time step by 1/2, and the differential form of Maxwell's equations is rewritten as the central difference form with two-order accuracy.

In the calculation, these correspond to six time varying scalar equations in each Yee's grid (i.e., corresponding to the three spatial components of the electric and magnetic fields, as shown in Figure 36.4), along with the propagation of the electromagnetic field, the electromagnetic field is constantly updated in the set period until convergence. At this time, the corresponding electromagnetic field (light) has a stable distribution in the whole calculation area. At this point, the absorbed power density and the distribution of carrier density are calculated by assuming that each of the photons absorbed could excite an electron–hole pair.

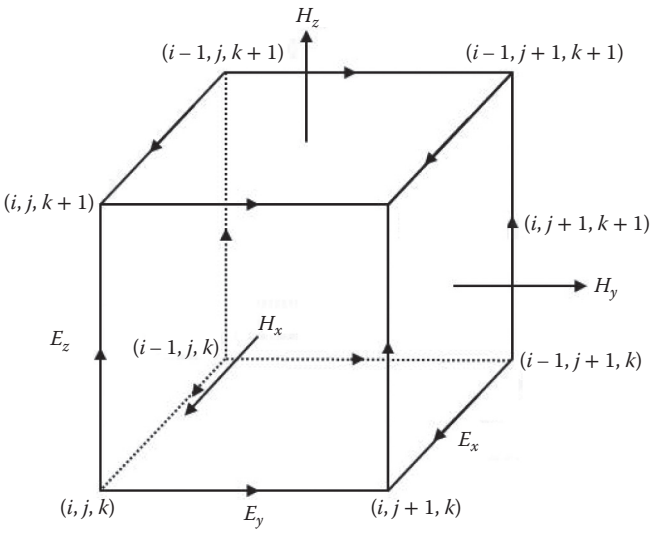

FIGURE 36.4 Structure of 3D Yee grid showing one cell of primary grid.

The differential forms of Maxwell equations are

$$\frac{\partial H}{\partial t} = -\frac{1}{\mu}\nabla \times E - \frac{\rho}{\mu}H \tag{36.21}$$

$$\frac{\partial E}{\partial t} = \frac{1}{\varepsilon}\nabla \times H - \frac{\sigma}{\varepsilon}E \tag{36.22}$$

Maxwell's curl equations can be discretized by the central difference method to achieve second-order accuracy. Equations 36.21 and 36.22 could be rewritten as formulas of electromagnetic field component:

$$\frac{H_x\big|_{i,j,k}^{n+1/2} + H_x\big|_{i,j,k}^{n-1/2}}{\Delta t} = \frac{1}{\mu_{i,j,k}}\left(\frac{E_y\big|_{i,j,k+1/2}^{n} - E_y\big|_{i,j,k-1/2}^{n}}{\Delta z} - \frac{E_z\big|_{i,j+1/2,k}^{n} - E_z\big|_{i,j-1/2,k}^{n}}{\Delta y} - \rho_{i,j,k}\cdot H_x\big|_{i,j,k}^{n}\right) \tag{36.23}$$

$$\frac{H_y\big|_{i,j,k}^{n+1/2} + H_y\big|_{i,j,k}^{n-1/2}}{\Delta t} = \frac{1}{\mu_{i,j,k}}\left(\frac{E_z\big|_{i+1/2,j,k}^{n} - E_z\big|_{i-1/2,j,k}^{n}}{\Delta x} - \frac{E_x\big|_{i,j,k+1/2}^{n} - E_x\big|_{i,j,k-1/2}^{n}}{\Delta z} - \rho_{i,j,k}\cdot H_y\big|_{i,j,k}^{n}\right) \tag{36.24}$$

$$\frac{H_z\big|_{i,j,k}^{n+1/2} + H_z\big|_{i,j,k}^{n-1/2}}{\Delta t} = \frac{1}{\mu_{i,j,k}}\left(\frac{E_x\big|_{i,j+1/2,k}^{n} - E_x\big|_{i,j-1/2,k}^{n}}{\Delta y} - \frac{E_y\big|_{i+1/2,j,k}^{n} - E_y\big|_{i-1/2,j,k}^{n}}{\Delta x} - \rho_{i,j,k}\cdot H_z\big|_{i,j,k}^{n}\right) \tag{36.25}$$

$$\frac{E_x\big|_{i,j,k}^{n+1} + E_x\big|_{i,j,k}^{n}}{\Delta t} = \frac{1}{\varepsilon_{i,j,k}}\left(\frac{H_z\big|_{i,j+1/2,k}^{n+1/2} - H_z\big|_{i,j-1/2,k}^{n+1/2}}{\Delta y} - \frac{H_y\big|_{i,j,k+1/2}^{n+1/2} - H_y\big|_{i,j,k-1/2}^{n+1/2}}{\Delta z} - \sigma_{i,j,k}\cdot E_x\big|_{i,j,k}^{n+1/2}\right) \tag{36.26}$$

$$\frac{E_y\big|_{i,j,k}^{n+1} + E_y\big|_{i,j,k}^{n}}{\Delta t} = \frac{1}{\varepsilon_{i,j,k}}\left(\frac{H_x\big|_{i,j,k+1/2}^{n+1/2} - H_x\big|_{i,j,k-1/2}^{n+1/2}}{\Delta z} - \frac{H_z\big|_{i+1/2,j,k}^{n+1/2} - H_z\big|_{i-1/2,j,k}^{n+1/2}}{\Delta x} - \sigma_{i,j,k}\cdot E_y\big|_{i,j,k}^{n+1/2}\right) \tag{36.27}$$

$$\frac{E_z\big|_{i,j,k}^{n+1} + E_z\big|_{i,j,k}^{n}}{\Delta t} = \frac{1}{\varepsilon_{i,j,k}}\left(\frac{H_y\big|_{i+1/2,j,k}^{n+1/2} - H_y\big|_{i-1/2,j,k}^{n+1/2}}{\Delta x} - \frac{H_x\big|_{i,j+1/2,k}^{n+1/2} - H_x\big|_{i,j-1/2,k}^{n+1/2}}{\Delta y} - \sigma_{i,j,k}\cdot E_z\big|_{i,j,k}^{n+1/2}\right) \tag{36.28}$$

Therefore, the electromagnetic field component can be solved by the Equations 36.23 through 36.28 for any given grid node at any moment.

For more detailed introduction of FDTD, see Chapter 52.

36.3.5 Beam Propagation Method

The BPM can be applied to find the light propagation and penetration into devices such as photodetectors. Its efficiency and relative accuracy are much better than the optical beam absorption method. The BPM solver is available for both 2D and 3D device geometries, where its computational efficiency compared with a full-wave approach becomes particularly apparent in 3Ds.

The BPM is based on the fast Fourier transform (FFT) and is a variant of the FFT BPM, which was developed by Feit and Fleck (1978).

The solution of the scalar Helmholtz equation is

$$\left[\nabla_t^2 + \frac{\partial^2}{\partial z^2} + k_0^2 n^2\left(x, y, z\right)\right]\phi\left(x, y, z\right) = 0 \tag{36.29}$$

At $z + \Delta z$ with $\nabla_t^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ and $n(x, y, z)$ being the complex refractive index can be written as:

$$\frac{\partial}{\partial z}\phi\left(x, y, z + \Delta z\right) = \mp i\sqrt{k_0^2 n^2 + \nabla_t^2}\phi\left(x, y, z\right) = \pm i\xi\phi\left(x, y, z\right) \tag{36.30}$$

In the paraxial approximation, the operator ξ reduces to

$$\xi = \sqrt{k_0^2 n_0^2 + \nabla_t^2} + k_0\delta n \tag{36.31}$$

where n_0 is taken as a constant reference refractive index in every transverse plane. By expressing the field ϕ at z as a spatial Fourier decomposition of plane waves, the solution to Equation 36.29 for forward-propagating waves reads

$$\phi\left(x, y, z + \Delta z\right) = \frac{1}{(2\pi)^2}\exp\left(ik_0\delta n\Delta z\right)\int_{-\infty}^{+\infty} d\vec{k_t}\exp\left(i\vec{k_t}\cdot\vec{r}\right)\exp\left(i\sqrt{k_0^2 n_0^2 \to \vec{k}_t^2}\Delta z\right)\tilde{\phi}\left(\vec{k}_t, z\right) \tag{36.32}$$

where $\tilde{\phi}$ denotes the transverse spatial Fourier transform. As can be seen from Equation 36.32, each Fourier component experiences a phase shift, which represents the propagation in a medium characterized by the reference refractive index n_0. The phase-shifted Fourier wave is then inverse transformed and given an additional phase shift to account for the refractive index inhomogeneity at each $\left(x, y, z + \Delta z\right)$ position. In the numeric implementation of Equation 36.32, an FFT algorithm is used to compute the forward and inverse Fourier transform.

For more detailed introduction of BPM, see Chapter Chapter 4.

36.4 Characteristic Input Parameters

The basic physical mechanism involved in numerical simulations of p-n junction photodiodes is very complicated and, depending on devices and use environments, different material parameters should be chosen for appropriate physical models. Some are related to bandgap such as effective mass, and another part of the parameters contributes to noise such as all of the recombination rate, while the absorption coefficient is dominant for the light absorption properties of device. In addition, the minority carrier lifetime of each region significantly influence the carrier transport. This chapter introduces some typical material parameters and related physical models (Table 36.2).

36.4.1 Minority Carrier Lifetimes

The minority carrier lifetime refers to the average survival time of minority carriers, and can be calculated as:

$$\tau = \frac{\Delta n}{U} \tag{36.33}$$

where Δn is the nonequilibrium carrier concentration and U is the carrier recombination rate.

TABLE 36.2 Parameters That Can Be Specified from Parameter File

Parameters	Units
Room-temperature bandgap, E_g	eV
Dielectric constant, ε_r	1
Density of states in conduction band, N_C	cm^{-3}
Density of states in valance band, N_V	cm^{-3}
Electron mobility, μ_e	$cm^2/V{\cdot}s$
Hole mobility, μ_h	$cm^2/V{\cdot}s$
Electron saturation velocity, v_{Se}	cm/s
Hole saturation velocity, v_{Sh}	cm/s
Electron effective mass (relative), m_e	1
Hole effective mass (relative), m_h	1
Electron Auger coefficient, R_{Ae}	cm^6/s
Hole Auger coefficient, R_{Ah}	cm^6/s
Electron SRH lifetime, τ_{Se}	s
Hole SRH lifetime, τ_{Sh}	s
Radiative coefficient, R_R	cm^3/s
Absorption spectra, A_1	cm^{-1}
Absorption spectra, A_2	cm^{-1}
Absorption spectra, E_1	eV
Absorption spectra, E_2	eV
Absorption spectra, P	1

Note: SRH, Shockley–Read–Hall.

The electron concentration n_0 and the hole concentration p_0 under the thermal equilibrium state fit the function:

$$n_0 p_0 = n_i^2 \tag{36.34}$$

where n_i is the intrinsic carrier concentration.

The generation-recombination rate of radiative recombination model is given by

$$U_R = r(np - n_i^2) \tag{36.35}$$

where r is the electron–hole recombination probability and n, p are the electron and hole concentration, respectively. The minority carrier lifetime is given by

$$\tau = \frac{1}{r(n_0 + p_0 + \Delta p)} \tag{36.36}$$

To the SRH recombination model, the generation-recombination rate can be written as

$$U_{SRH} = \frac{N_t r_n r_p (np - n_i^2)}{r_n (n + n_1) + r_p (p + p_1)} \tag{36.37}$$

with

$$n_1 = n_0 \exp\left(\frac{E_t - E_f}{k_0 T}\right) \tag{36.38}$$

$$p_1 = p_0 \exp\left(-\frac{E_t - E_f}{k_0 T}\right) \tag{36.39}$$

where r_n and r_p are the capture coefficient of electron and hole, N_t is the concentration of recombination center, E_t is the recombination center energy level, E_f is the Fermi energy level, and T is the lattice temperature. The SRH minority carrier lifetime can be calculated by

$$\tau = \frac{r_n\left(n + n_1\right) + r_p(p + p_1)}{N_t r_n r_p(n_0 + p_0 + \Delta p)} \tag{36.40}$$

When the surface recombination model is considered, a revised formula is used based on the SRH recombination model:

$$U_{Surf} = \frac{s_n s_p(np - n_i^2)}{s_p\left(n + n_1\right) + s_n(p + p_1)} \tag{36.41}$$

where s_n and s_p are the surface recombination velocity of electron and hole.

The rate of Auger recombination is given by

$$U_A = (\gamma_n n + \gamma_p p)(np - n_i^2) \tag{36.42}$$

where γ_n and γ_p are the auger recombination coefficient of electron and hole. The Auger recombination minority carrier lifetime can be written by

$$\tau = \frac{1}{\left(n_0 + p_0 + \Delta p\right)\left(\gamma_n n + \gamma_p p\right)} \tag{36.43}$$

36.4.2 Interface Traps and Fixed Charges

The following expressions for trap concentration versus the energy (E) define different types of distributions:

$$\begin{array}{ll}
N_0 \text{ for } E = E_0 & \text{Single-energy level} \\
N_0 \text{ for } E_0 - 0.5E_s < E < E_0 + 0.5E_s & \text{Uniform distribution} \\
N_0 \exp\left(-\left|\dfrac{E - E_0}{E_S}\right|\right) & \text{Exponential distribution} \\
N_0 \exp\left[-\dfrac{(E - E_0)^2}{2E_S^2}\right] & \text{Gaussian distribution} \\
\begin{cases} N_0 & \text{for} \quad E = E_1 \\ N_2 & \text{for} \quad E = E_2 \\ \cdots & \cdots \\ N_m & \text{for} \quad E = E_m \end{cases} & \text{User-defined table distribution}
\end{array} \tag{36.44}$$

Trapped electron and hole concentrations on one energy level are related to occupation probabilities for electrons (r_n) and holes (r_p) as follows:

$$\begin{array}{ll} n_{Dt} = N_{Dt} r_n & n_t = N_{Et} r_n \\ p_{At} = N_{At} r_p & p_t = N_{Ht} r_p \end{array} \tag{36.45}$$

where n_{Dt} is the electron concentration of the donor trap level, p_{At} is the hole concentration of the acceptor trap level, n_t is the electron concentration of the neutral electron trap level, and p_t is the hole concentration of the neutral hole trap level, N_{Dt} is the donor trap concentration, N_{At} is the acceptor trap concentration,

N_{Et} is the neutral electron trap concentration, N_{Ht} is the neutral hole trap concentration. It's obvious that $r_n = 1 - r_p$.

In consideration of the presence of traps, the Poisson equation is revised:

$$\nabla\varepsilon \cdot \nabla\psi = -q\left(p - n + N_{D^+} - N_{A^-} + \sum_{E_t} N_{Dt} - n_{Dt} - \sum_{E_t} N_{At} - p_{At} + \sum_{E_t} p_t - \sum_{E_t} n_t\right) \tag{36.46}$$

where E_t means trap energy levels.

The balance of the carrier flow to and from a trap level gives the following expression for the electron concentration of the trap level:

$$\frac{dn_t}{dt} = \nu_{th}^{n}\sigma_n N_{Et}\left[\frac{n_1}{g_n}r_n - n(1 - r_n)\right] + \nu_{th}^{p}\sigma_p N_{Et}\left[\frac{p_1}{g_p}(1 - r_n) - pr_n\right] \tag{36.47}$$

where ν_{th}^{n} and ν_{th}^{p} are the electron and hole thermal velocities, σ_n and σ_p are the electron and hole capture cross sections, and g_n and g_p are the electron and hole degeneracy factors, respectively.

In the steady state, the trap-assisted recombination process can be analyzed by standard SRH recombination; however, to the transient case, the recombination processes for electrons and holes are different, and the SRH form cannot be applied. The differential equations for the electron-related occupation probability at each energy level are

$$\frac{dr_n}{dt} + \left[\nu_{th}^{n}\sigma_n\left(n + \frac{n_1}{g_n}\right) + \nu_{th}^{p}\sigma_p\left(p + \frac{p_1}{g_p}\right)\right]r_n = \nu_{th}^{n}\sigma_n n + \nu_{th}^{p}\sigma_p\frac{p_1}{g_p} \tag{36.48}$$

This equation is solved self-consistently with the transport and Poisson equations.

The trap total recombination term in the continuity equations is the sum for all levels:

$$R_{trap} = \sum_{E_t} R_{Dt} + \sum_{E_t} R_{Et} + \sum_{E_t} R_{At} + \sum_{E_t} R_{Ht} \tag{36.49}$$

At nonheterointerface vertices, bulk traps are considered only from the region with the lowest bandgap. In cases where the bulk traps from other regions are important, assume appropriate value and solve nonlinear Poisson equation by changing the numeric damping of the trap charge, which properly considers bulk traps from all adjacent regions.

36.4.3 Absorption Coefficients

A widely used absorption coefficient model that fits all the materials is

$$\alpha\left(E_{ph}\right) = \begin{cases} \alpha_1 \exp\left[\dfrac{\left(E_{ph} - E_1\right)}{E_2}\right] & E_{ph} < E_1 \\ \alpha_1 + \alpha_2 \exp\left[\dfrac{\left(E_{ph} - E_1\right)}{E_2}\right]^p & E_{ph} \geq E_1 \end{cases} \tag{36.50}$$

where E_{ph} is the photon energy, and $\alpha_1, \alpha_2, E_1, E_2$, and p could be specified in the parameter file.

The abovementioned parameters can be obtained by fitting the data acquired in the experiment. A fitting result (Wang et al., 2015) for GaN and AlGaN are shown in Figures 36.5 and 36.6.

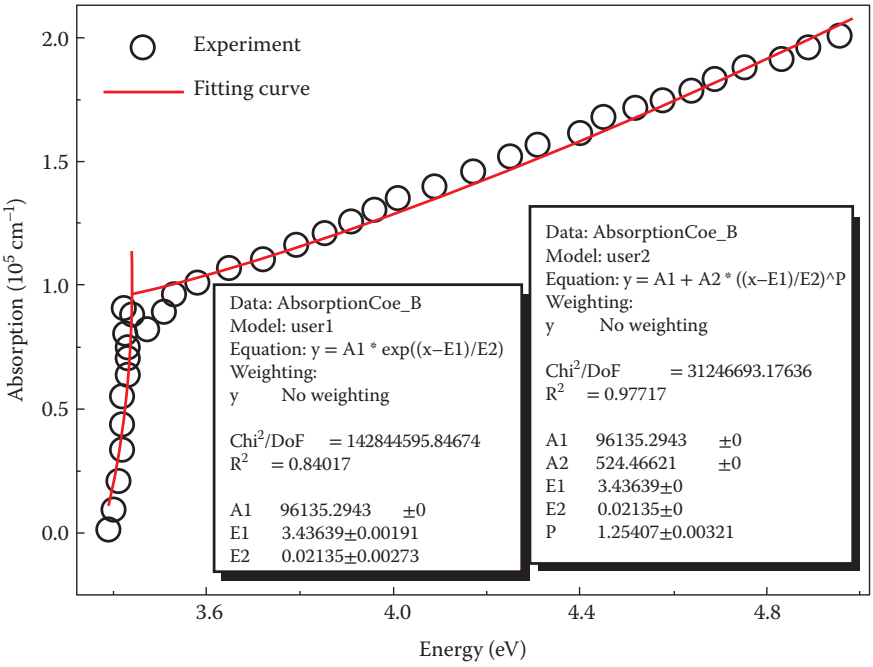

FIGURE 36.5 Fitting results of absorption spectrum for GaN at room temperature.

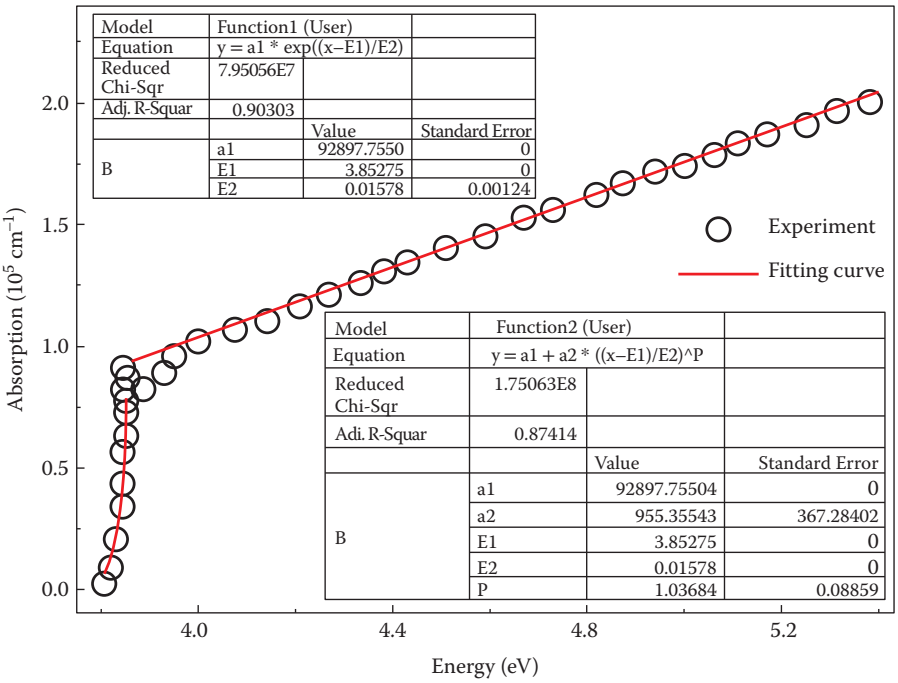

FIGURE 36.6 Fitting results of absorption spectrum for $Al_{0.15}Ga_{0.85}N$ at room temperature.

An absorption model (Rajkanan et al., 1979) is implemented for silicon and materials with similar bandgap structure:

$$\alpha(T) = \sum_{i=1}^{2}\sum_{j=1}^{2} C_i A_j \left\{ \frac{\left[hf - E_{gj}(T) + E_{pi}\right]^2}{e^{-\frac{E_{pi}}{kT}} - 1} + \frac{\left[hf - E_{gj}(T) - E_{pi}\right]^2}{1 - e^{-\frac{E_{pi}}{kT}}} \right\} + A_d \left[hf - E_{gd}(T)\right]^{1/2} \quad (36.51)$$

where

$$E_g(T) = E_g(0) - \frac{\beta T^2}{T + \gamma} \quad (36.52)$$

and T is the temperature. Parameter values used in this formula can be changed for some materials with bandgap structures similar to silicon or for silicon itself.

Instead, the complex refractive index model is adopted to replace the absorption coefficient model in some numerical simulation commercial software. The complex refractive index $\tilde{n}$ can be written as

$$\tilde{n} = n + i \cdot k \quad (36.53)$$

where n refers to the refractive index, and k to the extinction coefficient. The numerical conversion of the extinction coefficient k and the absorption coefficient α is

$$\alpha = \frac{4\pi k}{\lambda} \quad (36.54)$$

For specific extinction coefficients of different materials, one can refer to the website http://refractiveindex.info/.

36.5 Simulation Examples of P-N Junction Photodiodes

In the following section, examples of some widely used p-n junction photodiodes are introduced with the calculation results and parameters shown graphically. First, examples about visible photodiodes using different materials and structure are given. Second, photodiodes simulations applied to ultraviolet and infrared wave band are introduced. Finally, we modeled a two-color heterojunction photodiode, in which more complex mechanism and structure are included. All the absorption models involved in examples are based on Equation 36.50.

36.5.1 Physical Mechanism and Performance Parameters

In order to obtain comprehensive understanding of photoresponse (including dark current) mechanisms of p-n junction photodiodes and improve the device performance, these key parameters, such as electron field (when the effect of electric field intensity becomes dominant, the dark current will be increased and the photoresponse starts reducing); different kinds of junction types (see Figure 36.7a and b), which have a great effect on technological requirements, surface leakage current, and so on; dark current (see Figure 36.7c through f), which will influence our device performance; as well as responsivity and quantum efficiency (QE) should be taken into consideration. In addition, surface leakage current is another consideration of p-n junction photodiodes. A detailed description of the surface leakage current of HgCdTe infrared photodiode is shown in Figure 36.8.

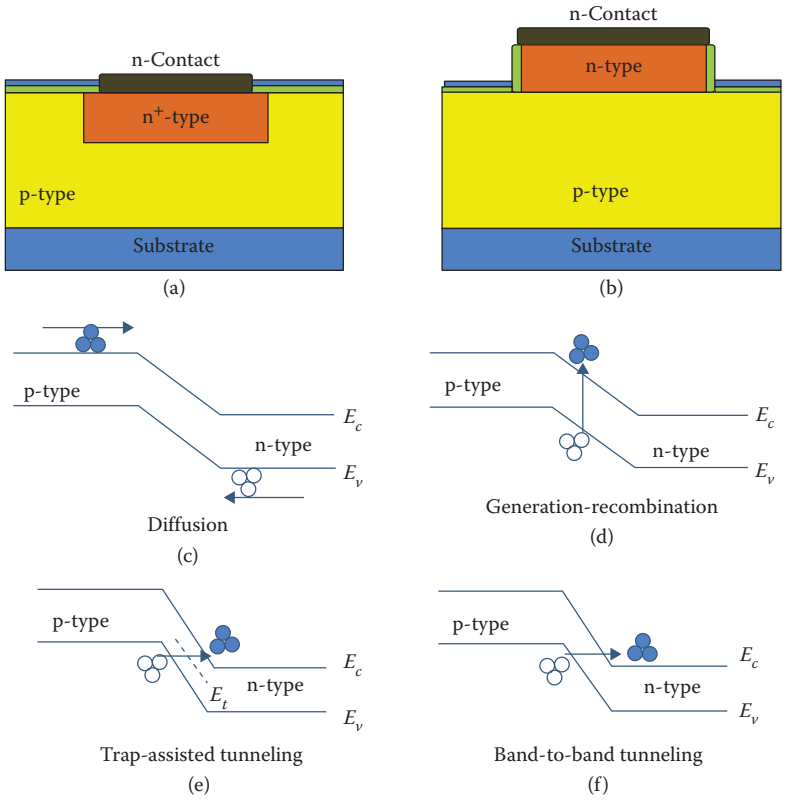

FIGURE 36.7 Schematic structures of two types of main structure of p-n junction for (a) planar junction type, (b) mesa junction type, and the main dark current components of the bulk p-n junction: (c) diffusion, (d) generation-recombination, (e) trap-assisted tunneling, and (f) direct band-to-band tunneling.

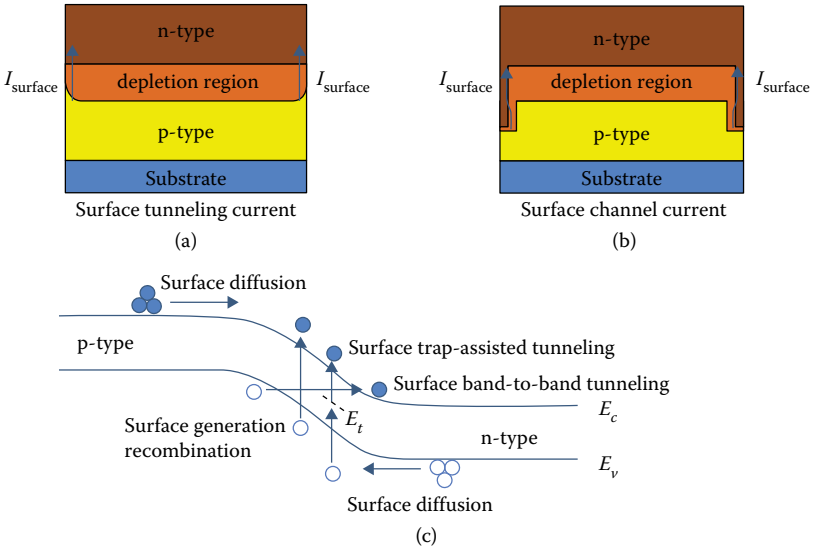

FIGURE 36.8 Surface leakage current of HgCdTe infrared photodiode: (a) surface tunneling current, (b) surface channel current, (c) the main dark current components of the surface-induced p-n junction surface diffusion current, surface generation-recombination, surface trap-assisted tunneling, and surface direct band-to-band tunneling.

The responsivity, quantum efficiency (QE) and cross-talk are index parameters measuring the ability of the device to acquire the target radiation signal. Responsivity is the light current formed by unit radiation power. The calculation formula is

$$R = \frac{I_0}{P_{\text{opt}}} \tag{36.55}$$

where R is the responsivity, I_0 is the optical current under zero bias, and P_{opt} is the power of incident light.

QE means the efficiency of the device with which the incident photons of light are converted to current:

$$\eta = \left(\frac{I_p}{q}\right)\left(\frac{P_{\text{opt}}}{h\nu}\right)^{-1} = \left(\frac{I_0}{P_{\text{opt}}}\right)h\nu = R\frac{1.24}{\lambda} = \frac{1.24R}{\lambda} \tag{36.56}$$

where λ is the wavelength of incident light and I_p is the light excitation current when the wavelength is λ (corresponding to the photon energy $h\nu$) and the power of incident light is P_{opt}.

Cross-talk is when the radiation signal in a unit produces a response signal R_0, the adjacent unit also appears in the response signal R_1. Cross-talk can be used to measure the image quality detection. The cross-talk is defined by:

$$C = \frac{R_1}{R_0} \times 100\% \tag{36.57}$$

36.5.2 Infrared Photodiodes

36.5.2.1 InP/$In_{0.53}Ga_{0.47}As$/InP PIN Photodiode

For an InP/$In_{0.53}Ga_{0.47}As$/InP PIN photodiode (Wang et al., 2008), the thickness of the absorption layer (T_{abs}) as a key structural parameter can determine, to a large extent, the magnitude of responsivity. The following discussion is concentrated on the effect of T_{abs} on the responsivity, and the following simulations are performed only at room temperature. Figure 36.9 shows the 2D structure of the device. The structure consists of a 0.5-μm heavily doped n-type InP buffer layer, a lightly doped n-type InGaAs absorption layer, and a 0.2-μm heavily doped p-type InP cap layer. The thickness of the InGaAs absorption layer (T_{abs}) is set as 1.5, 3.5, and 5.5 μm. Figure 36.10 presents the energy band of InP/InGaAs PIN device structure, respectively.

Here, the thickness of the InGaAs absorption layer is set as 1.5 μm. Vertical distributions of electric field with varying T_{abs} from 1.5 to 5.5 μm in 2-μm steps are investigated as shown in Figure 36.11; the dark and

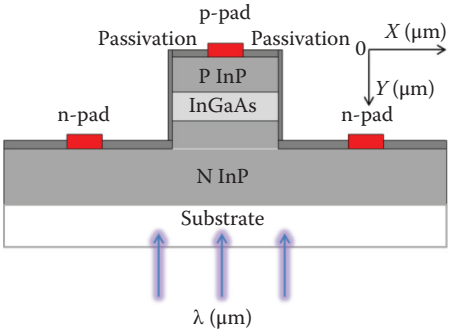

FIGURE 36.9 Schematic cross section of InP/InGaAs PIN photodiode.

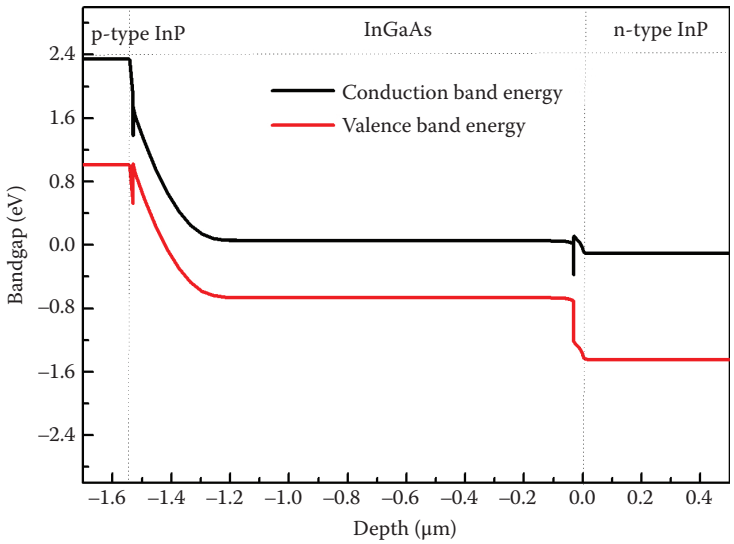

FIGURE 36.10 Energy band of the InP/InGaAs PIN photodiode.

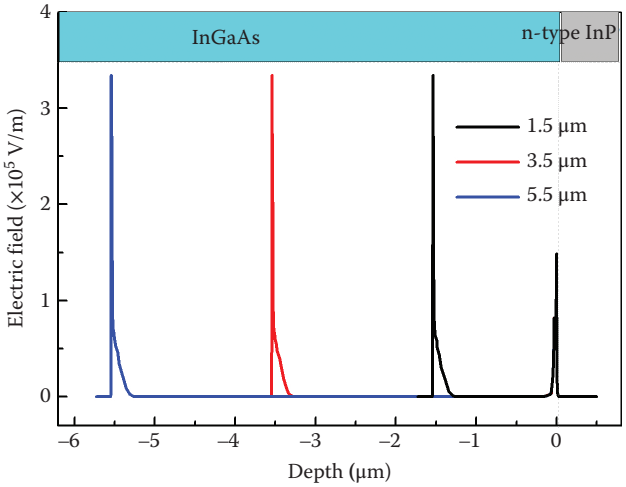

FIGURE 36.11 Vertical distributions of electric field with T_{abs} increasing from 1.5 to 5.5 μm in 2-μm steps.

light current characteristics with the absorption layer (T_{abs}) increasing from 1.5 to 5.5 μm in 1-μm steps are shown in Figure 36.12; the responsivity and QE, as a function of T_{abs} for different incident wavelengths, are shown in Figure 36.13. For clarity, the fitting parameters used in the simulation are listed in Table 36.3.

36.5.2.2 HgCdTe-Based P-N Junction Photodiode

Another example is the HgCdTe-based p-n junction photodiode infrared detector (Hu et al., 2008, 2010; Liang et al., 2014; Qiu and Hu, 2015; Yin et al., 2009; Wang et al., 2011a). The thickness of the n-region of the device has been chosen as a key structural parameter. This section discusses the effect of the thickness of the n-region on responsivity and QE of the device. The following simulations are performed at 77 K. Figure 36.14 presents a schematic diagram of the device where the white region represents the depletion region which is a space charge layer of a PN junction with positive charges in the n-type region and negative

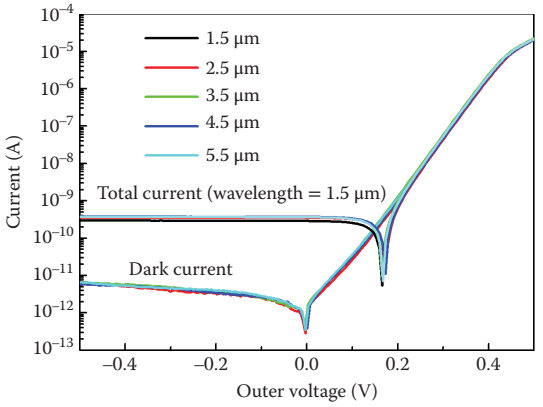

FIGURE 36.12 Dark and light current characteristics with the absorption layer (T_{abs}) increasing from 1.5 to 5.5 μm in 1-μm steps.

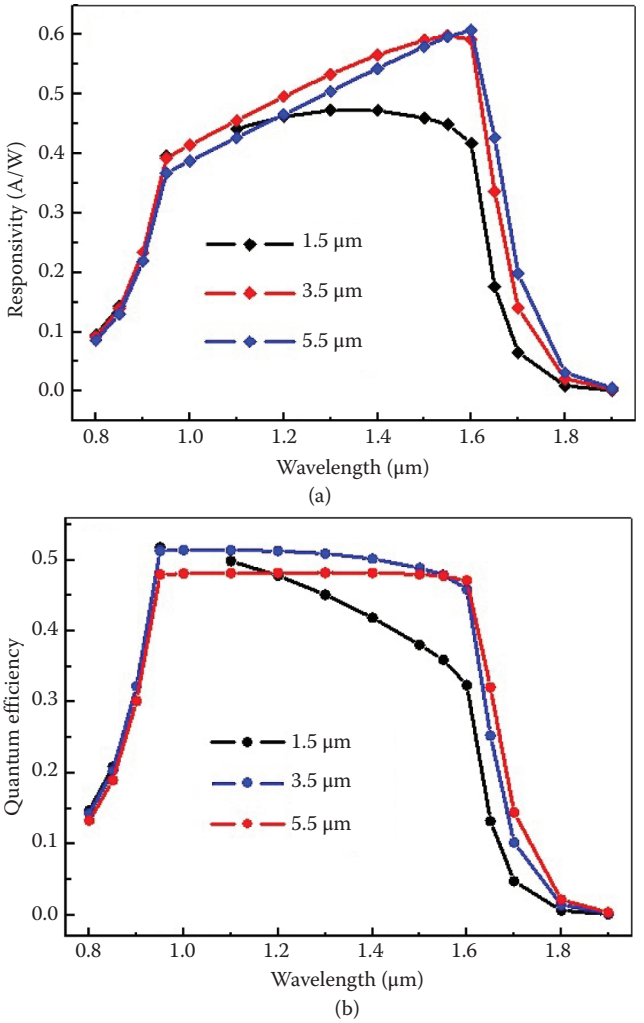

FIGURE 36.13 (a) Responsivity and (b) quantum efficiency for InP/InGaAs/InP PIN photodiode with the absorption layer (T_{abs}) increasing from 1.5 to 5.5 μm in 2-μm steps.

TABLE 36.3 InP/InGaAs/InP PIN Photodiode Simulation Parameter List

Parameters	Units	$In_{0.53}Ga_{0.47}As$	InP
Room-temperature bandgap, E_g	eV	0.78	1.34
Dielectric constant, ε_r	1	13.9	12.4
Density of states in conduction band, N_C	cm^{-3}	2.75×10^{17}	5.66×10^{17}
Density of states in valance band, N_V	cm^{-3}	7.62×10^{18}	2.03×10^{19}
Electron mobility, μ_e	$cm^2/V.\cdot s$	12,000	4,730
Hole mobility, μ_h	$cm^2/V.\cdot s$	450	151
Electron saturation velocity, v_{Se}	cm/s	8.9×10^6	2.6×10^7
Hole saturation velocity, v_{Sh}	cm/s	8.9×10^6	2.6×10^7
Electron effective mass (relative), m_e	1	0.0489	0.08
Hole effective mass (relative), m_h	1	0.45	0.861
Electron Auger coefficient, R_{Ae}	cm^6/s	3.2×10^{-28}	3.7×10^{-31}
Hole Auger coefficient, R_{Ah}	cm^6/s	3.2×10^{-28}	8.7×10^{-30}
Electron SRH lifetime, τ_{Se}	s	1×10^{-6}	1×10^{-9}
Hole SRH lifetime, τ_{Sh}	s	1×10^{-6}	1×10^{-11}
Radiative coefficient, R_R	cm^3/s	1.43×10^{-10}	2×10^{-11}
Absorption spectra, A_1	cm^{-1}	6687.08	6045.319
Absorption spectra, A_2	cm^{-1}	618.99	4008.099
Absorption spectra, E_1	eV	0.776	1.363
Absorption spectra, E_2	eV	0.012	0.013
Absorption spectra, P	1	1.013	0.65

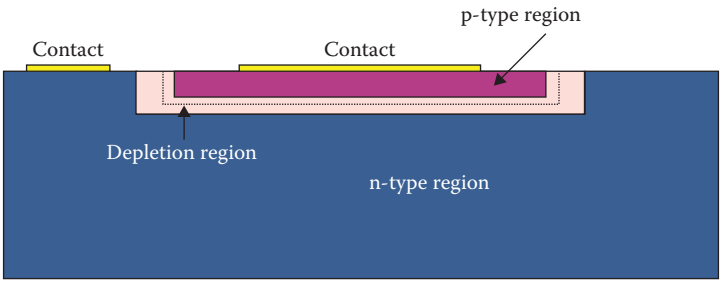

FIGURE 36.14 Schematic cross section of HgCdTe p-n junction photodiode.

charges in the p-type region. Figure 36.15 shows the energy band of the device with different thicknesses (d). Vertical distributions of electric field with the thickness of device (d) increasing from 3 to 7 in 2-μm steps under reverse bias of 200 mV are shown in Figure 36.16. Dark current and photocurrent with varying d from 3 to 7 μmin 2-μm steps under reverse bias are shown in Figure 36.17. The vertical distributions of the electric field of the device with three different conditions are very close. The responsivity and QE, as a function of d, for different incident wavelengths are shown in Figure 36.18. The responsivity and QE of the device with thickness of 5 μm are greater than those devices with thickness of 3 μm, because more photons can be absorbed by the material as the absorber region becomes thicker. However, the responsivity and QE of the device with thickness of 7 μm is smaller than that of the device with the thickness of 5 μm, because the recombination rate of the device with thickness of 7 μm increases, causing that photon-generated carriers to recombine before they arrive at the contacts. For clarity, the fitting parameters used in the simulations are listed in Table 36.4.

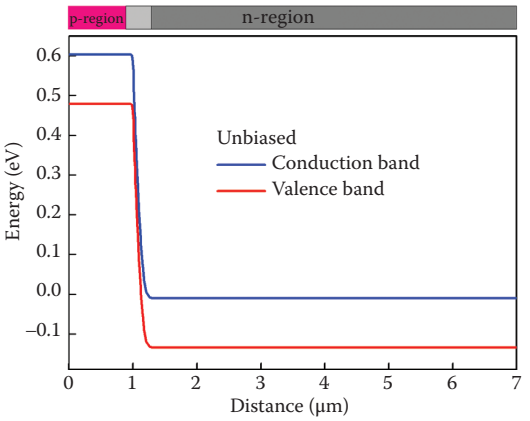

FIGURE 36.15 Energy band of HgCdTe p-n junction photodiode.

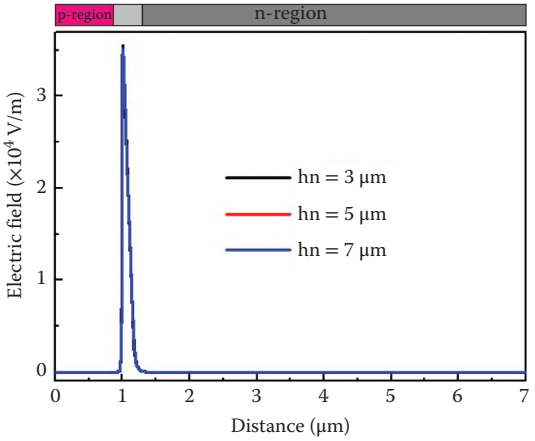

FIGURE 36.16 Vertical distributions of electric field with the height of n-region (hn) increasing from 3 to 7 μm in 2-μm steps.

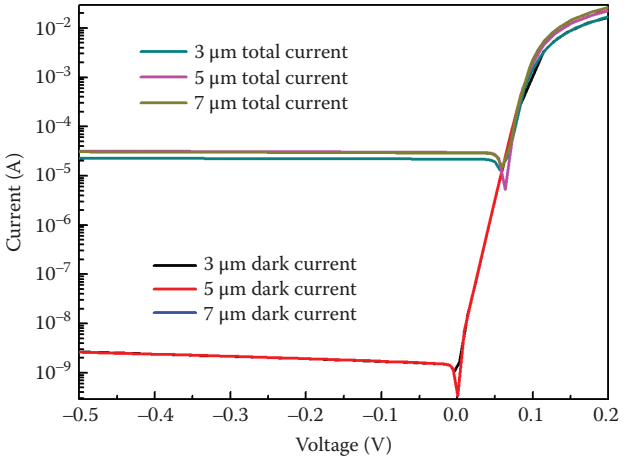

FIGURE 36.17 Dark and light current characteristics with the height of n-region increasing from 3 to 7 μm in 2-μm steps.

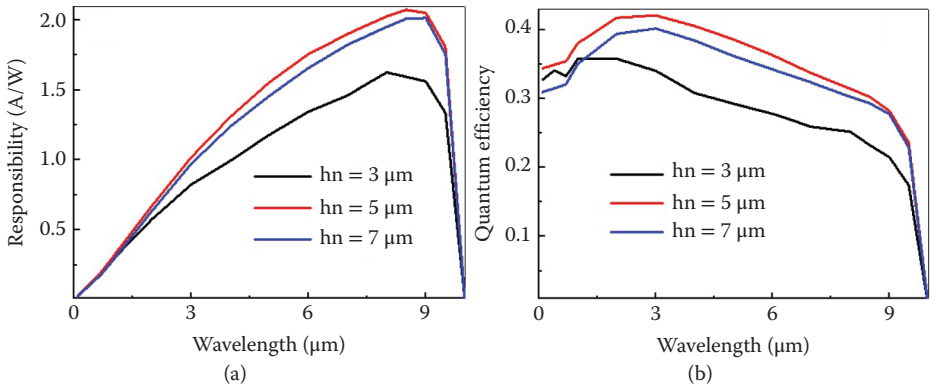

FIGURE 36.18 (a) Responsivity and (b) quantum efficiency for HgCdTe PN photodiode with the height of n-region (hn) increasing from 3 to 7 μm in 2-μm steps.

TABLE 36.4 List of the Key Parameters Used in the Simulation

Parameters	Units	HgCdTe
Bandgap	eV	0.12
Dielectric constant	1	17.4
Effective conduction band density of states	cm^{-3}	0.64×10^{19}
Effective valence band density of states	cm^{-3}	0.70×10^{19}
Electron mobility	$cm^2/V.\cdot s$	6,1367
Hole mobility	$cm^2/V.\cdot s$	610
Temperature exponent	1	0
Electron effective mass (relative)	1	1.09
Hole effective mass (relative)	1	1.15
Electron Auger coefficient	cm^6/s	2.3×10^{-25}
Hole Auger coefficient	cm^6/s	4.66×10^{-26}
Electron SRH lifetime	s	1×10^{-7}
Hole SRH lifetime	s	1×10^{-7}
Radiative recombination coefficient	cm^3/s	3.47×10^{-17}

36.5.3 Visible Photodiodes

36.5.3.1 Si P-N Junction Photodiode

In the Si p-n junction photodiode, the influence of p-region thickness, operating temperature, and doping is calculated. Figures 36.19 and 36.20 show a schematic diagram and the energy band of a silicon p-n junction device structure, respectively. Vertical distributions of electric field with varying doping concentration are investigated as shown in Figure 36.21. Figure 36.22 shows the light and dark current distribution as a function of different operating temperatures. The responsivity and QE as a function of incident wavelengths for different p-region absorption layer are shown in Figure 36.23. Parameters not described are all according to those given as (default) in parameter Table 36.5.

36.5.4 Ultraviolet Photodiodes

36.5.4.1 GaN/AlGaN PIN Photodiode

The structure of a visible blind GaN/AlGaN PIN photodiode (Wang et al., 2011b, 2014) consists of three epitaxial layers grown on a 2-inch-diameter sapphire substrate. The first layer is a 0.6-μm n^+-doped AlGaN

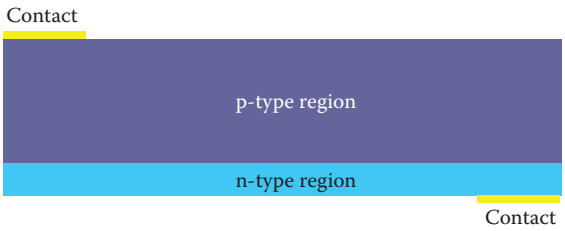

FIGURE 36.19 Schematic cross section of Si p-n junction photodetector.

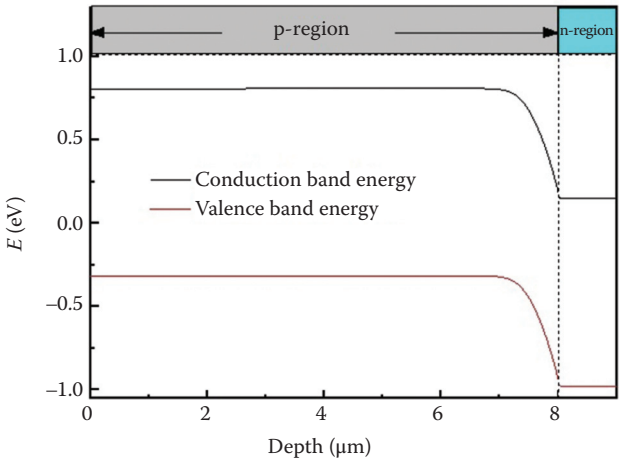

FIGURE 36.20 Energy band of Si p-n junction photodetector.

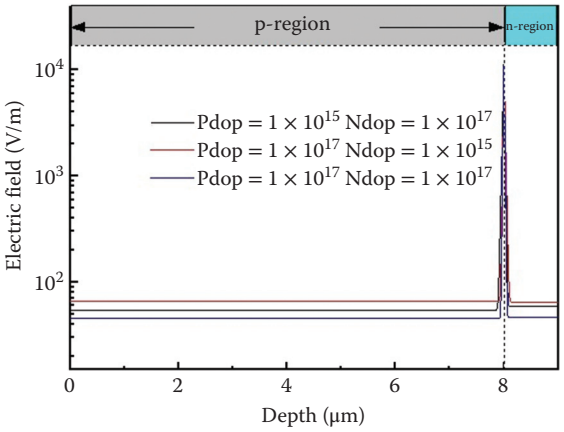

FIGURE 36.21 Vertical distributions of electric field with temperature increasing from 280 to 320 K in 20-K steps.

layer (n-layer) with a doping concentration of 3×10^{18} cm^{-3}. This layer is followed by a 0.1-μm unintentionally doped GaN absorption layer with an electron concentration of 5×10^{15} cm^{-3}, and a 0.15 μm p^+-doped GaN layer (p-layer) with a doping concentration of 3×10^{17} cm^{-3}. For simplicity, the thicknesses of the absorption layer and n-layer are defined as d_{abs} and d_n, respectively. The schematic cross section of

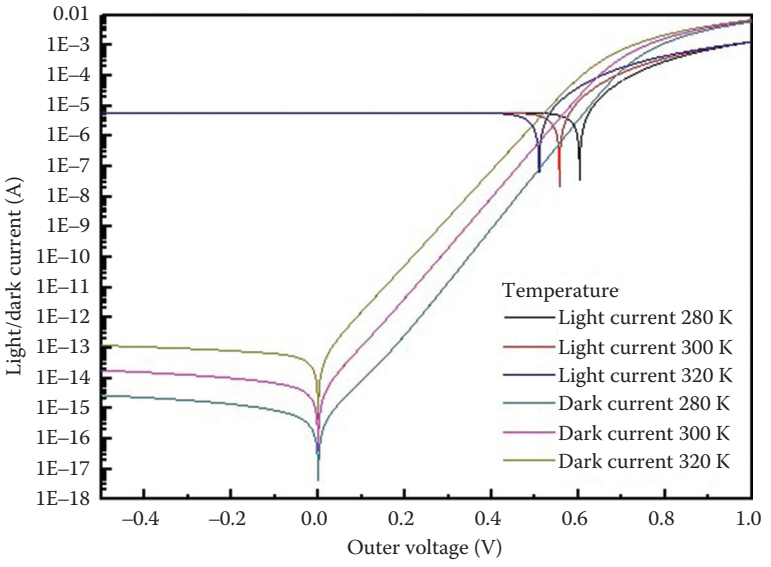

FIGURE 36.22 Dark and light current characteristics with temperature increasing from 280 to 320 K in 20-K steps.

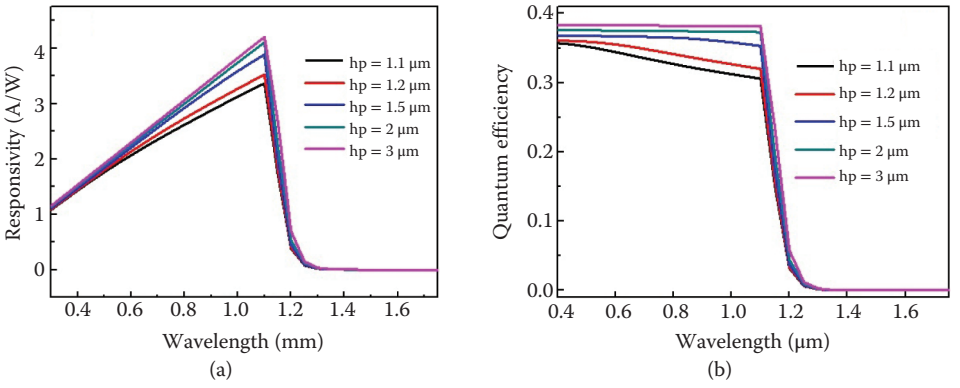

FIGURE 36.23 (a) Responsivity and (b) quantum efficiency for Si PN photodiode with the height of p-region (hp) under 1.1, 1.2, 1.5, 2, and 3 μm.

GaN/AlGaN PIN photodiode is shown in Figure 36.24. Figure 36.25 shows energy band as a function of vertical distance at equilibrium state, and the d_{abs} is set as 0.1 μm. Figure 36.26 describes responsivity versus the absorption layers (d_{abs}) with incident wavelengths of 0.30, 0.32, 0.34, and 0.36 μm, respectively, and electric field profile in the vertical direction with the absorption layers (d_{abs}) under 0.05, 0.2, 0.4, 0.6, and 0.8 μm.

The dark and light current characteristics with a bias voltage are shown in Figure 36.27. The absorption layer (d_{abs}) is set as 0.2 μm. Figure 36.28 shows responsivity and QE for GaN/AlGaN/GaN PIN photodiode with the absorption layer (d_{abs}) under 0.2, 0.3, 0.4, 0.6, and 1.1 μm. For clarity, the fitting parameters used in the simulation are listed in Table 36.6.

TABLE 36.5 List of the Key Parameters Used in the Simulation

Parameters	Units	Silicon
Bandgap	eV	1.124
Temperature (default)	K	300
Dielectric constant		11.7
p-type doping level (default)	cm^{-3}	10^{15}
n-type doping level (default)	cm^{-3}	10^{17}
p-type layer thickness (default)	μm	8
n-type layer thickness	μm	1
Lattice heat capacity	$J/K{\cdot}cm^3$	1.63
Electron affinity	eV	4.05
Electron's lifetime	s	10^{-5}
Hole's lifetime	s	10^{-6}
Electron effective mass		1.09
Hole effective mass		1.15
Electron mobility	$cm^2/V{\cdot}s$	1.42×10^3
Hole mobility	$cm^2/V{\cdot}s$	4.70×10^2

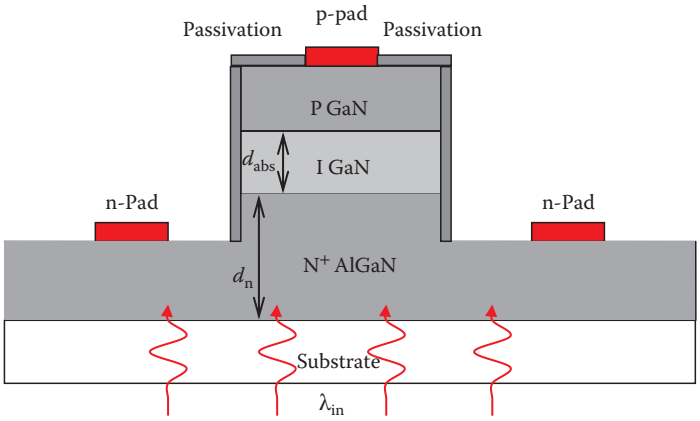

FIGURE 36.24 Schematic cross section of GaN/AlGaN/GaN PIN photodiode.

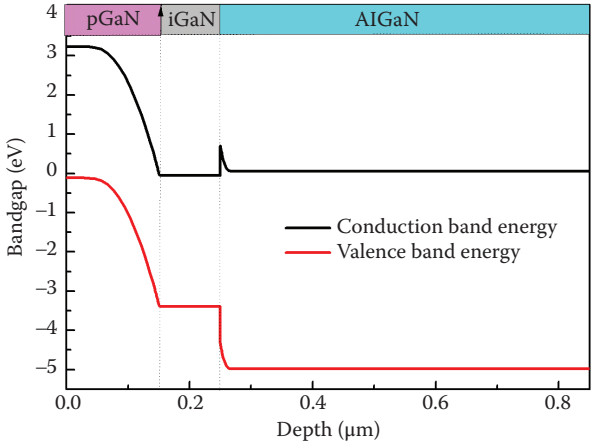

FIGURE 36.25 Energy band of GaN/AlGaN/GaN PIN photodiode.

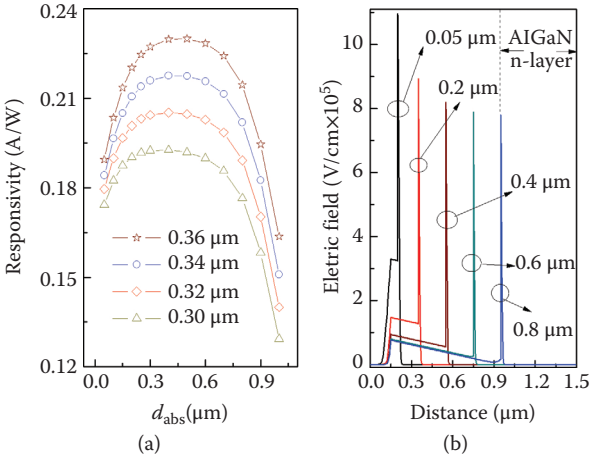

FIGURE 36.26 (a) Responsivity versus the absorption layers (d_{abs}) with incident wavelengths of 0.30, 0.32, 0.34, and 0.36 μm, respectively and (b) electric field profile in the vertical direction with the absorption layers (d_{abs}) changing from 0.05 to 0.8 μm.

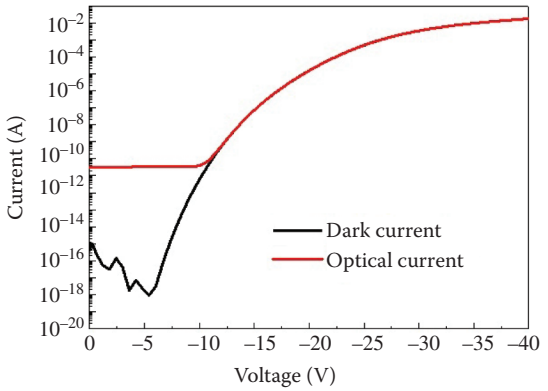

FIGURE 36.27 Dark and light current characteristics of GaN/AlGaN/GaN PIN photodiode.

36.5.5 Two-Color Heterojunction Photodiodes

36.5.5.1 Long-Wavelength/Mid-Wavelength Two-Color HgCdTe

There are two important parameters for the long-wavelength/mid-wavelength (LW/MW) two-color HgCdTe infrared focal plane array detector (Hu et al., 2009, 2010, 2011, 2012, 2013, 2014; Qiu et al., 2016): (1) absorption layer thickness and (2) opening dimensions of the groove (the length of the groove top, L_{top}; the length of the groove bottom, L_{bot}). In the simulation, the electrical cross-talk of both LW to MW ($C_{\lambda e\text{-LW-to-MW}}$) and MW to LW ($C_{\lambda e\text{-MW-to-LW}}$) contributes to the total spectral cross-talk dependent on thickness and band energy offset of the barrier layer (as shown in Figure 36.29). Figure 36.30 shows the spectral photoresponse for the HgCdTe two-color infrared detector with the proposed structure from the previous numerical simulations. Figure 36.31 shows the QE and cross-talk as a function of the thickness of the MW layer for the grooved n_1^+-p_1-P-p_2-n_2^+ HgCdTe two-color infrared detector. It is found that a minimum MW-to-LW cross-talk and maximum MW QE can be achieved with an MW layer of approximately 7 μm. To further optimize the groove structure, the QEs with different L_{top} and L_{bot} are calculated at cutoff wavelengths. It indicates that the MW QE at the MW cutoff wavelength of 4.8 μm increases with L_{bot} (no

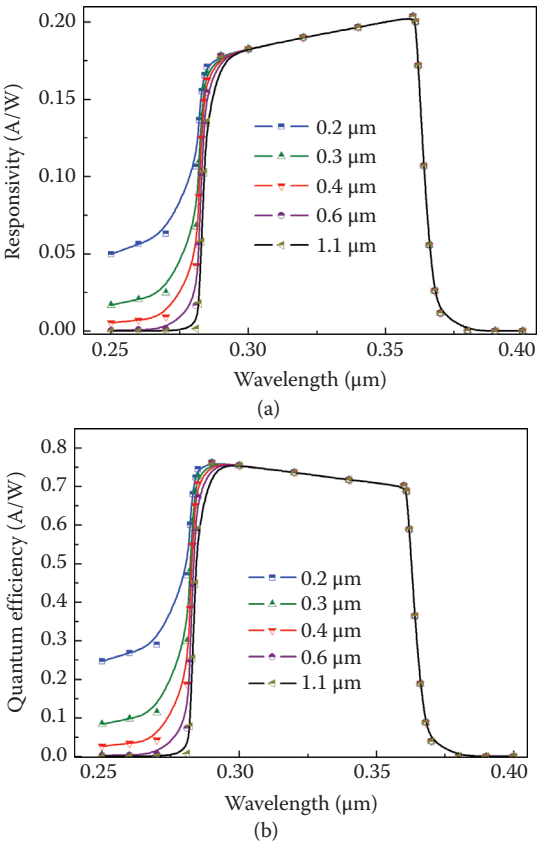

FIGURE 36.28 (a) Responsivity and (b) Quantum efficiency for GaN/AlGaN/GaN PIN photodiode with the absorption layer (Tabs) under 0.2,0.3,0.4,0.6, and 1.1 μm.

TABLE 36.6 List of the Key Parameters Used in the Simulation

Parameters	Units	$Al_{0.3}Ga_{0.7}N$	GaN
Band gap	eV	4.29	3.47
Dielectric constant	1	9.2	9.5
Effective conduction band density of states	cm^{-3}	3.09×10^{18}	2.65×10^{18}
Effective valence band density of states	cm^{-3}	1.03×10^{20}	2.5×10^{19}
Electron mobility	$cm^2/V{\cdot}s$	800	1010
Hole mobility	$cm^2/V{\cdot}s$	18	20
Saturation velocity	cm/s	1.5×10^{17}	1.5×10^{17}
Temperature exponent	1	0	0
Electron effective mass (relative)	1	0.25	0.222
Hole effective mass (relative)	1	2.2	1
Electron Auger coefficient	cm^6/s	2.1×10^{-29}	3.2×10^{-29}
Hole Auger coefficient	cm^6/s	6.5×10^{-30}	8.7×10^{-30}
Electron SRH lifetime	s	1×10^{-10}	1×10^{-9}
Hole SRH lifetime	s	1×10^{-10}	1×10^{-9}
Radiative recombination coefficient	cm^3/s	2×10^{-10}	2×10^{-10}

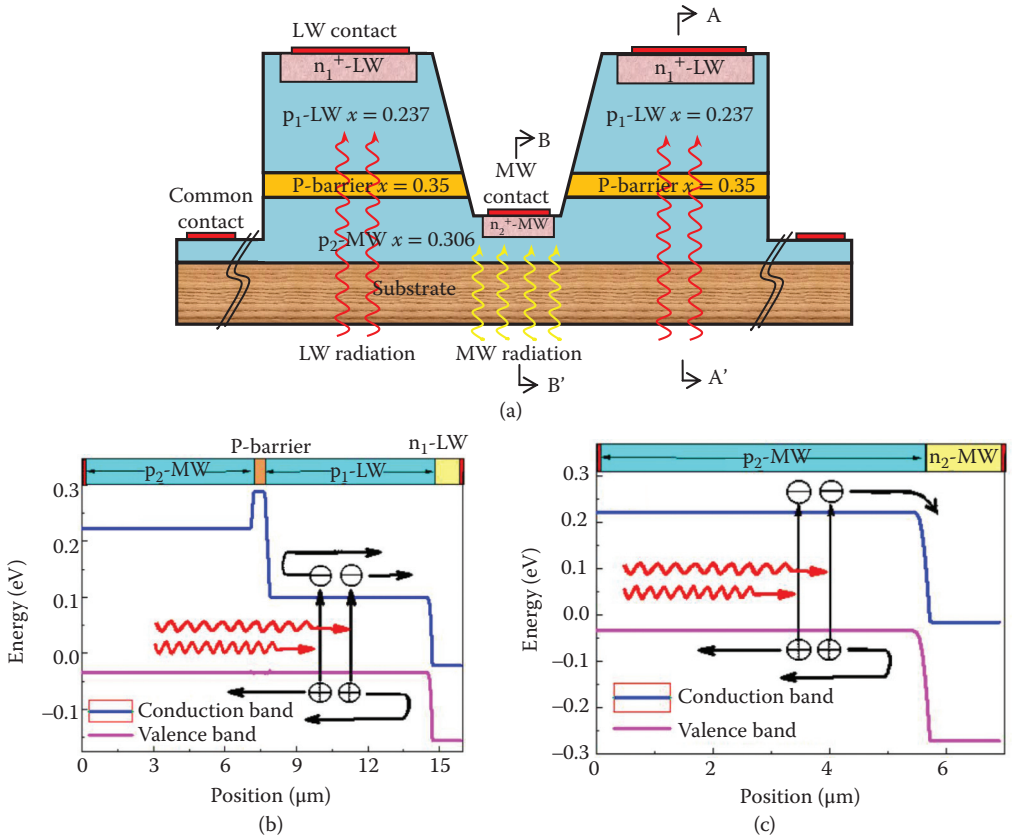

FIGURE 36.29 (a) Schematic of grooved HgCdTe two-color infrared detector, (b) equilibrium energy band diagram cut at A-A′, and (c) equilibrium energy band diagram cut at B-B′.

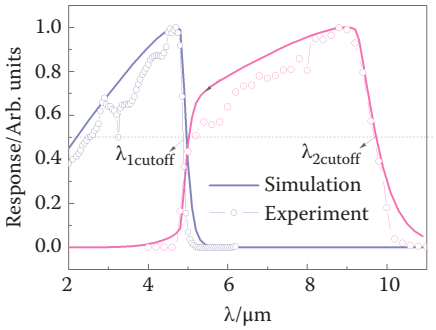

FIGURE 36.30 Experimental (dotted) and simulated (solid) spectral photoresponse of LW/MW HgCdTe two-color infrared detector with cutoff wavelengths of $\lambda_{1cutoff} = 4.8$ μm and $\lambda_{2cutoff} = 9.7$ μm for the LW and MW diodes respectively. A voltage of 0.01 V is used in simulation and experiment.

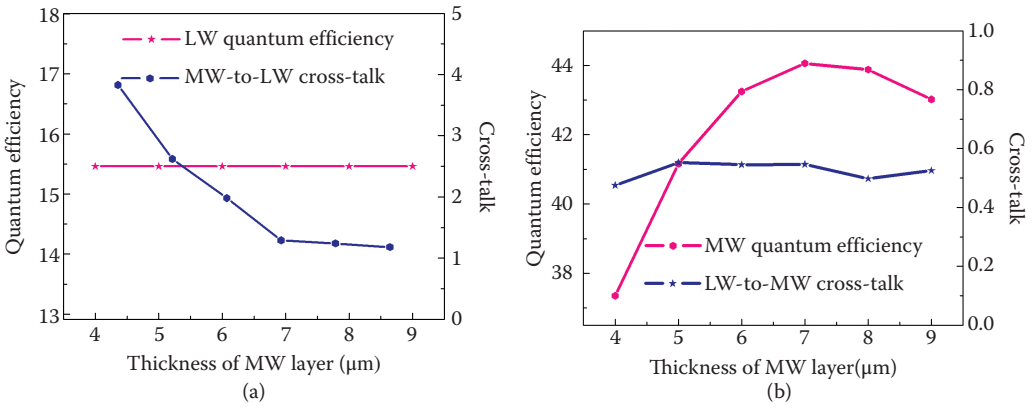

FIGURE 36.31 (a) LW quantum efficiency (QE) and MW-to-LW cross-talk and (b) MW QE and LW-to-MW cross-talk as a function of the thickness of MW layer for the infrared detector. The QE is calculated at cutoff wavelength.

TABLE 36.7 Material and Structural Parameters of the Two-color HgCdTe Photodiode at 77 K

Parameters	Value of LW (Units)	Value of MW (Units)	Value of Barrier (Units)
Cd molar fraction (x)	0.232	0.309	0.35
Doping density of n-region	1×10^{17} (cm^{-3})	1×10^{17} (cm^{-3})	–
Doping density of p-region	8×10^{15} (cm^{-3})	8×10^{15} (cm^{-3})	8×10^{15} (cm^{-3})
Thickness of n-region	1.2 (μm)	1.2 (μm)	–
Thickness of p-region	9 (μm)	5.8 (μm)	0.8 (μm)

dependence on L_{top}). When $L_{bot} \geq 14$ μm, the MW QE saturates at a maximum value as the photogenerated carriers with the diffuse length of ~10 μm are efficiently collected by the MW contact. The specific details of the detector are listed in Table 36.7.

36.6 Summary

With the rapid development of computational capabilities and algorithm, numerical simulation becomes a very important tool in both qualitative and quantitative studies of p-n junction photodiodes. In this chapter, physical equations, models, and parameters that play important roles in the numerical simulation have been introduced. First, for the construction of basic theoretical framework, the well-known Poisson and continuity equations are introduced. Meanwhile, the widespread used drift-diffusion model and thermodynamic model is also described. The results show the effects of the important device parameters such as structure size, doping concentration, and incident wavelength on the device performance such as dark current, photoresponsivity, QE, and so on. Second, we introduced and compared the characteristics of several common optical generation models in detail, which could achieve optoelectronic simulations of different precision and structural requirements. Third, the characteristic input parameters and widely used models in the simulations of typical p-n junction photodiodes are discussed and introduced. Different input parameters and models can be selected according to different optoelectronic materials, device structure, and microscopic physical mechanism. Finally, we introduced some key parameters and measures of diode simulation, and simulation examples of some typical p-n junction photodiodes are given. In addition, some important results are numerically extracted, and the parameter tables used in the simulations are provided.

References

Feit MD and Fleck JA (1978) Light propagation in graded-index optical fibers, *Applied Optics*, 17: 3990–3998.

Hu WD, Chen XS, Ye ZH, Chen YG, Yin F, Zhang B, and Lu W (2012) Polarity inversion and coupling of laser beam induced current in As-doped long-wavelength HgCdTe infrared detector pixel arrays: Experiment and simulation, *Applied Physics Letters*, 101: 181108.

Hu WD, Chen XS, Ye ZH, Feng AL, Yin F, Zhang B, Liao L, and Lu W (2013) Dependence of ion-implant-induced LBIC novel characteristic on excitation intensity for long-wavelength HgCdTe-based photovoltaic infrared detector pixel arrays, *IEEE Journal of Selected Topics in Quantum Electronics*, 19: 4100107.

Hu WD, Chen XS, Ye ZH, Lin C, Yin F, Zhang J, Li ZF, and Lu W (2010a) Accurate simulation of temperature-dependence of dark current in HgCdTe infrared detector assisted by analytical modeling, *Journal of Electronic Materials*, 39: 981–985.

Hu WD, Chen XS, Ye ZH, and Lu W (2010b) An improvement on short-wavelength photoresponse for heterostructure HgCdTe two-color infrared detector, *Semiconductor Science and Technology*, 25: 045028.

Hu WD, Chen XS, Ye ZH, and Lu W (2011) A hybrid surface passivation on HgCdTe long wave infrared detector with in-situ CdTe deposition and high-density hydrogen plasma modification, *Applied Physics Letters*, 99: 091101.

Hu WD, Chen XS, Yin F, Ye ZH, Lin C, Hu XN, Li ZF, and Lu W (2009) Numerical analysis of two-color HgCdTe infrared photovoltaic heterostructure detector, *Optical and Quantum Electronics*, 41: 699–704.

Hu WD, Chen XS, Yin F, Ye ZH, Lin C, Hu XN, Quan ZJ, Li ZF, and Lu W (2008) Simulation and design consideration of photoresponse for HgCdTe infrared photodiodes, *Optical and Quantum Electronics*, 40: 1255–1260.

Hu WD, Ye ZH, Liao L, Chen HL, Chen L, Ding RJ, He L, Chen XS, and Lu W (2014) A 128 × 128 long-wavelength/mid-wavelength two-color HgCdTe infrared focal plane array detector with ultra-low spectral crosstalk, *Optics Letters*, 39: 5130–5133.

Liang J, Hu WD, Ye ZH, Liao L, Li ZF, Chen XS, and Lu W (2014) Improved performance of HgCdTe infrared detector focal plane arrays by modulating light field based on photonic crystal structure, *Journal of Applied Physics*, 115:184504.

Qiu WC and Hu WD (2015) Laser beam induced current microscopy and photocurrent mapping for junction characterization of infrared photodetectors, *Science China-Physics Mechanics & Astronomy*, 58 (2): 1–13.

Qiu WC, Hu WD, Lin C, Chen XS, and Lu W (2016) Surface leakage current in 12.5 μm long-wavelength HgCdTe infrared photodiode arrays, *Optics Letters*, 41 (4): 828–831.

Rajkanan K, Singh R, and Shewchun J (1979) Absorption coefficient of silicon for solar cell calculations, *Solid-State Electronics*, 22: 793–795.

Shi M and Li ML (2014) *Semiconductor Devices Physics and Technology*, Soochow, China: Soochow University Press.

Wang XD, Chen XY, Hou LW, Wang BB, Xie W, and Pan M (2015) Role of n-type AlGaN layer in photoresponse mechanism for separate absorption and multiplication (SAM) GaN/AlGaN avalanche photodiode, *Optical and Quantum Electronics*, 47: 1357–1365.

Wang J, Chen XS, Hu WD, Wang L, Lu W, Xu FQ, Zhao J, Shi YL, and Ji RB (2011a) Amorphous HgCdTe infrared photoconductive detector with high detectivity above 200 K, *Applied Physics Letters*, 99: 113508.

Wang XD, Hu WD, Chen XS, Lu W, Tang HJ, Li T, and Gong HM (2008) Dark current simulation of InP/In0.53Ga0.47As/InP p-i-n photodiode, *Optical and Quantum Electronics*, 40: 1261–1266.

Wang XD, Hu WD, Chen XS, Xu JT, Li XY, and Lu W (2011b) Photoresponse study of visible blind GaN/AlGaN p-i-n ultraviolet photodetector, *Optical and Quantum Electronics*, 42: 755–764.

Wang XD, Hu WD, Pan M, Hou LW, Xie W, Xiu JT, Li XY, Chen XS, and Lu W (2014) Study of gain and photoresponse characteristics for back-illuminated separate absorption and multiplication GaN avalanche photodiodes, *Journal of Applied Physics*, 115:013103.

Yin F, Hu WD, Zhang B, Li ZF, Hu XN, Chen XS, and Lu W (2009) Simulation of laser beam induced current for HgCdTe photodiodes with leakage current, *Optical and Quantum Electronics*, 41: 805–810.

37

Quantum Well Infrared Photodetectors

Kwong-Kit Choi

37.1 Introduction

Quantum well infrared photodetectors (QWIPs) are based on optical transitions among quantized states in the conduction band of the quantum wells (QWs). They are widely used in long-wavelength infrared (LWIR) detection such as in gas sensing (Hinnrichs and Guptab, 2008) and in Earth observation from space (Jhabvala et al., 2011). The popularity of QWIPs is gained from the advantages in material quality and uniformity, detector sensitivity and stability, and array availability and cost. Another unique advantage of this technology is its predictable optoelectronic properties. In the absence of the extrinsic effects imposed by impurities and surfaces, the observed photocurrent and dark current of a QWIP are intrinsic to the detector structure and are well understood from fundamental physics. More specifically, its dark current is found to agree with the familiar thermionic emission (TE) process, while the photocurrent can be quantitatively predicted from first principle calculations on QW energy levels and electromagnetic (EM) field distributions inside the detector volume. With the known dark and photocurrents, the detection characteristics and performance of a QWIP array placed in the focal plane of an optical lens system can readily be predicted with very few system parameters. But because of the arbitrariness of the QW materials and the detector geometries, the associated physical equations can only be solved by numerical means. In this chapter, we describe the theoretical framework and numerical modeling of QWIPs.

The role of photocurrent I_p and dark current I_d in affecting the sensitivity of a QWIP focal plane array (FPA) has been analyzed (Choi et al., 1998). Here, it suffices to note that while the signal S of a detector is directly proportional to I_p, the noise N is directly proportional to $(I_p + I_d)^{1/2}$, which can be

derived from the Poisson statistics of charge collection. The signal-to-noise ratio of a QWIP is therefore proportional to

$$\left(\frac{S}{N}\right)_{\text{detector}} \propto \frac{I_p}{\sqrt{I_p + I_d}} = \left(\frac{I_p}{1 + I_d/I_p}\right)^{1/2} = \left(\frac{I_p}{1 + 1/r}\right)^{1/2}, \tag{37.1}$$

where r is the photocurrent to dark current ratio (I_p/I_d). From Equation 37.1, it is obvious that the S/N ratio is large only when I_p and r are large. In this chapter, we focus on the modeling of I_p and I_d in QWIP design and optimization.

37.2 Theoretical Model of QWIPs

A QWIP uses a multiple quantum well (MQW) structure to detect light (Levine et al., 1987). This structure is realized by growing alternate material layers of different bandgaps on a suitable substrate. The MQW creates a series of energy subbands, $E_1, E_2, \ldots$, in each QW unit. The QWs are doped n-type with a doping density N_D so that there are free electrons in the ground subband E_1 with a Fermi energy E_F. The band structure of a typical QWIP made of $Al_xGa_{1-x}As/In_yGa_{1-y}As$ is shown in Figure 37.1. In Figure 37.1, x is the aluminum mole fraction, y is the indium mole fraction, W is the well width, and B is the barrier thickness. An inter-subband absorption brings an electron from E_1 to E_n where the conductivity across the layers is higher. A photocurrent is thus created.

37.2.1 Dark Current

The dark current I_d of a QWIP can be qualitatively separated into three main components. They are direct tunneling (DI) between adjacent ground states, thermally assisted tunneling (TAT) near the top of the barrier, and TE over the barrier. These processes are indicated in Figure 37.2.

The TAT current originates from the finite thermal spreading of in-plane electron energy in E_1. Upon a scattering process, the in-plane energy combines with the out-of-plane energy, with which the electron tunnels more efficiently through the QW barrier. Therefore, the barrier transmission depends on the total energy rather than its energy components. Although each current component differs in details, such as the densities of initial and final states, the major factor in determining their magnitudes is the impedance of the QW barrier. By adopting an energy-dependent tunneling probability $\gamma(E, V)$ for the barrier, the total

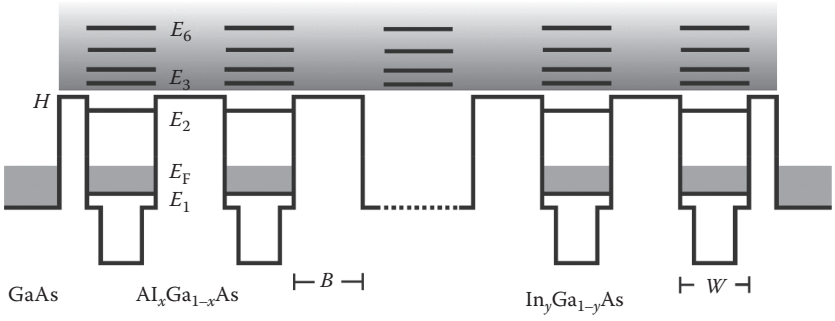

FIGURE 37.1 A typical energy diagram of an n-type QWIP along the z axis. E_1 to E_6 are the resonant QW states. The shaded region is the global conduction band.

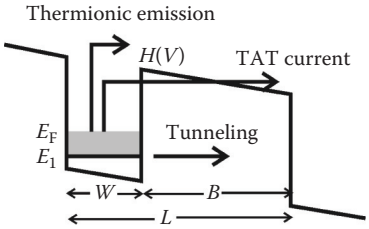

FIGURE 37.2 Different dark current transport processes in a QWIP. (TAT, thermally assisted tunneling.)

dark current can be obtained by summing over all electron energies (Levine et al., 1990):

$$\begin{aligned} J_{\mathrm{d}}(V,T) &= \int_{E_1}^{\infty} e\rho_{\mathrm{d}}(E,T)\, v_{\mathrm{d}}(V)\mathrm{d}E, \\ &= \int_{E_1}^{\infty} e\left[\frac{g_{2\mathrm{D}}}{L} f(E)\gamma(E,V)\right] v_{\mathrm{d}}(V)\mathrm{d}E, \\ &= \int_{E_1}^{\infty} e\frac{m^*}{\pi\hbar^2 L}\frac{\gamma(E,V)\, v_{\mathrm{d}}(V)}{1+\exp\left(\frac{E-E_F-E_1}{kT}\right)}\mathrm{d}E. \end{aligned} \tag{37.2}$$

In Equation 37.2, $J_{\mathrm{d}}(V,T)$ is the total dark current density under a bias of V per period at an operating temperature T, E the electron total energy, $\rho_{\mathrm{d}}(E)$ the three-dimensional (3D) electron density per electron-volt, $v_{\mathrm{d}}(V)$ the drift velocity, $g_{2\mathrm{D}}$ the 2D density of states, L the QW period length, $f(E)$ the Fermi–Dirac distribution, $\gamma(E,V)$ the tunneling probability of an electron with E under V, and m^* the electron effective mass in the well. The value of γ can be obtained from the usual Wentzel–Kramers–Brillouin (WKB) approximation (Choi, 1997) of a trapezoidal barrier with a barrier height H, a barrier thickness B, and an electron effective mass $m_{\mathrm{b}}{}^*$ in the barrier. Specifically,

$$\begin{aligned} \gamma(E,V) &= \exp\left\{\frac{-4B}{3e\hbar V}(2m_{\mathrm{b}}^*)\left[(H(V)-E)^{3/2}-(H(V)-eV-E)^{3/2}\right]\right\} \quad \text{for } E < H(V)-\mathrm{eV}\,, \\ &= \exp\left\{\frac{-4B}{3e\hbar V}(2m_{\mathrm{b}}^*)\,(H(V)-E)^{3/2}\right\} \quad \text{for } H(V) > E > H(V)-\mathrm{eV}, \\ &= 1 \text{ for } E > H(V), \end{aligned} \tag{37.3}$$

where H is generally a function of V because there is a potential drop across the well as depicted in Figure 37.2, which depends on the W/L ratio. The drift velocity was experimentally found to be linearly proportional to the electric field (V/L) with a low-field mobility μ initially and then becomes saturated at a constant velocity v_{sat} when the field is high. Therefore, $v_{\mathrm{d}}(V)$ can be represented as follows (Levine et al., 1990):

$$v_{\mathrm{d}}(V) = \frac{\mu V/L}{\sqrt{1+\left(\mu V/Lv_{\mathrm{sat}}\right)^2}}. \tag{37.4}$$

To fit the measured J_{d}, one can adjust H, μ, and v_{sat} in accounting for uncertainties in the material composition and quality.

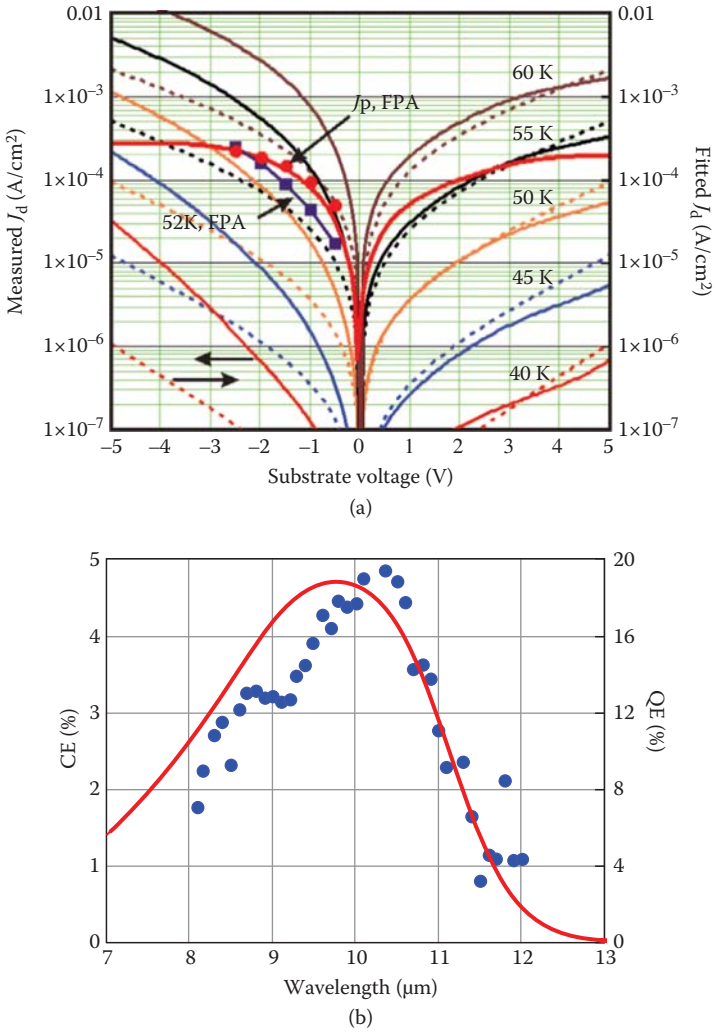

FIGURE 37.3 (a) The measured J_d (solid curves), fitted J_d (dashed curves), and measured J_p (thick solid curves) of a test detector are shown. The symbols are measured from a focal plane array (FPA). (b) The predicted and measured quantum efficiency (QE) and conversion efficiency (CE) of the FPA at −3 V based on angled sidewall optical coupling are shown.

Figure 37.3 shows the typical current–voltage (I–V) characteristics of a QWIP. The substrate voltage refers to the bias applied on the substrate while the top contact is grounded. They are measured from a QWIP with a cutoff wavelength $\lambda_c = 11.2$ μm (Choi et al., 2011). The QWIP is made up of 60 periods of 700 Å $Al_{0.19}Ga_{0.81}As$ and 50 Å GaAs. The wells are doped with N_D of 1.7×10^{18} cm^{-3}. By adopting a constant $H = 191$ meV, $\mu = 400$ cm^2/V · s, and $v_{sat} = 1 \times 10^7$ cm/s, the dark current I–V characteristics under positive bias can be explained in a large range of temperatures. In this example, H is independent of V because the W/L ratio (= 50/750 = 0.067) is very small. The adopted H is consistent with the expected band offset from the barrier composition, and the values of μ and v_{sat} are nearly the same for all experimental LWIR detectors. This consistency shows that Equation 37.2 is a satisfactory dark current model for QWIPs. According to the symmetrical band structure shown in Figure 37.1, the I–V characteristics of a QWIP should be symmetric with respect to bias polarity. The higher measured dark current shown in Figure 37.3

under the negative bias is attributed to dopant migration during the material growth, which lowers the barrier height in the barrier region immediately after the doping.

37.2.2 Photoconductive Gain

The photocurrent I_p of a QWIP can be expressed as $I_p = A\rho_p e v_d$, where A is the detector area, ρ_p the photoelectron density, and v_d the electron drift velocity. The parameter ρ_p can be further expressed as $G\tau/(A\ell)$, where G is the total photoelectron generation rate in the entire detector volume $A\ell$, ℓ the detector total thickness, and τ the average recombination lifetime of a photoelectron. After photoexcitation, a photoelectron only retains its energy in a period τ before it recombines into one of the QWs. Therefore, $G\tau/(A\ell)$ represents the time averaged photoelectron density. If we define the quantum efficiency (QE) η as the fraction of incident photon flux Φ being absorbed by the detector and converted into photoelectrons, the generation rate G will be $\eta\Phi$. Hence, I_p is given by

$$\begin{aligned} I_p &= A\rho_p e v_d, \\ &= A\left(\frac{G\tau}{A\ell}\right) e \left(\frac{\ell}{\tau_{tr}}\right) = \left(\frac{\eta\Phi\tau}{\ell}\right) e \left(\frac{\ell}{\tau_{tr}}\right) = \eta\Phi\left(\frac{\tau}{\tau_{tr}}\right) e, \\ &= \eta\Phi g e, \end{aligned} \tag{37.5}$$

where τ_{tr} is the time taken by a photoelectron to travel across the entire detector thickness ℓ, and $g = \tau/\tau_{tr} = \tau v_d/\ell$ is defined as the photoconductive gain. Equation 37.5 indicates that the measured number of photoelectrons (I_p/e) is g times the number of photons absorbed ($\eta\Phi$), resulted in a gain. Since the lifetime τ of a photoelectron is usually shorter than the transit time τ_{tr} in a QWIP, g is usually less than unity. Its photocurrent will then be lower than that of a photodiode with a gain of unity for the same QE. The value of g is usually deduced from noise measurements and theoretically fitted to the same functional form of v_d with τ as a fitting parameter. Experimentally, we found $\tau = 25$ ps for most LWIR QWIPs.

37.2.3 Photocurrent

Expanding the last line of Equation 37.5, the photocurrent is given by

$$\begin{aligned} I_p(V, T_B) &= eg(V)\int_{\lambda_1}^{\lambda_2} \eta(V,\lambda)\Phi(T_B,\lambda)\mathrm{d}\lambda, \\ &= eg(V)\int_{\lambda_1}^{\lambda_2} \eta(V,\lambda)\Omega A L(T_B,\lambda)\mathrm{d}\lambda, \\ &= \frac{\pi}{4F^2+1} A e g(V)\int_{\lambda_1}^{\lambda_2} \eta(V,\lambda) L(T_B,\lambda)\mathrm{d}\lambda. \end{aligned} \tag{37.6}$$

In Equation 37.6, Ω is the solid angle of the light cone sustained at the detector toward the lens. $L(T_B, \lambda)$ is the photon spectral radiance of the scene, which has a unit of number of photons arrived per unit time per unit detector area per unit solid angle per unit wavelength. Expressing Ω in terms of the usual specification for a lens system, the f-number F, which is the ratio between the focal length f and the lens diameter D, $\Omega = \pi \sin^2(\theta) = \pi(D/2)^2/[f^2 + (D/2)^2] = \pi/(4F^2 + 1)$, where θ is the half-cone angle.

The parameter $g(V)$ is the photoconductive gain at V, λ_1 and λ_2 are the lower and upper wavelengths encompassing the detector absorption spectrum, and $\eta(V,\lambda)$ is the detection QE spectrum. $\eta(V,\lambda)$ can be decomposed into $\eta_{abs}(\lambda)\gamma(V)$, where η_{abs} is the fraction of incident photons being absorbed and $\gamma(V)$ is the tunneling probability of the photoelectron out of the barrier. For the typical QWIP designs, the upper subbands are above the barrier and thus $\gamma(V)$ is close to unity. Therefore, we usually assume $\eta(V,\lambda) \approx \eta_{abs}(\lambda)$, and label $\eta_{abs}(\lambda)$ simply as $\eta(\lambda)$ unless the effect of γ is apparent.

37.2.4 Absorption Coefficient

The magnitude of the photocurrent I_p in Equation 37.6 depends on the QE spectrum $\eta(\lambda)$. In QWIPs, the optical transitions are generated from the transition dipole moment between two envelope eigenfunctions in the z-direction. The optical electric field E of the radiation must be oriented in this direction to drive the transitions. When the optical field is polarized in the x–y plane, the dipole moment is zero due to conservation of momentum, and therefore, there is no absorption. Therefore, the optical absorption of the material depends on the direction of light propagation, and the absorption coefficient α is anisotropic. The parameter α is defined as $(-1/\Phi)(d\Phi/dx)$, where Φ is the photon flux and x is the traveling distance. In general, when the light is traveling at an angle θ relative to the normal of the layer surface, only the vertical electric component $E_z = E\cdot \sin(\theta)$ in the transverse magnetic (TM) mode is able to initiate the transitions. The transverse electric mode, on the other hand, is totally uncoupled. Therefore, for an unpolarized light, only half of its intensity is being absorbed, and α of the remaining half varies as $\sin^2\theta$. When θ is zero under the usual detection orientation, i.e., under normal incidence, α will be zero unless there is an optical coupling structure to convert the incident electric polarization from horizontal to vertical. Overall, $\eta(\lambda)$ of a QWIP depends on both the material intrinsic absorption properties and the detector physical structure.

For a parallel propagating light with vertical polarization, the absorption coefficient $\alpha(\hbar\omega)$ depends on the oscillator strength between the ground state and one of the upper states and the number of free carriers in the ground state. Quantitatively, $\alpha(\hbar\omega)$ is given by (Choi, 1997)

$$\alpha(\hbar\omega) = \frac{N_s}{L}\frac{\pi e^2 \hbar}{2m^*\sqrt{\varepsilon_r}\,\varepsilon_0 c}\sum_{n=2}^{\infty} f_n \rho_n(\hbar\omega + E_1), \tag{37.7}$$

where $N_s = N_D W$ is the 2D doping density per QW period, L the length of a QW period, m^* the effective electron mass of individual layers, ε_r the real part of the dielectric constant, f_n the oscillator strength between E_1 and E_n, and ρ_n the normalized density of states due to the presence of line broadening mechanisms. The oscillator strength f_n is given by

$$f_n = \frac{2\hbar^2}{m^*\left(E_n - E_1\right)}\left|\langle \psi_n \left| \frac{\partial}{\partial z} \right| \psi_1 \rangle\right|^2, \tag{37.8}$$

where ψ_n is the eigenfunction of E_n. The overall absorption lineshape depends mostly on the energy distribution of E_n and the respective f_n with a weaker dependence on the individual ρ_n. We assume a Gaussian shape for ρ_n, i.e.,

$$\rho_n(E) = \frac{1}{\sqrt{2\pi}\sigma_n}\exp\left(-\frac{\left(E - E_n\right)^2}{2\sigma_n^2}\right), \tag{37.9}$$

where σ_n is the line broadening parameter. In the next section, we show modeling procedures for $\alpha(\hbar\omega)$.

37.3 Material Absorption Modeling

The eigenfunction ψ_n and eigenvalue E_n of an MQW can be obtained from the transfer matrix method (TMM) (Choi et al., 2002). In this method, ψ_j in the jth material layer is written as two counter-propagating plane wave states:

$$\psi_j = A_j\, e^{i k_j z} + B_j\, e^{-i k_j z}, \tag{37.10}$$

where A_j and B_j are two constants to be determined, and k_j is the wavevector given by

$$k_j(E) = \frac{\sqrt{2\, m_j^*(E)}}{\hbar}\sqrt{E - \Delta E_j}, \tag{37.11}$$

where $m_j^*(E)$ is the energy-dependent effective mass, E and ΔE_j are the electron energy and the conduction band offset measured from the GaAs conduction band edge, respectively. The constants A_j and B_j relate to A_{j+1} and B_{j+1} of the adjacent $(j+1)$th layer through the boundary conditions that both ψ and $(1/m^*)\mathrm{d}\psi/\mathrm{d}z$ are continuous across the interface:

$$\begin{aligned}\begin{bmatrix} A_j \\ B_j \end{bmatrix} &= \frac{1}{2}\begin{bmatrix} (1+\gamma_{j,j+1})\, e^{i(k_{j+1}-k_j)\, d_{j,j+1}} & (1-\gamma_{j,j+1})\, e^{-i(k_{j+1}+k_j)\, d_{j,j+1}} \\ (1-\gamma_{j,j+1})\, e^{i(k_{j+1}+k_j)\, d_{j,j+1}} & (1+\gamma_{j,j+1})\, e^{-i(k_{j+1}-k_j)\, d_{j,j+1}} \end{bmatrix}\begin{bmatrix} A_{j+1} \\ B_{j+1} \end{bmatrix} \\ &= \frac{1}{2}\, M_{j,j+1}\begin{bmatrix} A_{j+1} \\ B_{j+1} \end{bmatrix},\end{aligned} \tag{37.12}$$

where $\gamma_{j,j+1} = (m_j{}^* k_{j+1})/(m_{j+1}{}^* k_j)$ and $d_{j,j+1}$ is the spatial coordinate of the interface. The energy level structure of a multilayer material system can be obtained from the maxima of the global transmission coefficient T_G of a plane wave through the material. When the incident energy of the wave coincides with one of E_n's, strong resonant transmission occurs and T_G attains a local maximum. Assuming the plane wave incident from the left, the coefficients A_1 and B_1 for the first material layer on the left and the coefficients A_p and B_p of the last pth layer are connected by

$$\begin{bmatrix} A_1 \\ B_1 \end{bmatrix} = \frac{1}{2^{p-1}}\, M_{1,p}\begin{bmatrix} A_p \\ B_p \end{bmatrix}, \tag{37.13}$$

where $M_{1,p} = M_{1,2}M_{2,3}\ldots M_{p-1,p}$, and $M_{j,j+1}$ is defined in Equation 37.12. By assuming that there is no wave traveling to the left in the last period, one can set $A_p = 1$ and $B_p = 0$, with which all the coefficients in the layers are determined. The value of T_G at a given energy E is then equal to

$$T_G(E) = \frac{1}{\left|A_1(E)\right|^2}\frac{v_p(E)}{v_1(E)} = \frac{2^{2p-2}}{\left|a_{11}(E)\right|^2}\frac{m_1^*(E)}{m_p^*(E)}\frac{k_p(E)}{k_1(E)}, \tag{37.14}$$

where v is the electron group velocity and a_{11} the first diagonal element of $M_{1,p}$. The local maxima in $T_G(E)$ determine the locations of E_n. In general, these local maxima are less than unity unless the material structure is symmetrical. Calculating $A_n(E)$ and $B_n(E)$ at $E = E_n$ will yield ψ_n.

37.3.1 Material Parameters

The TMM calculation can be handled efficiently by a personal computer. The material parameters required for the InGaAs/AlGaAs material system had been collected and summarized (Shi and Goldys, 1999).

To begin with, the unstrained bandgap of the ith layer in the unit of meV is

$$E^u_{g,j} = 1424 + 1186\,x_j + 370\,x_j^2 - 1499\,y_j + 429\,y_j^2, \tag{37.15}$$

where x_j and y_j are the Al and In molar ratios in the jth layer, respectively. (In the present calculation, x and y both cannot be nonzero in a given layer, i.e., AlInGaAs quaternary compound is not considered.) Since the material layers are under strain, the values of bandgaps are modified. Assuming the strain is accommodated exclusively in the InGaAs well layers, which are usually much thinner than the GaAs/AlGaAs layers, the modified bandgap for $y_j > 0$ is

$$E^s_{g,j} = E^u_{g,j} + 2\,a\left(1 - \frac{c_{12}}{c_{11}}\right)\varepsilon_{11}, \tag{37.16}$$

where $a = -8700 + 2700y_j$ meV is the deformation potential, $c_{12}/c_{11} = (5.32 - 0.79y_j)/(11.81 - 3.48y_j)$ is the ratio of the elastic constants, and $\varepsilon_{11} = -0.405y_j/(5.6533 - 0.405y_j)$ is the hydrostatic component of the strain induced by the adjacent GaAs/AlGaAs material layers. Therefore, in the present assumption, the bandgap of InGaAs is widened while that of GaAs and AlGaAs are unchanged. The band offset ΔE_j is taken as

$$\begin{aligned} \Delta E_j &= \beta_x\,(E^u_{g,j} - 1424) \quad \text{if} \quad x_j > 0 \\ &= \beta_y\,(E^s_{g,j} - 1424) \quad \text{if} \quad y_j > 0, \end{aligned} \tag{37.17}$$

where β_x is the band offset parameter for the GaAs/AlGaAs interface and β_y is for the InGaAs/GaAs interface. From fitting detectors with different wavelengths, we notice that β_x increases from 0.60 to 0.77 when x decreases from 0.4 to 0.16, and β_y is 0.63 when $y < 0.34$. The effective mass at the conduction band edge is given by

$$m^*_j = 0.0665\,(1 + 1.256\,x_j - 0.579\,y_j). \tag{37.18}$$

Taking the band nonparabolicity into account, the effective mass $m_j^*(E)$ can be expressed in the form of

$$m^*_j(E) = m^*_j\left(1 + \frac{E - \Delta E_j}{E_{g,j}}\right), \tag{37.19}$$

where $E_{g,j}$ is the strain modified bandgap. With all these material parameters, the input parameters in Equation 37.14 are known. Since m^* varies from layer to layer, strictly speaking, Equations 37.7 and 37.8 need to be evaluated individually in each layer. For simplicity, we assume m^* to be the effective mass at the GaAs conduction edge since optical transitions mostly occur in the GaAs layers.

37.3.2 Numerical Examples

In this section, we present some examples and compare them to experiments. To measure the α lineshape, infrared (IR) light is incident on a 45° polished facet at the edge of a substrate (referred to as the edge coupling method), so that the optical electric field pointing at the z-direction is a factor of $1/\sqrt{2}$ of that of the TM component. The measured quantity is the responsivity (R) or the conversion efficiency ($CE \equiv \eta g$). The values of R and CE are related by

$$R(\lambda) = \frac{I_p}{P(\lambda)} = \frac{e\eta g\Phi(\lambda)}{\frac{hc}{\lambda}\Phi(\lambda)} = \frac{\lambda}{\frac{hc}{e}}\eta g = \frac{\lambda}{1.24\times 10^{-6}}CE(\lambda) \quad \left[\frac{\text{A}}{\text{W}}\right], \tag{37.20}$$

where $P(\lambda)$ is the incident optical power. Since in the weak absorption limit, η is directly proportional to α (see Section 37.4.4 for details on this coupling scheme), we can compare either the experimental $CE(\lambda)$ lineshape to the theoretical $\alpha(\lambda)$ or $R(\lambda)$ lineshape to $\lambda\alpha(\lambda)$. In Figure 37.4, we show the calculated $f_n(\lambda)$ of a typical QWIP, labeled as Det. A, which is made of 21 periods of 500 Å $Al_{0.27}Ga_{0.73}As$/46 Å GaAs QWs. The TMM calculation was performed on a single QW unit. The result will be the same if the calculation is performed on the entire MQW when E_1 of individual wells are not coupled (Choi et al., 2002), and in this case, the total f is independent of B.

Substituting the f spectrum into Equation 37.7, $\alpha(\lambda)$ can be obtained. Figure 37.5 shows the calculated $\alpha(\lambda)$ for Det. A along with three other GaAs/AlGaAs detectors having different cutoff wavelength λ_c and doping density N_D. Figure 37.5 also shows the measured CE lineshape of each detector in arbitrary units. To fit the measurements, the values of β_x needs to vary from 0.665 to 0.715 for the nominal x from 0.27

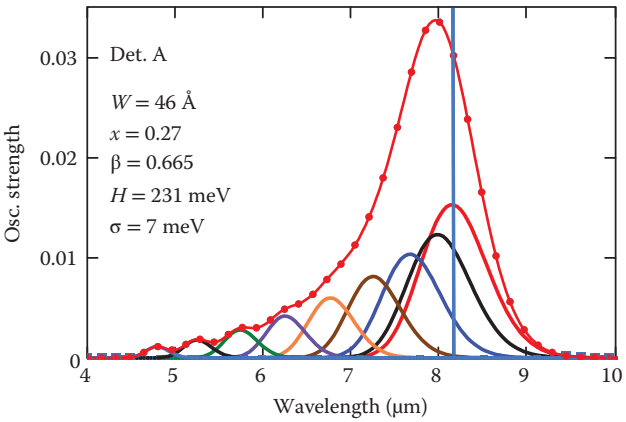

FIGURE 37.4 The calculated oscillator strengths f_n of a QWIP with $W = 46$ Å, $B = 500$ Å, $x = 0.27$, and $\sigma = 7$ meV. The curve with circles is the total oscillator strength. The vertical line shows λ at which the final state energy $= H - E_1$.

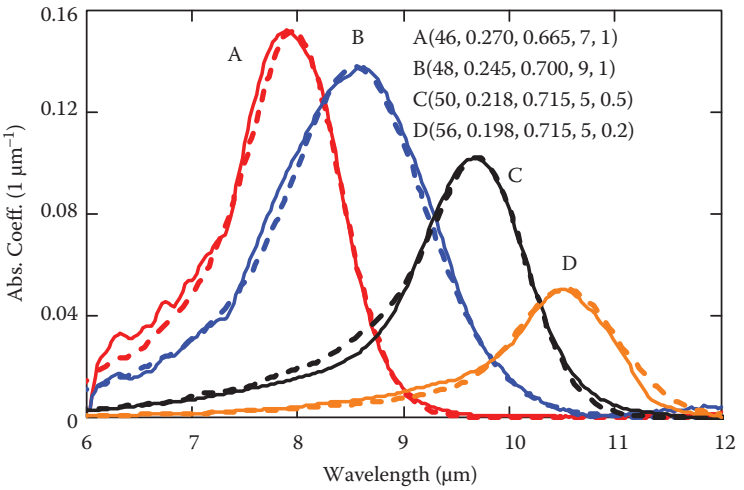

FIGURE 37.5 The measured CE lineshapes (solid curves) from four materials. The dashed curves are the calculated α assuming $(W, x, \beta_x, \sigma, N_D)$ shown in the figure, where N_D is in 10^{18} cm^{-3}. The barrier thickness is 500 Å for all detectors except 600 Å for Det. D.

to 0.198. The broadening parameter σ ranges from 5 to 9 meV for the first upper state, and it increases by about 0.5 meV for each successive upper state. With the two fitting parameters, β_x and σ, the intrinsic absorption lineshape can be satisfactorily explained.

The bandwidths shown in Figure 37.5 are rather limited. To broaden the bandwidth, one approach is to use superlattices (SLs) to create miniband-to-miniband (M–M) transitions. When the barrier thickness B is small, E_n states among different wells are coupled to form minibands M_n. Since the degenerate E_1 states develop into a range of occupied states in M_1, the absorption can cover a wider spectral range. This bandwidth can further be expanded if the QWs in the SL have different well widths (Choi et al., 2002).

To cover the entire 8- to 14-μm LWIR window, we designed a structure (Li et al., 2006) that contains seven SLs, and they are separated by 600-Å $Al_{0.17}Ga_{0.83}As$ barriers. Each SL contains seven 30-Å $Al_{0.23}Ga_{0.77}As$ barriers and four 70-Å and four 75-Å wells alternately placed together. The doping is 4×10^{17} cm^{-3} in the wells while the barriers are undoped. Since the eight miniband ground states within M_1 have different energies, their electron populations are also different. In this example, the first state in M_1 is at 39.4 meV and the last state is at 55.7 meV, while the common Fermi energy for the present doping is at 57.8 meV. Therefore, the electron population in the first state is 8.8 times larger than the last state, resulting in a significant modification of the lineshape from that of MQWs. Calculating Equation 37.7 with a different effective 3D doping density (N_s/L) for each miniband ground state according to the different Fermi levels, the spectral lineshape of the SL can be obtained. Figure 37.6 shows the calculated lineshape for the present SL. It agrees well with the measurement using a constant σ of 7 meV. The spectral responsivity is seen to cover the entire LWIR window. Following this material design approach, a QWIP with any cut-on and cutoff wavelengths can be designed.

By using InGaAs/AlGaAs QWs to increase the combined barrier height, a mid-wavelength IR QWIP can be developed (Choi et al., 2002). The illustrated structure contains eight $In_{0.34}Ga_{0.66}As/Al_{0.4}Ga_{0.6}As$ SL units separated by 500-Å $Al_{0.4}Ga_{0.6}As$ blocking barriers. Each unit consists of five wells and four barriers. The wells are made of 5-Å GaAs/25 Å InGaAs/5Å GaAs layers, and the barriers are made of 30-Å AlGaAs layers. There are top and bottom GaAs contact layers. All SL layers are doped to 2×10^{18} cm^{-3}. The modeled responsivity curve matches the experiment if $\sigma = 27.5$ meV as shown in Figure 37.7. This much larger σ is usually observed in the MWIR regime.

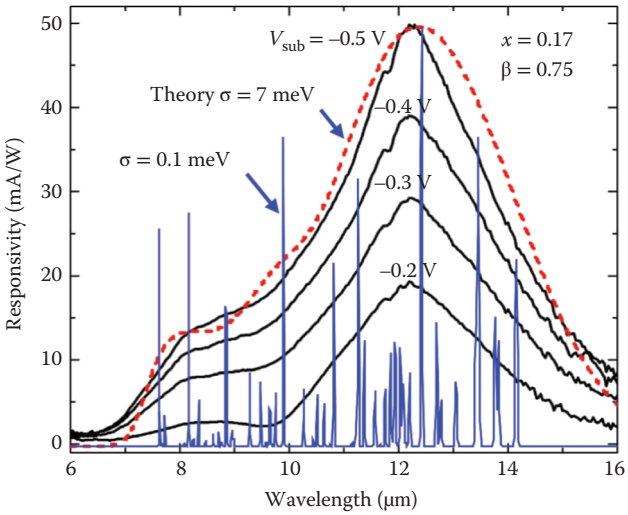

FIGURE 37.6 The measured responsivity from an SL structure with alternate well widths (solid curves). The fitted spectral lineshape of λα(λ) with σ = 7 meV is shown as the dashed curve. The curve with σ = 0.1 meV is to reveal the spectral distribution of individual M-M transitions.

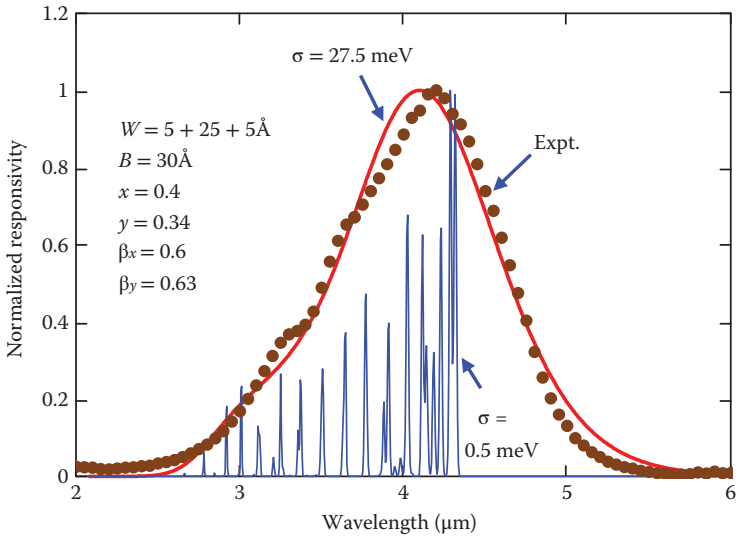

FIGURE 37.7 The measured normalized responsivity (circles) and the calculated $\lambda\alpha(\lambda)$ with $\sigma = 27.5$ meV (solid curve) for the combined transitions in the SL with blocking barriers. The curve with $\sigma = 0.5$ meV is to reveal the spectral distribution of individual M-M transitions.

37.4 Detector QE Modeling

Although $\alpha(\lambda)$ for parallel propagating light can be modeled accurately, the detector QE under normal incidence is still far from definite because it also depends on the optical coupling scheme. The search for an optimum coupling structure for a given application occupies the bulk of QWIP research to date. To yield the theoretical QE, one has to know the EM field distribution inside the detector volume in the presence of a polarization-dependent absorption coefficient. This task can only be solved by employing 3D numerical EM modeling. Since this capability was not readily available in the past, the common QWIP structures are not optimized.

With the advent of computer hardware and software technologies, one can now perform realistic and sophisticated EM modeling using commercial software on a personal computer. These advances have greatly facilitated the QWIP development. There are many numerical methods in solving EM problems. The more common ones are finite element method (FEM), finite-difference time-domain method, finite boundary method, and modal transmission-line (MTL) method, etc. In this chapter, we describe using FEM to model the theoretical QE of QWIPs.

FEM solves Maxwell's equation,

$$\nabla \times \left(\mu_r^{-1} \nabla \times \boldsymbol{E}\right) - k_0^2 \varepsilon_{rc} \boldsymbol{E} = 0, \tag{37.21}$$

via trial basis functions under a user-defined boundary condition. In Equation 37.21, μ_r and $\varepsilon_{rc} = \varepsilon_r - j\varepsilon_i$ are the relative permeability and relative permittivity, respectively, and $k_0 = 2\pi/\lambda$ is the free-space wavevector. Through variational principles, FEM transforms Equation 37.21 into a large set of linear equations, which can be solved by matrix multiplications (Sadiku, 2001). To yield a valid solution, it is important to present a correct boundary condition in the model.

The present modeling is performed in the RF Module of a commercial EM solver, COMSOL Multiphysics version 3.5 (COMSOL AB, Nov 2008). The modeling procedures involve selecting the EM analysis mode, building the 2D or 3D detector geometry, defining constants, variables, and functions, inputting

subdomain properties, selecting appropriate boundary conditions, building mesh structures, setting solver parameters, performing computation, and using postprocessing to yield the required information.

To model the QE of a QWIP, we note that η can be expressed as (Choi et al., 2012, 2013a)

$$\eta(\lambda) = \frac{1}{P_0}\int_V \mathrm{d}I(\mathbf{r}) = \frac{1}{P_0}\int_V \alpha(\lambda) I(\mathbf{r},\lambda) d^3r$$
$$= \frac{\alpha(\lambda)}{A\frac{c\varepsilon_0}{2}E_0^2}\int_V \frac{n(\lambda)c\varepsilon_0}{2}\left|E_z(\mathbf{r},\lambda)\right|^2 d^3r = \frac{n(\lambda)\alpha(\lambda)}{AE_0^2}\int_V \left|E_z(\mathbf{r})\right|^2 d^3r, \tag{37.22}$$

where P_0 is the incident optical power, V the detector active volume, I the optical intensity associated with E_z, α the absorption coefficient, A the detector area, and n the material refractive index. Equation 37.22 states that the QE of a QWIP can be calculated from the volume integral of $|E_z|^2$ in the presence of a finite α. If E_z in Equation 37.22 is replaced by the total E, it will be applicable to detectors with isotropic absorption as well. Since E_0 and E_z are linearly proportional to each other, E_0 can be set arbitrarily. Besides E_0, the only free input parameter in Equation 37.22 is the wavelength-dependent α(λ) (if $n(\lambda)$ of the material is known accurately). For a known α(λ) from Equation 37.7, there will be no free parameters, and the value of η(λ) for a given combination of material and geometry can be uniquely and unambiguously determined.

37.4.1 Modeling Validation and Phase Coherence

Since numerical EM modeling involves many steps, assumptions, and idealizations of materials and structures, it is critical to validate the model through theoretical and experimental means. One important assumption in the model is the complete phase coherence of radiation. While it is true for radio waves, it is generally not the case in the IR wavelengths and the rationale for EM modeling needs further explanation. If the optical field distribution in a detector indeed depends on the degree of coherence of light, the present modeling clearly will not be relevant. It follows that a QWIP will not have a definite QE, and its value depends on the light source characteristics. For a blackbody source, it is well known that the radiation is incoherent and its coherence length ℓ_c is only about $0.8\lambda_{max}$, where λ_{max} is the free-space wavelength at the intensity maximum (Donges, 1998). At a blackbody temperature $T_B = 300$ K, λ_{max} is 10 μm, and ℓ_c is 8 μm in free space and 2.7 μm in GaAs. Such a short ℓ_c would suggest phase coherence exists only in a very small detector volume and coherent EM modeling would not be meaningful.

The abovementioned picture is certainly true when considering all of the radiations from the blackbody. Consistent with this picture, it also follows that the modeled coherent QE at λ_{max} will not be valid for all the blackbody radiations, which is also certainly the case. Nonetheless, when all the incoherent wave packets from the blackbody are Fourier transformed into the plane wave components, which happen to assume the Planck distribution, each individual wavelength component is in fact phase coherent. (By the definition of a plane wave with a definite λ, its coherence length is infinite.) Since the detector absorbs one photon with a definite λ at a time, not a wave packet with a broad spectrum at a time, the detector is interacting with each Fourier component individually, and because of this specific absorption feature, the spectrally dispersed radiation appears to be completely coherent to the detector at the individual wavelength component basis. Following this reasoning, there will only be one QE value for each λ, and it is the value evaluated under phase coherent condition. The modeled QE is thus an intrinsic property of the detector independent of the light source. Consequently, the detector maintains the same optical properties when detecting objects with different coherence lengths caused by, for example, different scene temperatures.

For experimental verification, when one of the Fourier components is separated out from a blackbody by an instrument, either using a grating or a filter, its coherence length $\ell_c = \lambda^2/\Delta\lambda$ becomes finite, where Δλ is bandwidth of the instrument (Fowles, 1975). It is more convenient to use a light source with a long ℓ_c, i.e., a small Δλ, so that no QE convolution is need. However, even if Δλ is substantial in comparison

with the variation of modeled $\eta(\lambda)$, one can still make comparison by characterizing the spectral power density $P(\lambda)$ of the source around the nominal incident wavelength λ_0. The weighted theoretical $\eta(\lambda_0) = \int \eta(\lambda - \lambda_0)P(\lambda - \lambda_0)d\lambda / \int P(\lambda - \lambda_0)d\lambda$ can then be compared to experiment.

It is interesting to mention that when $\Delta\lambda$ of the instrument is broader than the detector absorption bandwidth Γ, the detector can only sense part of each wave packet emitted by the instrument, with which the coherence length appears to be longer at λ^2/Γ instead of $\lambda^2/\Delta\lambda$ for the original wave packet. Therefore, a detector with a given Γ will not be able to perceive the true phase coherence of a source if the source ℓ_c is shorter than λ^2/Γ. In the following, we present examples of modeling validation.

37.4.2 Planar Solar Cell

A planar GaAs solar cell provides a simple test for the modeling since it lacks any x–y spatial variations. Its material structure includes an active GaAs absorbing layer, a front antireflection coating, and a back metal reflector as shown in Figure 37.8 (Grandidier et al., 2012). With the known complex refractive indices of GaAs, and assuming 100% transmission for the antireflection (AR)-coating and 100% reflection for the reflector, a simple classical model for QE using ray-tracing technique is represented by the dash curve in Figure 37.9. On the other hand, using the complex refractive indices from all the constituent materials, the 2D modeling based on Equation 37.22 with isotropic absorption yields the solid curve. It coincides

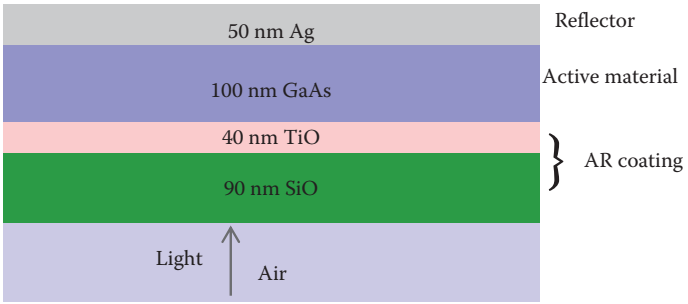

FIGURE 37.8 The modeled layer structure of a planar solar cell.

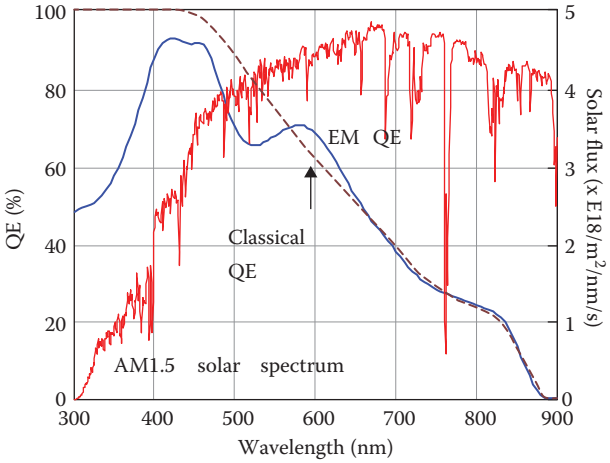

FIGURE 37.9 The quantum efficiency (QE) obtained by raytracing (dashed curve) and FEM (solid curve). The figure also shows the solar flux under 1-sun. (EM, electromagnetic.)

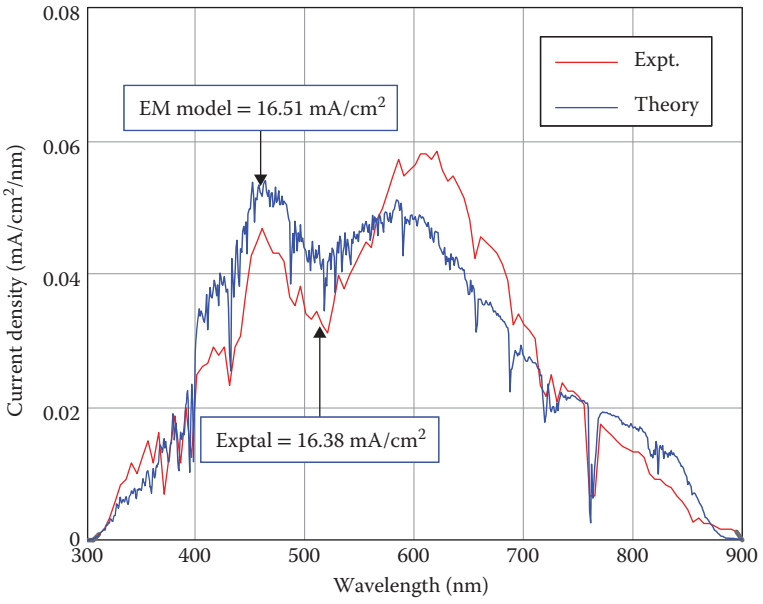

FIGURE 37.10 The measured and modeled photocurrent density under 1-sun without fitting parameters. (EM, electromagnetic.)

with the classical solution when the coatings in the modeling are effective. At the same time, the predicted photocurrent spectrum in Figure 37.10 is very similar to the measured spectrum (Grandidier et al., 2012) without any fitting parameters. In particular, the model accurately predicts two photocurrent peaks at 430 and 600 nm, which are absent in the simplified classical model. Furthermore, the model predicts the integrated photocurrent to be 16.51 mA/cm^2 under 1-sun condition, which is in excellent agreement with the measured 16.38 mA/cm^2. Therefore, without knowing the actual material properties in the experimental study, the agreement between theory and experiment is satisfactory for this device structure.

37.4.3 InAs/GaInSb IR Photodiode

When the detector material is fabricated into detector pixels, it acquires spatial variations. In this example, we show an IR photodiode in Figure 37.11 with a linear dimension of 22 μm in a 25-μm pitch array. It consists of a 2-μm thick p-InAs/$Ga_{0.75}In_{0.25}Sb$ strained SL (p-SLS2) layer and a 0.5-μm thick wider bandgap n-SLS1 layer to form a photodiode. The photodiode is sandwiched between a 0.5-μm thick wider bandgap p-contact layer (p-SLS3) and a 2-μm thick n-$InAs_{0.9}Sb_{0.1}$ contact layer. On top of the p-contact, there is a 12-μm wide gold layer for ohmic contact. Light is incident from the n-contact layer. The detector is assumed to be surrounded by a semi-infinite epoxy material with $n = 1.5$ and with negligible IR absorption.

Since there is no AR-coating in this detector, Fabry–Pérot (FP) oscillations arise from optical interference between the top and bottom material interfaces. Because of the inhomogeneous phase change along the top surface, the FP oscillations are irregular. In Figure 37.12, we show the theoretical α spectrum of the p-SLS2 layer and the deduced α spectrum of n-$InAs_{0.9}Sb_{0.1}$ contact layer from open literature. The latter spectrum is obtained by shifting the experimental n-InAs spectrum by $\lambda = 0.7$ μm in accounting for the bandgap reduction (Miles et al., 1990). We further assume α of the n-SLS1 and p-SLS3 to be the same as that of p-SLS2 but at the shorter wavelengths (Dixon et al, 1961). The $n(\lambda)$ value in Equation 37.22 is taken to be a constant of 3.5, which is the average between 3.4 for InAs and 3.7 for GaSb at 10 μm. All IR absorption is assumed to be isotropic. Since the absorption of all the layers except the p-SLS2 layer will be dissipated as heat and no photocurrents will be generated in these layers, their absorption reduces the QE of the p-SLS2 layer in IR detection.

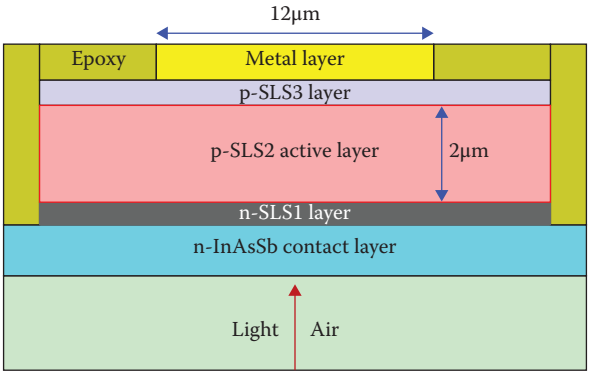

FIGURE 37.11 The material layer structure of a SLS photodiode. The pixel pitch is 25 μm.

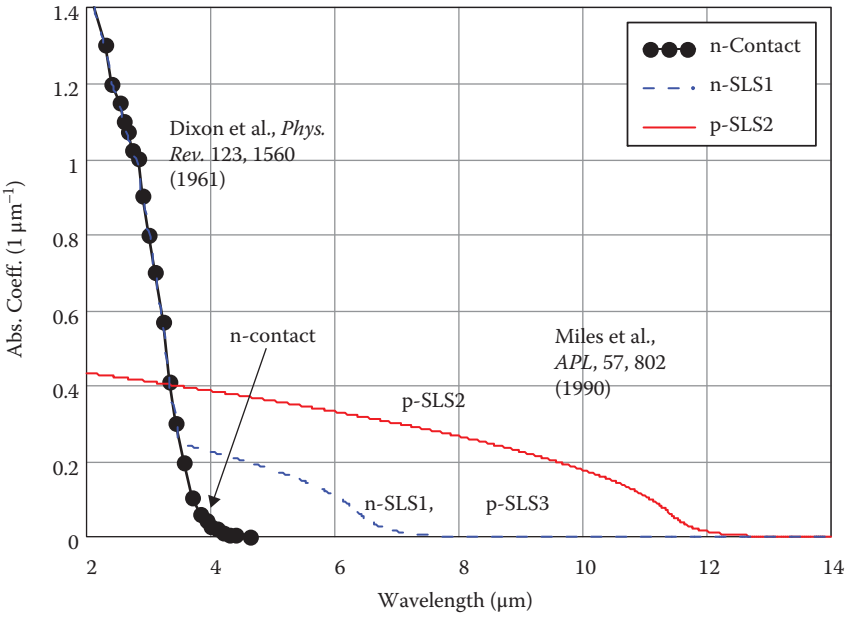

FIGURE 37.12 The theoretical and experimental α of individual layers obtained from open literature.

With these material and detector structures, the detector QE spectrum is modeled and the result is shown in Figure 37.13. Without adjusting any input parameters, the theoretical prediction matches satisfactorily with the experiment both in magnitude and in spectral lineshape in the entire spectral range. This example shows the validity of the modeling in the presence of spatial variations.

37.4.4 Edge Coupling for QWIPs

In Section 37.3.2, we assumed optical coupling efficiency using a 45° facet to be λ-independent in order to employ this coupling scheme for material characterization. It is worth checking its validity (Choi et al., 2012). Figure 37.14 shows the detector geometry and the calculated E fields with $E_0 = 377$ V/m. The detector has a 3-μm thick active QWIP layer, a 1-μm top contact layer, and a 1-μm gold layer. The detector is 100-μm wide. The value of α is assumed to be 0.15 μm^{-1}, and it is λ independent.

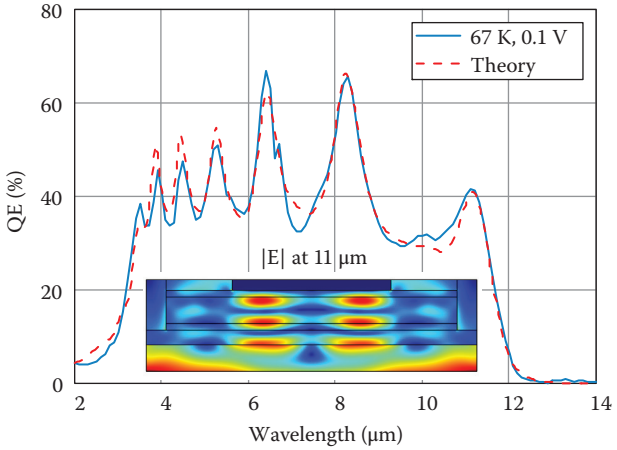

FIGURE 37.13 The calculated and the measured quantum efficiency (QE) of the p-SLS2 layer. The experimental data are provided by M. Jhabvala at NASA Goddard Space Flight Center. The insert shows the color plot of the $|E|$ distribution at $\lambda = 11$ μm.

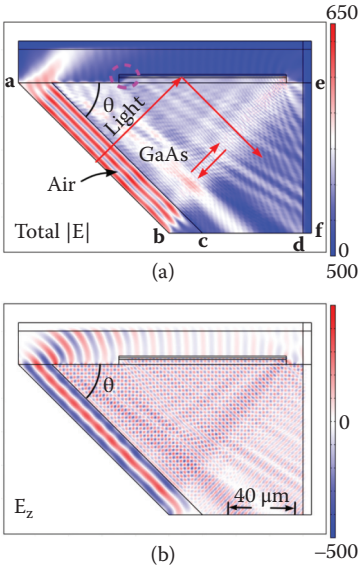

FIGURE 37.14 (a) The detector geometry at wedge angle θ of 45°. The arrows show the typical light paths. (b) The corresponding E_z is shown.

The 2D modeling with anisotropic absorption shows that QE at $\theta = 45°$ as well as other angles are nearly independent of λ as shown in Figure 37.15, justifying the experimental approach. The small oscillations are due to FP oscillations established in one of the detector corners shown in Figure 37.14a, but they will not be noticeable in the presence of edge roughness. In Figure 37.15, we also show the classical QE calculated from

$$\eta_c(\theta) = \frac{4n}{(1+n)^2}\frac{\cos\theta}{2}\left[1 - \exp\left(-\alpha \sin^2\theta \cdot \frac{2\ell}{\cos\theta}\right)\right], \tag{37.23}$$

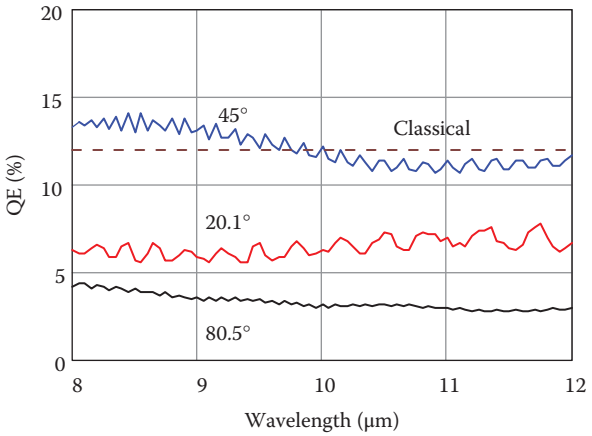

FIGURE 37.15 The calculated quantum efficiencies (QEs) at three different wedge angles. The dashed lines show the classical value at $\theta = 45°$.

where $\ell = 3$ μm is the active layer thickness and $\theta = 45°$ is the angle of the polished wedge, which is also the angle of incidence measured from the detector normal. In Equation 37.23, $4n/(1 + n)^2$ is the transmission coefficient of the GaAs substrate. The term $\cos(\theta)/2$ accounts for the smaller projected detector area in the direction of light and the fact that only the TM mode is coupled. The factor $\alpha \cdot \sin^2(\theta)$ is the absorption coefficient under oblique incidence, and $2\ell / \cos(\theta)$ is the total optical path length inside the detector accounting for the internal reflection at the top of the detector. In this classical model, if α is constant respect to the wavelength, η_c will also be wavelength independent. In Figure 37.15, η_c is shown to agree with the mean value from EM modeling. Therefore, EM modeling indeed supports the edge coupling approach in material characterization.

37.4.5 Quantum Grid Infrared Photodetector

Equipped with a quantitative model, we can revisit some of the QWIP coupling structures used in the past. Figure 37.16a shows the cross section of a structure known as the quantum grid infrared photodetector (QGIP) (Choi et al., 2004). It uses a metal layer on top of long and linear grid lines to scatter light incident from the bottom. The width w of the grid line acts as a half-wave dipole antenna and its dimension determines the detection wavelength. This detector geometry can also be modeled by 2D MTL method (Tamir and Zhang, 1996). Figure 37.16b compares the MTL solution and the present FEM solution for the same QGIP structure, assuming the same constant ε_i of $9.722 + i$. The close agreement in the QE spectrum between these two methods and the agreement between the methods and the experiment (Choi et al., 2004; not shown here) confirm the validity of both EM models. The E_z field distribution under the individual metal grid lines shows the expected scattering pattern from a dipole antenna.

37.4.6 Plasmonic-Enhanced QWIP

The structures described earlier were modeled in 2D. In this example, we perform 3D modeling as the structure cannot be reduced to 2D. The present structure, shown in Figure 37.17, uses surface plasmon to couple incident light into a QWIP (Wu et al., 2010). Surface plasmon is the resonant oscillation of conduction electrons at the interface between a metal and a semiconductor under optical excitation. The detector in this case contains a 400-Å thick gold film deposited on a 0.528-μm-thick InGaAs/InP active material.

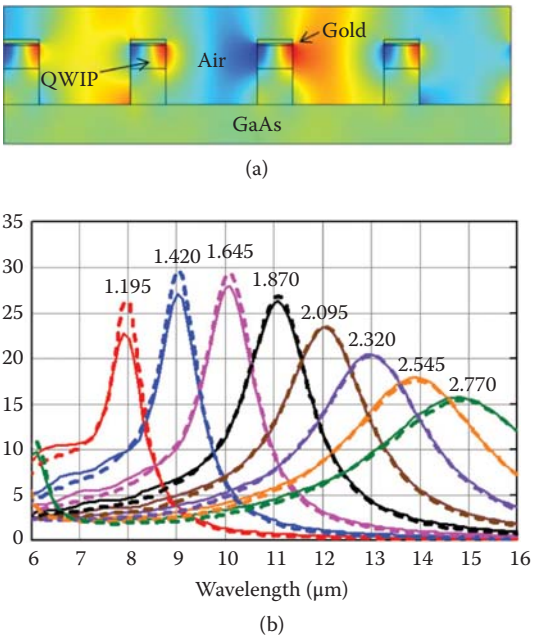

FIGURE 37.16 (a) The QGIP structure and the E_z distribution for $w = 1.87$ μm and $\lambda = 11.1$ μm. (b) The quantum efficiency (QE) spectra calculated based on FEM (solid curves) and MTL (dashed curves) having the same detector parameters. The numbers are w in microns. (QWIP, quantum well infrared photodetectors.)

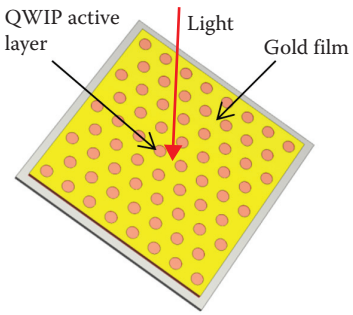

FIGURE 37.17 The 3D perspective of a plasmonic-enhanced quantum well infrared photodetectors (QWIP). Light is shining through the perforated gold layer from the top.

The gold layer is perforated with circular holes in a hexagonal lattice with spacing of 2.9 μm, and the hole diameter is 1.4 μm. The light incident on the gold layer excites surface plasmons in the opposite side with a resonant wavelength determined by the hole pattern. The resonant surface plasmons then create an intense vertical field at the gold/semiconductor interface. The active material has a low doping, with which the peak α is calculated to be 0.05 μm^{-1}. The modeled QE assuming this constant α for E_z is shown in Figure 37.18. It has a narrow peak at $\lambda = 8.3$ μm. Meanwhile, from the measured responsivity spectrum (Wu et al., 2010) and the estimated gain of 8.5 from a similar detector structure (Eker et al., 2010), the experimental peak QE is deduced to be 12.6%, which agrees with the modeling to within 10%, and the two spectra have similar lineshapes.

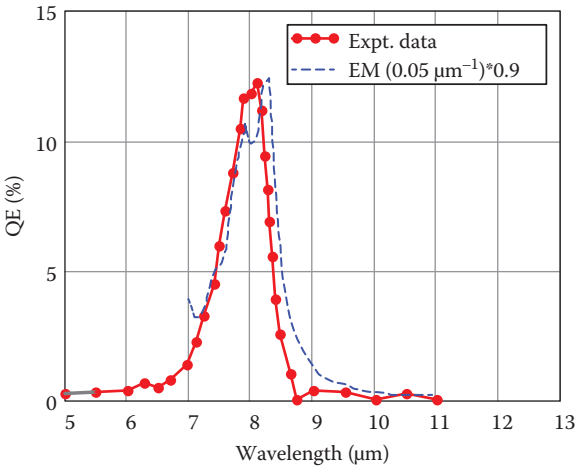

FIGURE 37.18 The measured (solid curve) and modeled (dashed curve) quantum efficiency (QE) of the plasmonic-enhanced QWIP. (EM, electromagnetic.)

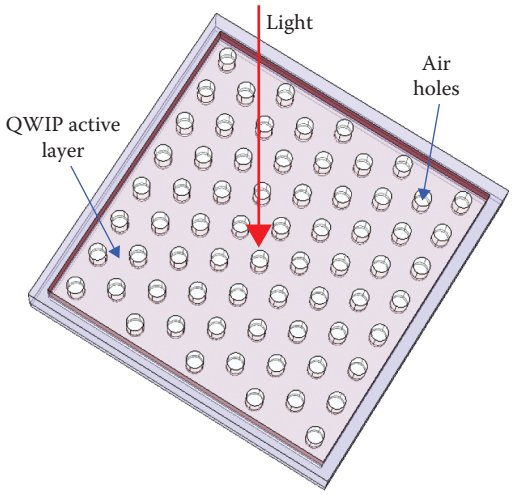

FIGURE 37.19 The 3D perspective of the PCS-QWIP. (QWIP, quantum well infrared photodetectors.)

37.4.7 Photonic Crystal Slab QWIP

To show another 3D optical structure, Figure 37.19 shows a photonic crystal slab QWIP (PCS-QWIP) studied (Kalchmair et al., 2011). The PCS is made of QWIP material, which is filled with periodic holes. These holes create a periodic modulation of refractive index in the x- and y-directions. Together with refractive index difference between GaAs and air in the z-direction, the PCS creates a series of characteristic photonic bands and bandgaps for the parallel propagating light. The normal incident light excites the stationary photonic eigenstates at the Γ-point of these bands in strong and sharp resonances. The present detector consists of an array of holes with spacing of 3.1 μm and diameter of 1.24 μm etched through the active and bottom contact layers. There is no metal layer in the optical path. The active material thickness is 1.5 μm, and the bottom contact thickness is 0.5 μm. The PCS is suspended in the air at a height t_{air} of 2.0 μm above the GaAs substrate.

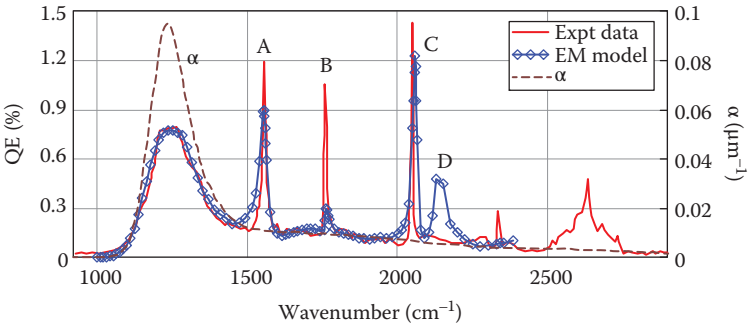

FIGURE 37.20 The measured photocurrent spectrum in arbitrary unit (solid curve) and the calculated QE spectrum based on the displayed α spectrum. (EM, electromagnetic.)

The measured photocurrent spectrum is shown in Figure 37.20. Based on the material α spectrum shown in Figure 37.20, which is deduced from the edge-coupled detector, the QE spectrum is modeled and it indeed contains the characteristic sharp peaks. However, the calculated peaks do not align exactly with the measurement. To obtain a better alignment, the theoretical hole spacing is reduced slightly from 3.1 to 2.9 μm as shown in Figure 37.20. This discrepancy could be due to the assumption of an average GaAs refractive index n to be 3.239 over a wide range of wavelengths. In reality, the value of n varies from 3.34 at 4 μm to 3.04 at 11 μm. Furthermore, the magnitudes of these peaks depend weakly on t_{air}. To obtain a larger peak at position B in Figure 37.20, t_{air} is adjusted from 2.0 to 0.9 μm. A different t_{air} could be due to the sagging of the PCS in the air at the operating temperature. From the modeling, the sharpness of these peaks is caused by two factors. One is the nearly symmetrical detector structure, both in vertical and horizontal directions, which induces strong resonances. The second is the weak material absorption at the peak wavelengths, which introduces only small damping effects on the resonances. The close match of the main peak and some of the side peaks lends support to the present modeling approach.

The large responsivity peaks A to C in the absorption tail illustrate that a large relative photoresponse can be obtained by using resonance in spite of a small intrinsic material absorption. Besides the examples presented earlier, we have also modeled other coupling structures, and the agreements between theory and experiment are similar to those presented (Choi et al., 2013a). They include air resonant cavities, linear and cross gratings, random-gratings, corrugated-QWIPs (C-QWIPs), etc. After studying all these examples, we found that the present FEM is highly accurate when comparing with other theoretical models. It also agrees well with experiments in both magnitudes and spectral lineshapes when both modeling and experimental uncertainties are taken into account.

37.5 EM Design of Resonator-QWIPs

The present FEM solves Maxwell equations numerically and thus should be able to optimize any detectors mentioned earlier as well as exploring new ideas. Among the many possible mechanisms and geometries, one structure that can yield a large QE is a photon trap. The function of a photon trap is to capture and hold incident light inside the detector until it is absorbed by the detector.

37.5.1 Resonator-QWIP Design

One realization of a photon trap is the resonator-QWIP (R-QWIP) as shown in Figure 37.21 (Choi et al., 2013a). It consists of an active layer, a thin GaAs substrate layer, and a number of diffractive elements (DEs) made of the same GaAs material on top. The topmost surface is then covered with a metal layer, and the detector is encapsulated with a low refractive index material.

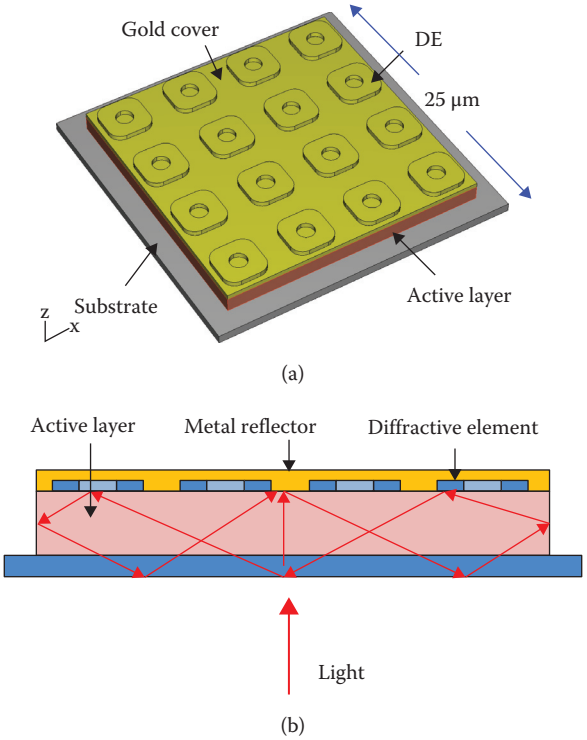

FIGURE 37.21 (a) A 3D perspective of an R-QWIP. Light is incident from the bottom. (b) The cross-section and one possible optical path.

The function of the DEs is to diffract incident light such that the subsequent angles of incidence at all detector boundaries are larger than the critical angle for total internal reflection (~18° for GaAs/air interface). If this is the case, the light will be totally confined in the detector. To account for interference effects, the size and shape of the detector volume are adjusted such that the scattered optical paths form a constructive interference pattern inside the detector. Under this condition, the newly incident light will be able to reinforce the light already under circulation, and the optical energy can be accumulated and stored in the detector as in a resonator. Therefore, by designing the detector into a resonator with a diffractive surface, i.e., an R-QWIP structure, an effective photon trap can be obtained.

To demonstrate the light storage capability, we first set α to zero so that there is no optical absorption in the detector. We then adjust the detector dimensions such that the intensity enhancement ratio (r_{ie}) given by

$$r_{ie} = \frac{1}{\frac{c\varepsilon_0}{2}E_0^2}\frac{nc\varepsilon_0}{2}\frac{1}{V}\int_V |E|^2 d^3r \tag{37.24}$$

reaches a maximum. The detector dimensions include the shape, size, and spacing of the DEs, the DE distribution pattern, the material layer thicknesses, the detector areal dimensions, etc. The quantity r_{ie} compares the average intensity built up inside the active volume relative to the incident intensity in the air. Figure 37.22 shows that r_{ie} can attain a value of 20 at 9.6 μm with the optimized structure shown in Figure 37.21a. For QWIPs, only the vertical intensity initiates absorption. The corresponding r_{ie} in Figure 37.22 for the vertical polarization component is 11. Such a large vertical r_{ie} produces a QE of 71% when α becomes finite and equals to 0.2 μm^{-1}, despite the very thin active thickness of 1.2 μm is adopted.

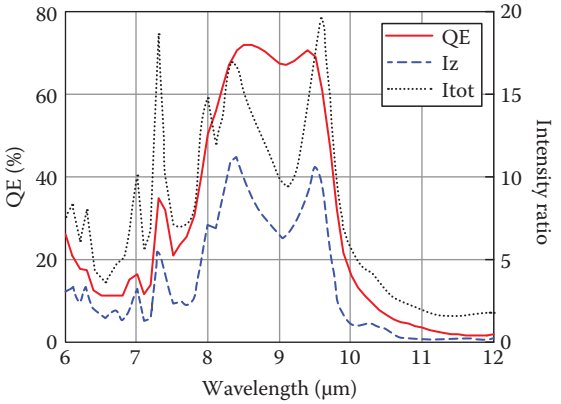

FIGURE 37.22 The calculated r_{ie} for the total intensity (dotted curve) and the vertical intensity (dashed curve) assuming $\alpha = 0$. The figure also shows the calculated QE (solid curve) if $\alpha = 0.2\ \mu m^{-1}$. The detector dimension is 22 μm. (DE, diffractive element.)

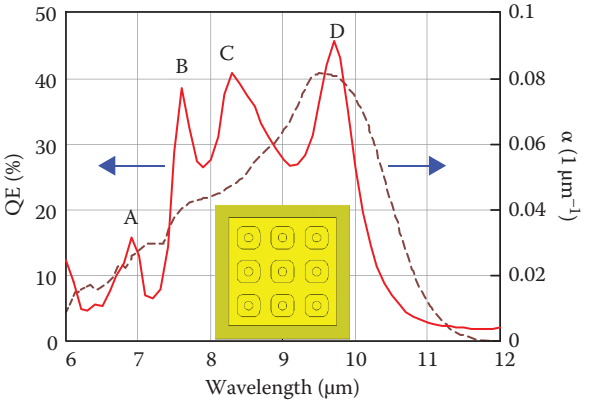

FIGURE 37.23 The measured α spectrum and the modeled quantum efficiency (QE) of an R-QWIP if $\alpha = 0.08\ \mu m^{-1}$. The pixel size is 20 μm, and the pixel pitch is 25 μm. The insert shows the DE pattern used for optical coupling.

37.5.2 R-QWIP Realization

We have conducted experimental research into R-QWIPs, and found QE as high as 70% in materials with doping larger than $1 \times 10^{18}\ cm^{-3}$ (Choi et al., 2013b, 2014, 2015). In this section, we present a more recent R-QWIP to show its advantages. The material consists of 19 QW periods. Each period is made of 50-Å GaAs/60 Å $Al_{0.22}Ga_{0.78}As$/50-Å GaAs/500-Å $Al_{0.22}Ga_{0.78}As$ double-barrier (DW) structure and each well is doped to $N_D = 0.5 \times 10^{18}\ cm^{-3}$. The adoption of the DW structure is to widen the absorption width, and the theoretical peak α is $0.08\ \mu m^{-1}$ for this N_D. The measured α lineshape based on edge coupling is shown in Figure 37.23. Its bandwidth of 3 μm is about twice that of the single well design. This experimental lineshape and a theoretical peak value of $0.08\ \mu m^{-1}$ is used for QE modeling. A matching resonator structure for this spectral range is shown in the insert of Figure 37.23. Assuming a λ-independent α of $0.08\ \mu m^{-1}$ and $n = 3.05$, the resonator shows four resonant peaks (A, B, C, and D) in Figure 37.23, and the highest peak value is 45%.

The measured I–V characteristics of the R-QWIP averaged over 1600 pixels are shown in Figure 37.24a, and the QE spectral lineshape measured at −1.1 V averaged over 1000 pixels is shown in Figure 37.24b. Figure 37.24b also shows the calculated spectral lineshape using the material α(λ). It is seen that the relative

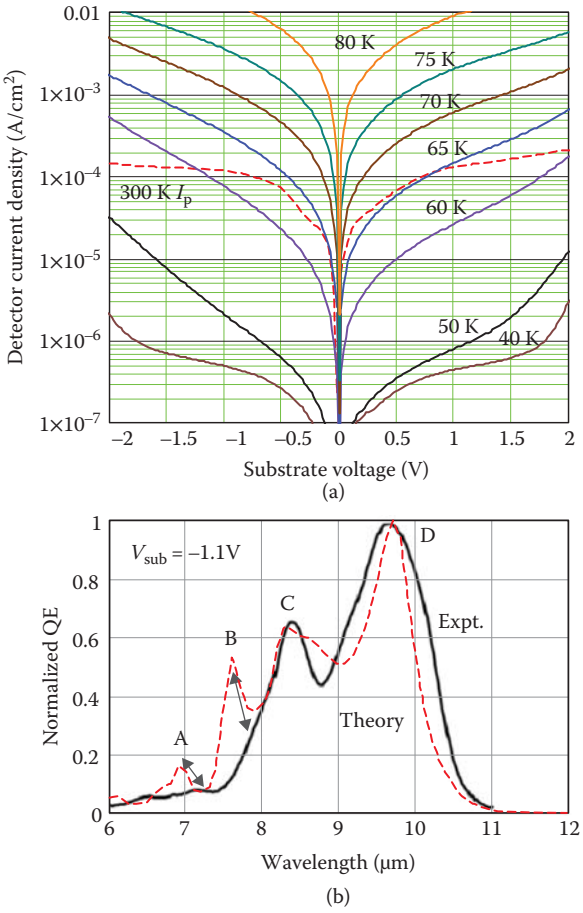

FIGURE 37.24 (a) The average I–V characteristics of 1600 R-QWIP pixels. (b) The measured average quantum efficiency (QE) lineshape of 1000 pixels in an FPA at −1.1 V and the modeled lineshape.

magnitudes of the experimental C and D peaks are as predicted while those of A and B peaks are much lower. The lower peaks also shift to longer wavelengths. This wavelength shift is expected when a constant n of 3.05 suitable only for 10 μm is used in modeling, while the actual n is known to increase toward shorter wavelengths. In addition to the less accurate n, the E_z distributions shown in Figure 37.25 at λ for A and B peaks indicate that they are higher order resonant modes that require higher precisions in the global structural dimensions. With the presence of processing nonuniformity, these resonances will be reduced.

For the QE magnitude, the main peak at 9.7 μm is measured to be 20% at −2 V and 37% at +2 V. The QE at positive bias is thus consistent with the modeling in this case. The polarity asymmetry is usually observed in R-QWIPs and is related to the highly localized optical intensity created in these detector structures as shown in Figure 37.26. For the two main peaks C and D, E_z is much stronger at the top of the active layer, where the QW layers are more active under positive bias. Therefore, this polarity produces a larger detection QE (Choi et al., 2014). In our simple dark current model described in Section 2.1, we assumed a linear potential drop in the MQW structure. In reality, the potential drop is slightly curved with potential drop near the cathode larger than the anode (Choi, 1997). This nonlinearity makes $\gamma(V)$ in $\eta_{abs}(\lambda)\gamma(V)$ larger near the top contact region than the substrate under positive bias. While the change of $\gamma(V)$ along z does not affect the average I_d with uniformly excited thermal electrons, it does promote larger I_p under positive bias when the optical intensity is localized. Therefore, R-QWIPs usually prefer positive bias operation. For future modeling efforts, the inclusion of $n(\lambda)$ and $\gamma(z)$ are clearly needed.

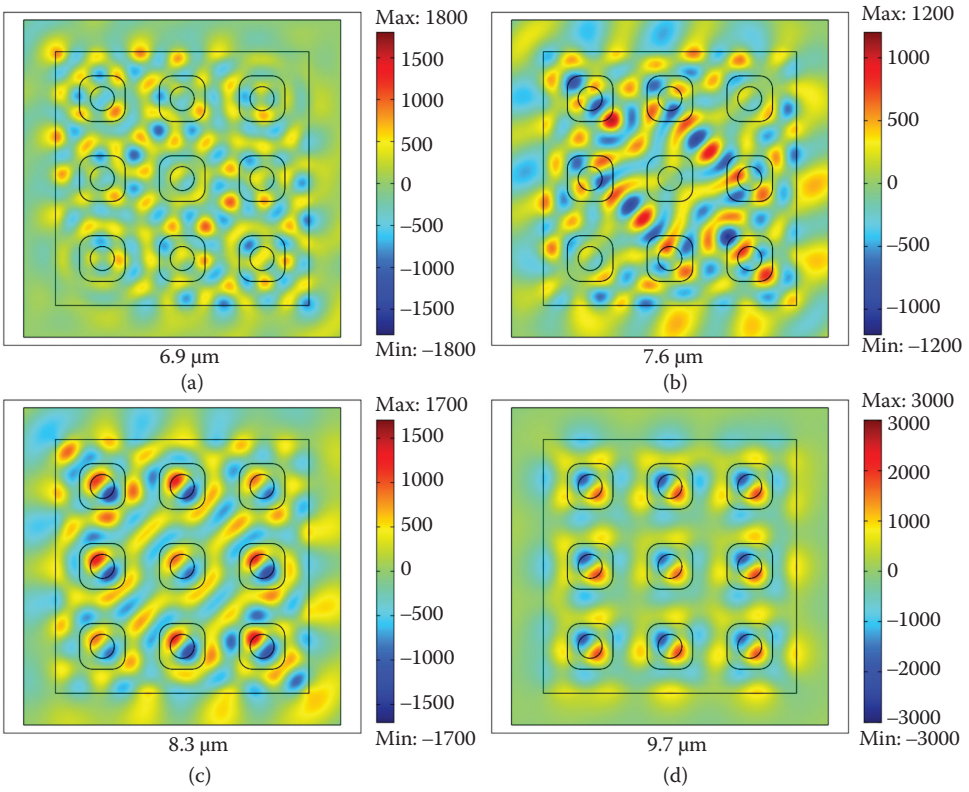

FIGURE 37.25 The calculated E_z at the four peak wavelengths in the middle plane of the active layer. E_0 is 377 V/m. Note the change of scale in the plots.

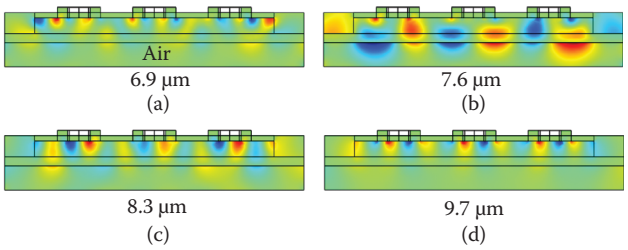

FIGURE 37.26 E_z at the center cross-section of the R-QWIP with distribution in (a), (b), (c), and (d) corresponding to peak A, B, C, and D, respectively. The value of E_z decreases exponentially from the top toward the substrate in (a), (c), and (d).

The advantages of the R-QWIPs can be seen by comparing the corrugated-QWIP performance (C-QWIP) in Figure 37.3a and the R-QWIP performance in Figure 37.24a. The C-QWIP uses angled sidewalls to couple normal incident light (Choi et al., 2011, 2012). The cutoff wavelength λ_c of these two materials are similar, which is 11.2 μm for the C-QWIP and 10.6 μm for the R-QWIP. The C-QWIP shows $J_d = 9 \times 10^{-4}$ A/cm^2 at 3 V and 60 K and $J_p = 1.5 \times 10^{-4}$ A/cm^2 at 3 V. For the R-QWIP, $J_d = 2.5 \times 10^{-5}$ A/cm^2 at 1 V and 60 K and $J_p = 1.5 \times 10^{-4}$ A/cm^2 at 1 V. While the difference in the dark currents of the two detectors is a factor of 36, the photocurrents are the same. Accounting for the longer λ_c for the C-QWIP, the R-QWIP improves the I_p/I_d ratio by approximately a factor of 10. Figure 37.27 shows an IR image taken by the present R-QWIP FPA, which shows its excellent sensitivity and uniformity. The FPA operating condition is as follows: temperature = 61 K, applied bias = −1.1 V, $F/\# = 2.5$, signal integration

FIGURE 37.27 This image was taken by the R-QWIP FPA with 1024 × 1024 pixels.

time = 3.06 ms, and signal processing = 1-point background subtraction. Under this operating condition, the pixel operability is 99.5% and thermal sensitivity is 45 mK. Even higher sensitivity is expected at a lower temperature or at positive bias.

37.6 Conclusion

In this chapter, we have discussed the theoretical methods in modeling the optoelectronic properties of QWIPs. We found that analytical methods can be used to explain the detector dark current and photoconductive gain. The intrinsic material absorption properties can be calculated from the TMM, and the detector QE can be solved by performing finite-element EM modeling. With these methods, all of the relevant optoelectronic properties of a QWIP can be predicted with very few empirical parameters. This ability greatly facilitates QWIP FPA production, since it allows more consistent and capable products be designed and manufactured.

Due to limited space, some of the advantages of EM modeling have not been discussed. For example, EM modeling can be used to design structures that use thinner materials without antireflection coatings or optical filters to reduce manufacturing cost and improve uniformity; structures that are more immune to fabrication nonuniformity and have less pixel spatial and spectral cross-talk; structures that have special functionality such as extremely narrowband detection, two-color detection, broadband detection, and polarization detection; and structures that are very small in size but are still sensitive so that ultra-high-density FPAs can be produced. For the last benefit, the model predicts that a QE of 56% can be obtained with an absorption coefficient of 0.2 μm^{-1} and pixel size of $5 \times 5\ \mu m^2$. Its realization will enable the production of very high-resolution FPAs with much smaller system size, weight, power, and cost. Besides their usefulness to QWIPs, EM modeling and the photon trap concept are also applicable to other sensors and energy convertors in the IR and other nearby wavelengths. Therefore, one can foresee the increasing vital role played by EM modeling in advancing different optoelectronic technologies in the future.

References

Choi KK (1997) *The Physics of Quantum Well Infrared Photodetectors*. River Edge, NJ: World Scientific.

Choi KK (2012) Electromagnetic modeling of edge coupled quantum well infrared photodetectors. *J. Appl. Phys*. 111:124507.

Choi KK, Bandara SV, Gunapala SD, Liu WK, and Fastenau JM (2002) Detection wavelength of InGaAs/AlGaAs quantum wells and superlattices. *J. Appl. Phys*. 91:551–564.

Choi KK, Chen CR, Goldberg AC, Chang WH, and Tsui DC (1998) Performance of corrugated quantum well infrared photodetectors. *Proc. SPIE* 3379:441–452.

Choi KK, Dang G, Little JW, Leung KM, and Tamir T (2004) Quantum grid infrared spectrometer. *Appl. Phys. Lett*. 84:4439–4441.

Choi KK, Jhabvala, MD, Forrai D, Sun J, and Endres D (2011) Corrugated quantum well infrared photodetectors for far infrared detection. *Opt. Eng*. 50:061005-1–6.

Choi KK, Jhabvala MD, Forrai DP, Waczynski A, Sun J, and Jones R (2012) Electromagnetic modeling of quantum well infrared photodetectors. *IEEE J. Quant. Electron*. 48:384–393.

Choi KK, Jhabvala MD, Forrai DP, Waczynski A, Sun J, and Jones R (2013a) Electromagnetic modeling and design of quantum well infrared photodetectors. *IEEE J. Sel. Top. Quant. Electron*. 19:1–10.

Choi KK, Jhabvala MD, Sun J, Jhabvala CA, Waczynski A, and Olver K (2013b) Resonator-quantum well infrared photodetectors. *Appl. Phys. Lett*. 103:201113.

Choi KK, Jhabvala MD, Sun J, Jhabvala CA, Waczynski A, and Olver K (2014) Resonator-QWIPs and FPAs. *Proc. SPIE* 9070:907037.

Choi KK, Sun J, and Olver K (2015) Resonator-QWIP FPA development. *Proc. SPIE* 9451:94512K.

Dixon JR and Ellis JM (1961) Optical Properties of n-type indium arsenide in the fundamental absorption edge region. *Phys. Rev*. 123:1560–1566.

Donges A (1998) The coherence length of black-body radiation. *Eur. J. Phys*. 19:245–249.

Eker SU, Arslan Y, Onuk AE, and Besikci C (2010) High conversion efficiency InP/InGaAs strained quantum well infrared photodetector focal plane array with 9.7 μm cut-off for high-speed thermal imaging. *IEEE J. Quantum Electron*. 46:164–168.

Fowles GR (1975) *Introduction to Modern Optics*, 2nd Edition. New York, NY: Holt, Rinehart, and Winston.

Grandidier J, Callahan DM, Munday JN, and Atwater HA (2012) Gallium arsenide solar cell absorption enhancement using whispering gallery modes of dielectric nanospheres. *IEEE J. Photovolt*. 2:123–128.

Hinnrichs M and Guptab N (2008) Comparison of QWIP to HgCdTe detectors for gas imaging. *Proc. SPIE* 6940:69401Q.

Jhabvala MD et al. (2011) Performance of the QWIP focal plane arrays for NASA's Landsat Data Continuity Mission. *Proc. SPIE* 8012:80120Q.

Kalchmair S, Detz, H, Cole GD, Andrews AM, Klang P, Nobile M, Gansch R, Ostermaier C, Schrenk W, and Strasser G (2011) Photonic crystal slab quantum well infrared photodetector. *Appl. Phys. Lett*. 98:011105.

Levine BF, Bethea CG, Hasnain G, Shen VO, Pelve E, Abbott RR, and Hsieh SJ (1990) High sensitivity low dark current 10 μm GaAs quantum well infrared photodetectors. *Appl. Phys. Lett*. 56:851–863.

Levine BF, Choi KK, Bethea CG, Walker J, and Malik RJ (1987) A new 10 micron infrared detector using inter-subband absorption in resonant tunneling GaAs/GaAlAs superlattices. *Appl. Phys. Lett*. 50:1092–1094.

Li J, Choi KK, Klem JF, Reno JL, and Tsui DC (2006) High Gain, broadband InGaAs/InGaAsP quantum well infrared photodetectors. *Appl. Phys. Lett*. 89:081128.

Miles RH, Chow DH, Schulman JN, and McGill TC (1990) Infrared optical characterization of InAs/Ga_{1-x}InxSb superlattices. *Appl. Phys. Lett*. 57:801–803.

Sadiku MNO (2001) Numerical techniques in electromagnetics, 2nd Edition. Boca Raton, FL: CRS Press.

Shi JJ and Goldys EM (1999) Intersubband optical absorption in strained double barrier quantum well infrared photodetectors. *IEEE Trans. Elect. Dev*. 46:83–88.

Tamir T and Zhang S (1996) Modal transmission-line theory of multilayered grating structures. *J. Lightwave Technol*. 14:914–927.

Wu W, Bonakdar A, and Mosheni H (2010) Plasmonic enhanced quantum well infrared photodetector with high detectivity. *Appl. Phys. Lett*. 96:161107.

38

Optical Modulators

Dominic F. G. Gallagher

and

Dmitry Labukhin

38.1 Introduction

The modulation of an optical signal is a key requirement of modern optical communications. Even when the transmitter is a laser diode, which can itself be directly modulated, there are advantages of using an unmodulated laser and an external modulator on the output.

Optical modulators can be constructed using many different physical effects. The main ones of interest are as follows:

1. Linear electro-optic, or Pockel's effect—an electrostatic field is used to modulate the refractive index of certain crystals.
2. The plasma effect—the alteration of the free carrier density in a semiconductor can change both the absorption and the refractive index.
3. The Franz–Keldysh (FK) effect and the related quantum-confined Stark effect (QCSE)—where an electrostatic field changes the band structure of a semiconductor, altering both the absorption spectrum and the refractive index.

In choosing a modulator, the following factors should be taken into account:

1. Do you want a phase modulator or an absorption modulator?
2. The modulation bandwidth needed.
3. Its energy efficiency, typically measured in pJ/bit.
4. The optical bandwidth required—some designs are based on resonant structures and can modulate only a narrow range of wavelengths, others can work over very wide optical bandwidths.
5. Size.

There is no such thing as a pure phase modulator or pure absorption modulator. The Kramers–Kronig relationship states that the dispersion spectrum (real part of the refractive index) can be known given the

absorption spectrum over all frequencies, and vice versa:

$$\varepsilon'(\omega) = 1 + \frac{2}{\pi} P \int_0^\infty \frac{\omega' \varepsilon''(\omega')}{\omega'^2 - \omega^2} d\omega' \tag{38.1}$$

$$\varepsilon''(\omega) = -\frac{2\omega}{\pi} P \int_0^\infty \frac{\varepsilon'(\omega') - 1}{\omega'^2 - \omega^2} d\omega' \tag{38.2}$$

where the dielectric constant $\varepsilon = \varepsilon' + i \cdot \varepsilon''$, and P is the *Cauchy principal value*. The Kramers–Kronig relationships tell us that the largest change in real index occurs near a change in absorption. Thus the most efficient phase modulators are likely to have absorption at nearby wavelengths.

38.2 QCSE-Based Modulators

38.2.1 Figures of Merit

Design of an electroabsorption modulator (EAM) requires accurate computation of its key characteristics. Among most important figures of merit are (a) on/off ratio defined as the output light power ratio between on and off states of the EAM; (b) the change in absorption coefficient per unit applied voltage; and (c) maximum data transmission speed. Calculating (a) and (b) is based on accurate modeling of absorption spectrum as a function of applied voltage, which will be the main focus of this chapter. Computing (c) involves modeling carrier dynamics. A detailed account of the latter can be found in [1–3].

38.2.2 Structure

Let us consider a typical traveling-wave EAM shown in Figure 38.1. This type of structure is used in this chapter as an example. Although other configurations have been reported as well (e.g., VCSEL-type EAM [4]), the traveling-wave type is the most common and the general principles of its operation can be extended to others with some modifications.

In a traveling-wave configuration, the light propagation dimension (z in Figure 38.1) is much greater than light- and carrier-confining dimensions (x and y in Figure 38.1) and, therefore, z can be separated from the x and y for computational efficiency. This means that we can split the device along axis z into a finite

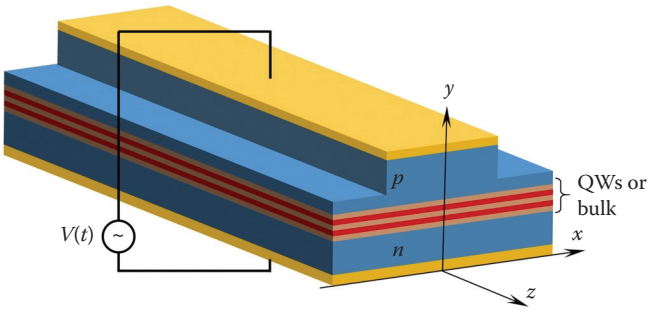

FIGURE 38.1 Traveling-wave EAM structure. The light signal propagates in the z-direction. Light confinement in the x–y plane can be realized via various configurations, the rigde waveguide shown in the figure being just one example. Epitaxial layers are grown in the y-direction. Top and bottom layers are metallic contacts. QWs and barriers comprise the active region. Next to the metallic contact layers are the substrate (n-doped) and cladding (p-doped). (QW=quantum well.)

number of short segments and assume optical properties within each segment to be z-independent, thus reducing the problem of modeling the optical properties to two transversal dimensions, x and y. Furthermore, in many cases, when the variation of the optical field and carrier density in one transversal direction (e.g., y in Figure 38.1) is much greater than in the other, the problem becomes one-dimensional. Then the calculated optical properties, namely, *material absorption* and *refractive index*, are used as input to an optical mode solver to compute the *effective index* and *modal absorption* of a particular waveguide mode that carries the optical signal. Finally, the result is passed as input to a one-dimensional z-propagation algorithm to obtain the device's characteristics. The first stage of this sequence, modeling material absorption and refractive index, is the focus of this chapter.

The optical absorption in EAMs is caused by several processes, among which are carrier transitions between conduction and valence band, free-carrier absorption (FCA), intraband transitions, photon–phonon interaction, etc. By far, the strongest contribution within the relevant spectral bandwidth is made by the first of these. It can be controlled by electrical bias applied across the epitaxial structure via one of the two main physical mechanisms: the FK effect in bulk layer structures or the QCSE in quantum wells (QWs). It is the latter one that is used in the majority of EAMs due to its higher efficiency. However, the FK effect is briefly examined here for the benefit of general understanding of these two phenomena.

38.2.3 Franz–Keldysh Effect

The Franz–Keldysh (FK) effect is the change in photon absorption spectrum induced by reverse bias applied to a bulk semiconductor structure (Figure 38.1). In an unbiased bulk layer, the electron and hole states are represented by space-harmonic functions. The probability of a photon-absorbing transition between two states is proportional to the overlap integral between their respective wavefunctions. When electric field F is applied across spatial direction y, the electron and hole wavefunctions change their shape and become Airy functions as shown in Figure 38.2b.

As the overlap integral between them changes, the probability of the photon absorption changes accordingly. Furthermore, the minimum absorption frequency falls below the bandgap, as can be seen from Figure 38.2b. This enables the control of the transmitted power of the light traveling through the EAM with applied voltage at a given wavelength (Figure 38.3). However, absorption change per applied voltage is relatively low and QW structures are preferred due to the higher efficiency of the QCSE.

38.2.4 Quantum-Confined Stark Effect

The QCSE is a counterpart of the FK effect in QW structures. It has, however, an additional feature: the presence of thin peaks in the absorption spectrum caused by the excitonic transitions. Indeed, unlike the bulk structures, QWs facilitate strong electron and hole confinement (see Figure 38.4), which causes

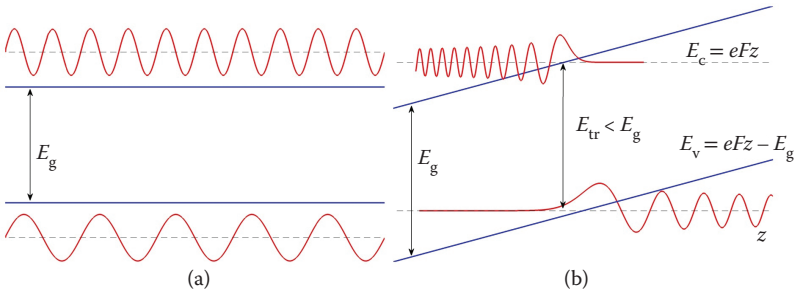

FIGURE 38.2 Electron and hole wavefunction in (a) an unbiased bulk semiconductor with bandgap E_g and when (b) a constant electric field F is applied across spatial direction y.

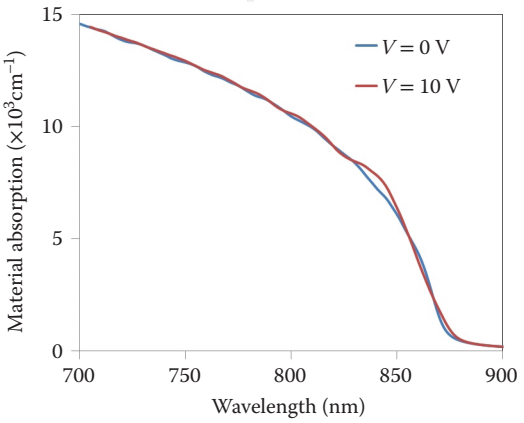

FIGURE 38.3 Change in absorption spectrum in 70-nm bulk AlGaAs under reverse bias due to the FK effect.

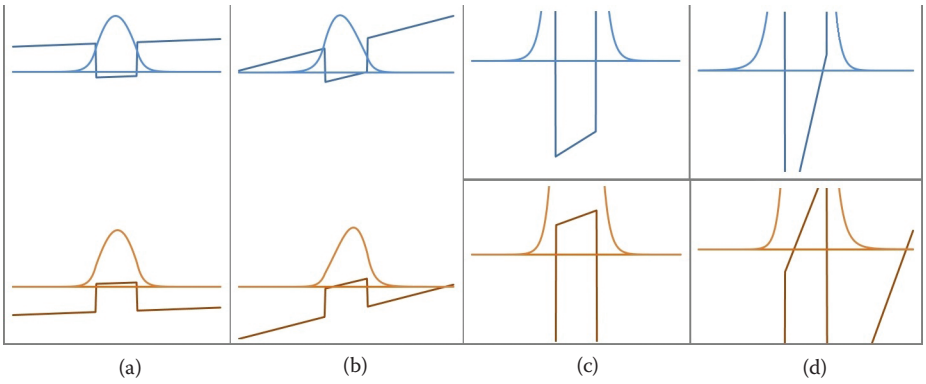

FIGURE 38.4 QW potential, electron and hole wavefunctions (a) without and (b) with a reverse bias. Plots (c) and (d) are (a) and (b), respectively, zoomed in to demonstrate the redshift in the energy levels.

the two carrier types to form bound states, *excitons*, similar to those between a proton and an electron in a hydrogen atom. These states introduce new discrete energy levels in addition to those of the QWs: but whereas the contribution to the absorption spectrum from the QW transitions has a staircase-like shape, the exciton contribution appears as peaks in the absorption spectrum.

Like in bulk layers, applying a reverse bias to the QW potential creates a double effect (see Figure 38.4). First, the electron and hole wavefunctions start "escaping" through the lower potential barrier, which decreases their overlap and, consequently, the probability of photon absorption. Second, their energy levels move closer to one another, which results in the redshift of the respective absorption peaks.

It can be seen from Figure 38.5 that if the absorption is measured at wavelength λ_0 near the exciton peak, the net effect of the two phenomena is the *increase* in absorption. This enables the control of the amount of light passing through the EAM by varying electric bias applied to the device. The material absorption change is much higher than in the case of the bulk structure. However, the material absorption is factored by the optical confinement factor, which is much smaller in QWs. To overcome this problem, multiple quantum wells (MQWs) are normally used to increase the efficiency of the device.

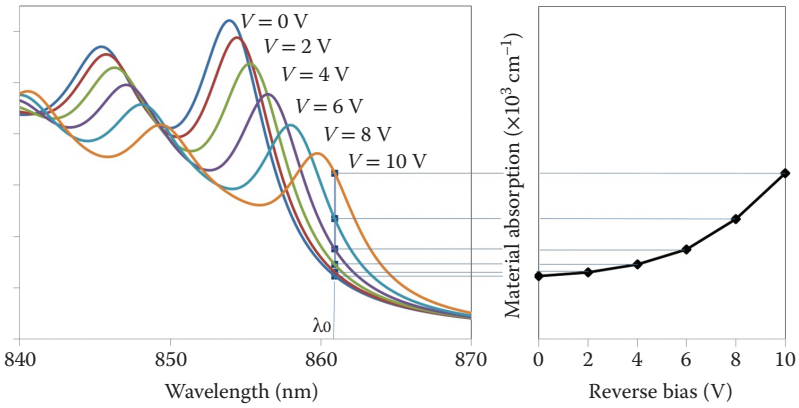

FIGURE 38.5 Change in material absorption spectrum in a 10-nm single QW AlGaAs EAM under reverse bias due to the QCSE effect. The peaks represent exciton's positions.

38.2.5 Modeling Absorption Spectrum

Modeling of the absorption spectrum as a function of voltage bias consists of several stages. Two most important effects contributing to the absorption is the carrier transitions between (a) energy levels created by QW confinement and (b) exciton binding energy levels. Photon absorption rate is proportional to the probability of the transition between an occupied state n in the valence band (hole) and an unoccupied state m in the conduction band (electron) multiplied by the Fermi population level $\alpha_{mn} \propto P_{mn}(f_m - f_n)$. The total absorption is given by the double sum over all relevant transitions mn, while each probability P_{mn} is determined from the overlap between the respective electron and hole wavefunctions. Therefore, the central part of the modeling is solving the Schrödinger equation in the QW potential and with an excitonic term added to its Hamiltonian. The QW potential, in its turn, is created by the juxtaposition of epitaxial layers with different bandgaps and applied electric field. It can be calculated using the Poisson-drift-diffusion model. In the case of EAMs, the coupling between the Schrödinger and Poisson equations can be neglected, which significantly reduces computational load.

38.2.6 Poisson-Drift-Diffusion Model

The Poisson-drift-diffusion model is based on three first-order differential equations: the carrier continuity equations for electrons and holes and the Gauss's law (see Chapter 3.1 in [5]),

$$\begin{aligned} \frac{\partial n}{\partial t} &= G_n - R_n - \nabla \cdot (\mu_n n \nabla \phi - D_n \nabla n) \\ \frac{\partial p}{\partial t} &= G_p - R_p - \nabla \cdot (\mu_p p \nabla \phi - D_p \nabla p) \end{aligned} \tag{38.3}$$

$$0 = \nabla \cdot \varepsilon \nabla \phi + q_e(p - n + N_D + N_A) \tag{38.4}$$

where $n = n(\boldsymbol{\rho}_{xy})$, $p = p(\boldsymbol{\rho}_{xy})$ are the electron and hole densities; $\boldsymbol{\rho}_{xy} = \{x, y\}$ is the spatial coordinate; $G_{n,p}(\boldsymbol{\rho}_{xy})$ and $R_{n,p}(\boldsymbol{\rho}_{xy})$ are generation and recombination rates, respectively; $\mu_{n,p}(\boldsymbol{\rho}_{xy})$ is electron and hole mobilities; $D_{n,p}(\boldsymbol{\rho}_{xy})$ is the electron and hole diffusion coefficients; $\varphi(\boldsymbol{\rho}_{xy})$ is the static field potential; $\varepsilon(\boldsymbol{\rho}_{xy})$ is the electric permittivity; $N_{D,A}(\boldsymbol{\rho}_{xy})$ is the donor and acceptor doping levels; and q_e is the electron charge.

Since in most EAMs the carrier density is very low, the generation and recombination terms can be discarded. With the reverse bias, the currents are also assumed negligible and, therefore, for the steady-state solution, the time derivatives are set to zero:

$$\begin{aligned} \nabla \cdot (\mu_n n \nabla \phi - D_n \nabla n) &= 0 \\ \nabla \cdot (\mu_p p \nabla \phi - D_p \nabla p) &= 0 \\ \nabla \cdot \varepsilon \nabla \phi + q_e (p - n + N_D + N_A) &= 0 \end{aligned} \tag{38.5}$$

Equation 38.5 is solved with respect to $\{n, p, \varphi\}$, whereas $\mu_{n,p}$, $D_{n,p}$, ε, and $N_{D,A}$ are usually assumed to be independent of $\{n, p, \varphi\}$ and are precomputed from the epitaxial layer composition. The potential φ is then used to find the spatial profile of the band edges:

$$E_{c,v}(\boldsymbol{\rho}_{xy}) = E_{c0,v0}(\boldsymbol{\rho}_{xy}) - \phi(\boldsymbol{\rho}_{xy}) \tag{38.6}$$

In addition to static field potential, the Poisson-drift-diffusion model calculates the distribution of carriers and, consequently, the distribution of charge in the PIN-junction as a function of applied bias. This is essential for calculating the device's characteristics in an electric circuit, such as capacitance, resistance, and, eventually, the modulation response.

38.2.7 Schrödinger Equation for QW States and Excitons

The next step after finding the band edge potentials $E_{c,v}(\boldsymbol{\rho}_{xy})$ is to solve the Schrödinger equation in order to find energy levels and spatial distributions of electron and hole wavefunctions. There exist several methods to do so, with different tradeoffs between accuracy and computational efficiency. The general form of the Schrödinger equation is

$$\hat{H}\Psi = E\Psi \tag{38.7}$$

The Hamiltonian operator $\hat{H}$ in the presence of excitonic effect is split into electron, hole, and exciton parts [6]:

$$\begin{aligned} \hat{H} &= \hat{H}^e + \hat{H}^h + \hat{H}^{exc} \\ \hat{H}^e &= \hat{H}^e_{kin} + E_c(\boldsymbol{r}_e) \\ \hat{H}^h &= \hat{H}^h_{LK} + E_v(\boldsymbol{r}_h) \end{aligned} \tag{38.8}$$

where $\hat{H}^e_{kin} \propto \nabla^2_e$ is the kinetic part of the electron Hamiltonian, $E_{c,v}(\mathbf{r}_{e,h})$ is the QW potentials with applied bias for conduction and valence bands, respectively, as calculated by solving Equations 38.5 and 38.6. Note that the spatial coordinates for electrons and holes are different. $\hat{H}^h_{LK}$ is the Luttinger–Kohn Hamiltonian for holes. The latter can be a 4×4 matrix operator, in case if only heavy and light holes are taken into account [6]. When highly strained structures are considered, the influence of the spin-orbit split-off band becomes significant and it needs to be included in $\hat{H}^h_{LK}$, increasing its size to a 6×6 matrix operator [29,30]. The exciton part, $\hat{H}^{exc}$, is due to the Coulomb interaction between electron and holes. The additional term in the Hamiltonian of the system is, therefore,

$$\hat{H}^{exc} = \frac{q_e}{4\pi\varepsilon_0\varepsilon} \frac{1}{|r_e - r_h|} \tag{38.9}$$

As the first step, the energies and the wavefunctions of the conduction and valence bands without the exciton contribution are obtained by solving the Schrödinger equation with respective Hamiltonians:

$$\begin{aligned}\hat{H}^{\mathrm{e}}\psi_m^{\mathrm{e}}(\mathbf{k},\mathbf{r}) &= E_m^{\mathrm{e}}\psi_m^{\mathrm{e}}(\mathbf{k},\mathbf{r})\\ \hat{H}_{\mathrm{LK}}^{\mathrm{h}}\psi_n^{\mathrm{h}}(\mathbf{k},\mathbf{r}) &= E_n^{\mathrm{h}}\psi_n^{\mathrm{h}}(\mathbf{k},\mathbf{r})\end{aligned} \tag{38.10}$$

It is assumed that the scale of the spatial variation of parameters in the x- and z-directions is insignificant in comparison to that in y-direction. The solution of these equations is, therefore, obtained by separating their spatial dependence: the wavefunction is assumed to be harmonic in the x- and z-directions and (Equation 38.10) is reduced to a one-dimensional eigenvalue problem with a y-dependent Hamiltonian only (note that in some references quoted here the perpendicular axis is denoted z, not y).

The one-dimensional Schrödinger equation can be solved using a finite-difference method or a so-called "shooting" method. In the case of the effective mass approximation (parabolic dispersion relation $E(\mathbf{k}_{xz})$), it only needs to be solved for $\mathbf{k}_{xz} = 0$, where $\mathbf{k}_{xz}$ is an in-plane wave vector (see, e.g., [7]).

If nonparabolicity of the dispersion curve needs to be taken into account, Equation 38.10 must be solved for a range of $\mathbf{k}_{xz}$ in the vicinity of zero. This can present a considerable computational challenge. The axial approximation removes the dependency of $\hat{H}_{\mathrm{LK}}^{\mathrm{h}}$ on azimuthal angle θ and the two-dimensional vector $\mathbf{k}_{xz}$ is replaced with its magnitude k, which significantly reduces computational time (see, e.g., [6]).

After the energies and wavefunctions in Equation 38.10 are found, the second step is to compute the exciton transition energy levels and wavefunctions with the full Hamiltonian:

$$\hat{H}\,\Psi^X(\mathbf{r}_{\mathrm{e}},\mathbf{r}_{\mathrm{h}}) = E^X\Psi^X(\mathbf{r}_{\mathrm{e}},\mathbf{r}_{\mathrm{h}}) \tag{38.11}$$

Here X denotes different exciton states, such as 1s, 2s, 2p, etc., called so by analogy with the atomic states. Two main methods of solving this problem are the momentum-space method [6] and the variation method [7].

In the momentum-space method, the exciton function is expanded in terms of conduction and valence wavefunctions $\psi_n^{\mathrm{e}}(\mathbf{k}_{xz}, y_{\mathrm{e}})$ and $\psi_m^{\mathrm{h}}(\mathbf{k}_{xz}, y_{\mathrm{h}})$ (found by solving Equation 38.10) in the $\mathbf{k}_{xz}$ space. As a result, Equation 38.11 is reduced to an eigenvalue problem with the integral operator on the right-hand side (see Equations 38.18 and 38.19 in [6]). Similarly to Equation 38.10, the axial approximation can be applied to full Hamiltonian $\hat{H}$: vector $\mathbf{k}_{xz}$ is replaced with its absolute value k, and the dependence of the wavefunctions on the azimuthal angle is confined to the phase term only: $\varphi(\mathbf{k}_{xz}) \rightarrow \varphi(k)e^{j\theta l}$. As mentioned earlier, the exciton states with $l = 0, \pm1, \pm2$, and ±3, are s, p, d, and f.

The exact mathematical details can be found in [6], but it is important to mention here that the final equation will contain a quadruple integral: over the spatial domain of the electron wavefunctions y_{e}, over that of the hole wavefunctions y_{h} spatial domain, and over the momentum angle of the in-plane vector $\mathbf{k}$; and over its magnitude k. The method was demonstrated to be in excellent agreement with the experimental results [6]. However, computing the quadruple integral for large structures (on the order of tens of nanometers in y-direction) can become prohibitively slow.

The variational method [7] is used to increase computational efficiency, albeit at the cost of accuracy. The exciton function in Equation 38.11 is expanded in terms of conduction and valence wavefunctions in real space ρ, not $\mathbf{k}$. Only the 1s exciton state is taken into account (so that the θ dependence is dropped, $\boldsymbol{\rho} \rightarrow \rho$):

$$\Psi^{1s}(\rho, y_{\mathrm{e}}, y_{\mathrm{h}}) = \phi_{mn}^{1s}(\rho)\psi_m^{\mathrm{e}}(y_{\mathrm{e}})\psi_n^{\mathrm{h}}(y_{\mathrm{h}}) \tag{38.12}$$

where $\phi_{mn}^{1s}(\rho)$ is a 1s trial function with variational parameter λ:

$$\phi_{mn}^{1s}(\rho) = \sqrt{\frac{2}{\pi}}\frac{e^{-\rho/\lambda}}{\lambda} \tag{38.13}$$

38.2.8 Permittivity Spectrum

Equations 38.7 through 38.13 serve to calculate the spectral positions of the exciton and QW transitions. The next step is to calculate the strength and the shape of the absorption $\alpha(\omega)$ as a function of light frequency ω due to these transitions and the corresponding change in the refractive index $n_{re}(\omega)$, both making part of the complex permittivity $\varepsilon(\omega) = [n_{re}(\omega) + i\alpha(\omega)c_0/(2\omega)]^2$, where ω is the angular frequency of light and c_0 is the speed of light in vacuum. The expression for total complex permittivity can be written as

$$\varepsilon(\omega) = \varepsilon^{opt} + \sum_{m(e)} \sum_{n(h)} (\varepsilon_{mn}^{QW}(\omega) + \varepsilon_{mn}^{exc}(\omega)) \tag{38.14}$$

where ε_{opt} is the permittivity for low optical frequencies, $\varepsilon_{mn}^{QW}(\omega)$ and $\varepsilon_{mn}^{exc}(\omega)$ is the permittivity change due to the QW and exciton transition between hole state n and electron state m. Whereas the QW contribution is a staircase-like function with steps located at energies $(E_m - E_n)$, the exciton transition is a delta function found at energies $(E_m - E_n - E_X)$. The exact formulae for the momentum-space and variational methods can be found in [6,7], respectively.

Two things must be pointed out. First, as mentioned earlier, the contribution both to the QW "steps" and exciton peaks is proportional to the overlap integral between the respective hole and electron functions:

$$\varepsilon_{mn}^{QW,exc}(\omega) \sim \int \psi_m^e(y)\psi_n^{h*}(y)dy \tag{38.15}$$

Second, due to the finite lifetime of the exciton states, these peaks have a spectral width $\simeq 1/\tau_{exc}$. Similarly, the steps of the staircase-like function of the QW contribution are smoothed out by a QW linewidth function, whose width is inversely proportional to the lifetime of the QW states. Mathematically, it is expressed as a convolution integral between $\varepsilon_{mn}^{QW,exc}(\omega)$ and respective linewidth functions $L_{mn}^{QW,exc}(\omega)$. The latter are complex-valued functions satisfying Kramers–Kronig relations (Equations 38.1 and 38.2) and they contribute to both the real and imaginary parts of the complex permittivity. Strictly speaking, each electron–hole pair should have its own $L_{mn}^{QW,exc}(\omega)$, but often in practice the same function is used for all pairs mn. There have been a few publications dedicated to calculating the shape of $L^{QW,exc}(\omega)$ (see, e.g., [9]). Still in many cases, a simple Lorenzian shape with an experimentally measured spectral width is used in calculations with satisfying accuracy [6].

38.2.9 Numerical Results

Below, experimental and modeling results are shown for several structures. Figure 38.6 shows the comparison between experimental results obtained for an SiGe modulator [10] and those simulated using the variational method. The variational method demonstrates excellent agreement in positions and shapes of the primary excitonic peaks and their dependence on the applied electric bias. It is not as accurate, however, in modeling higher energy spectra.

Furthermore, experimental results for an AlGaAs EAM, reported in [6], are compared with curves computed by the variational method and the momentum-space method (Figure 38.7).

It can be seen that the latter captures the general shape of the curve much more accurately, especially at higher energies. It can be argued, however, that the working frequency of the EAM is below the bandgap (to the left of the excitonic peak at zero bias) and it is, therefore, the most important spectral area to model. Moreover, the behavior in this area is largely dependent on the correct modeling of the exciton shape, not covered by either of the methods (see [9]). Therefore, in many cases, the variational method is preferred as a computationally fast alternative of the momentum-space method, especially for large MQW structures.

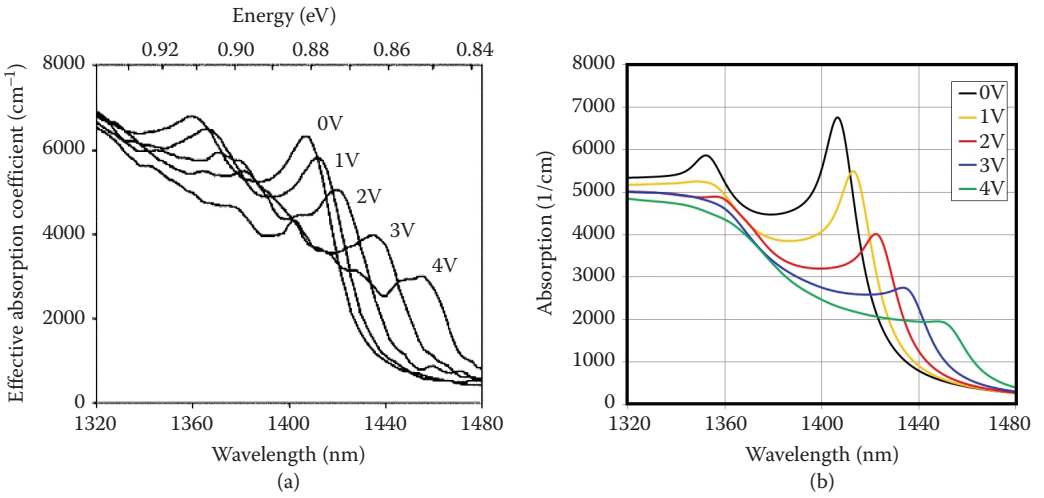

FIGURE 38.6 SiGe modulator. Transverse electric (TE) material absorption spectra for various values of the bias voltage: (left) experimental data from [10] (see detailed description of the structure in Section II of [10]) and (right) variational method by Photon Design Harold EAM.

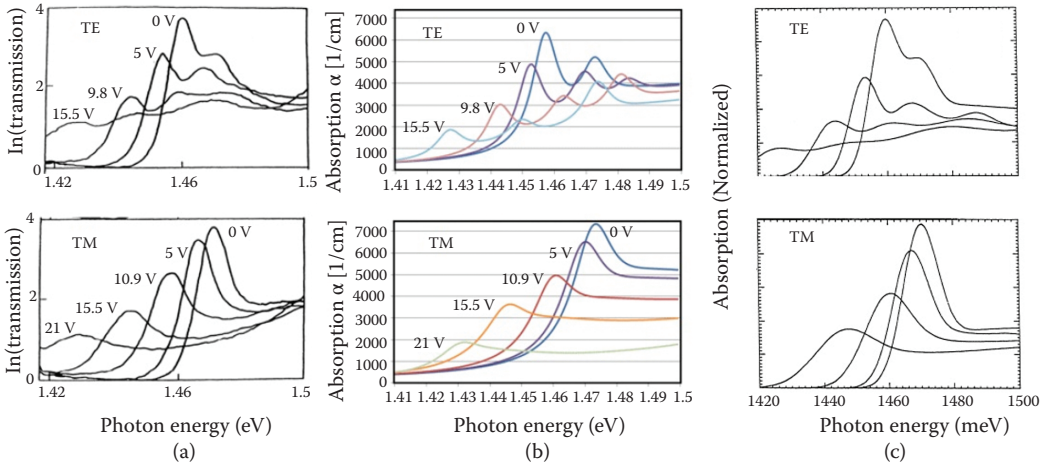

FIGURE 38.7 AlGaAs modulator [6] ([9.4 nm AlGaAs]/$Al_{0.3}Ga_{0.7}As$ QW, see detailed description of the structure in [31]). Absorption spectra for various values of the bias voltage: (a) experimental data from; (b) simulated curves using variational method for (top) TE polarization and (bottom) TM polarization; and (c) simulated curves using momentum-space method.

38.3 Electro-Optic Effect Modulators

The linear electro-optic effect, otherwise known as the Pockel's effect is present in crystals that do not have an inversion symmetry. This means they need an anisotropic dielectric tensor. Popular materials exploiting the Pockel's effect include $LiNbO_3$, $LiTaO_3$,and β-barium borate (BBO). The Pockel's effect is effectively instantaneous, once the electrostatic field has been established, making it suitable for exceedingly high-speed modulators.

$LiNbO_3$ has been a very popular material for electrorefractive modulators for many years because of its relatively large electro-optic coefficient [11,12], i.e., variation of refractive index with an applied electrostatic field. The rest of this section focuses on $LiNbO_3$ not because it is the only material available, but because of its dominant industrial position in photonics.

$LiNbO_3$ is strongly anisotropic implying a nontrivial dielectric tensor ε_{ij}. Since the tensor is symmetric, the double indexing conventionally simplified as

$$\begin{pmatrix} D_x \\ D_y \\ D_z \end{pmatrix} = \begin{pmatrix} \varepsilon_1 & \varepsilon_6 & \varepsilon_5 \\ \varepsilon_6 & \varepsilon_2 & \varepsilon_4 \\ \varepsilon_5 & \varepsilon_4 & \varepsilon_3 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} \tag{38.16}$$

In addition, the variation of each ε_{ij} with applied electrostatic field depends on the direction of the electrostatic field E, leading to a second-rank tensor:

$$\Delta \left(\frac{1}{\varepsilon_i} \right) = \sigma_{ik} E_k \tag{38.17}$$

This implies that different optical polarizations will experience different modulations. In $LiNbO_3$, only a few of the electro-optic tensor coefficients are significant:

$$\sigma_{13} = 8.6 \text{ pm/V}, \sigma_{22} = 3.4 \text{ pm/V}, \sigma_{33} = 30.8 \text{ pm/V}, \sigma_{51} = 28 \text{ pm/V} \tag{38.18}$$

Unlike for example, silicon modulators, there is no macroscopic movement of charge except that associated with the capacitance of the dielectric: $C = \epsilon_r \epsilon_0 A/d$; d is thickness, A is area. This leads to a very high-speed modulator capable of working to 30 GHz or more [13]. In fact, the change in refractive index is near instantaneous once the electric field has been established and the speed of an $LiNbO_3$ modulator is limited almost entirely by it capacitance and the speed of the electrical circuit driving it.

As the electro-optic tensor coefficients given earlier imply, the effect is anisotropic in both the direction of the applied field and in the modulation of the optical dielectric tensor. The crystal must be carefully oriented to obtain the desired effect. The two favored orientations are denoted x-cut and z-cut. An x-cut crystal will create a strong modulation of a TE-like polarized optical mode in response to an electrostatic field in the same transverse axis as the optical one, exploiting the σ_{33} coefficient. An x-cut crystal requires a vertical electrostatic field to modulate the same TE-like polarized optical mode and exploits the σ_{51} coefficient.

38.3.1 The Mach–Zehnder Modulator

The change in refractive index can be converted to a change in signal amplitude using a Mach–Zehnder interferometer design (Figure 38.8). With sufficient voltage, a 180°phase shift between the two optical paths is created by the time the waveguides combine, causing full cancellation of the signal.

A modulator can be made using a field between an electrode placed on top of the $LiNb0_3$ chip and a ground plane underneath. However, since the chip is rather thick, this leads to a low field strength, so typically both electrodes are placed on top of the chip either side of the waveguide, as shown in Figure 38.9.

The design in Figure 38.9 has one big flaw. The electrodes are very close to the optical mode; this causes substantial optical losses especially in the TM-like polarization. To cure this, a cap layer can be deposited on the $LiNbO_3$ before the electrodes are deposited. The buffer has another advantage in that it alters the velocity of the electrical signal along the modulator, allowing it to better match the velocity of the optical signal. This substantially improves the modulation bandwidth.

The disadvantage of the buffer is a reduction in electric field strength in the vicinity of the optical mode, requiring a longer modulator to achieve the same phase change. In Figure 38.10, you can see the voltage contours of a design with a 100-nm SiO_2 buffer. Notice that a lot of potential is dropped across the buffer layer. This typically leads to a ~30% drop in modulator efficiency.

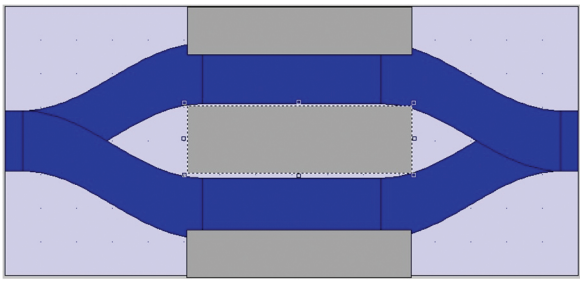

FIGURE 38.8 An $LiNbO_3$ modulator based on a Mach–Zehnder interferometer. A 3-dB splitter divides the input signal between two arms. One arm is given a positive electric field and the other arm a negative field, creating an increase in mode effective index in one branch and a similar decrease in effective index in the other.

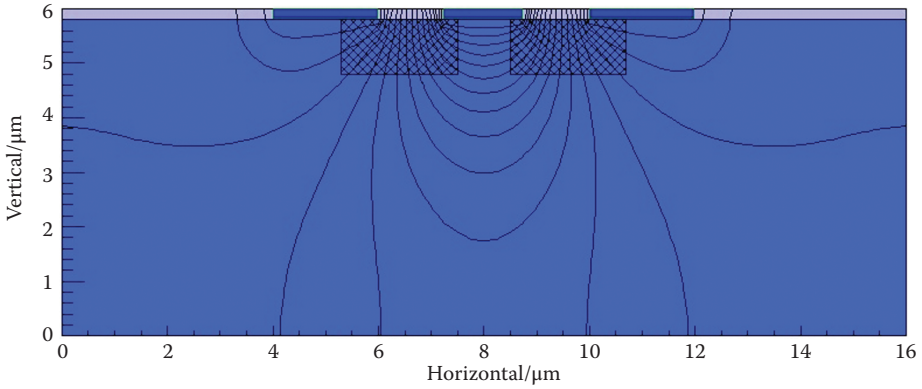

FIGURE 38.9 Cross section of a typical x-cut $LiNbO_3$ Mach–Zehnder modulator showing the Ti-diffused waveguides (hashed), three electrodes on top of the structure and electric potential contour lines. Typically, the outer electrodes are grounded and the central one carries the modulating signal. The opposite electrostatic field direction across each waveguide creates a natural push-pull effect for the Mach–Zehnder modulator. Note how the strongest field is generated in the middle of the waveguides.

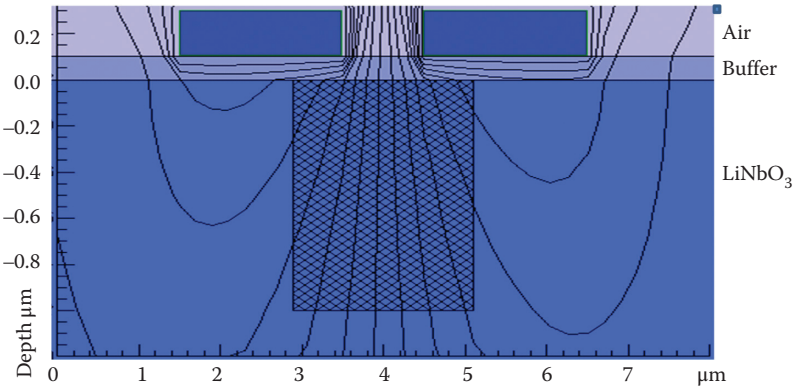

FIGURE 38.10 An x-cut $LiNbO_3$ modulator cross section with a cap layer creating an offstand between the electrodes and the optical mode. But note the strong field drop across the cap layer. This will reduce the electric field in the vicinity of the optical mode.

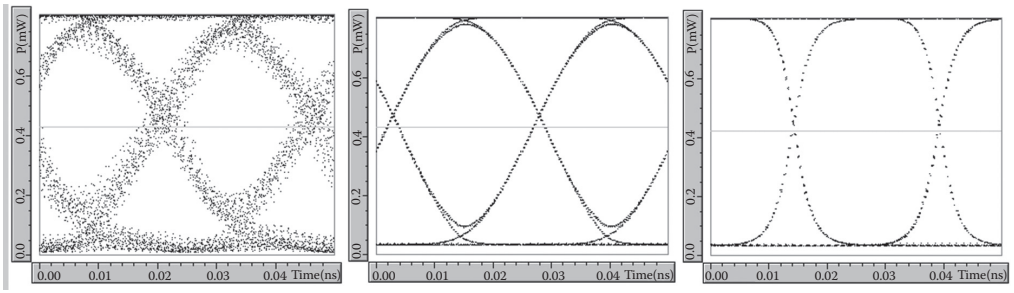

FIGURE 38.11 Eye diagrams of a traveling-wave $LiNbO_3$ MZI modulator at 40 Gb/s, computed with an optical circuit simulator PicWave [14] that includes traveling-wave time domain models for both the optical field and the electrodes: (left) with an unmatched electrode contacted a few millimeters from the end of the 20-mm-long modulator; (middle) contacted at the end; and (right) with a velocity-matched electrode.

38.3.2 Traveling-Wave Design

Because the change in effective index with voltage is typically very small, ~0.0001 per volt, the modulator has to be very long. This means that both the optical signal and the electrical signal will take a significant time to travel along the device. In order to maximize the modulation bandwidth, it is advantageous to match the velocity of the optical and electrical signals by appropriate design of the electrodes. Figure 38.11 shows some simulations of a traveling-wave $LiNbO_3$ Mach–Zehnder modulator under 40 Gb/s NRZ data stream. The figures show "eye diagrams" of the modulation response. To create an eye diagram the circuit is driven by a random bit sequence, and response of each clock period is superimposed. A clear open "eye" indicates a good signal.

38.4 Plasma-Effect Modulators

The plasma effect is the name given to the interaction of free electrons with an optical field. A plasma naturally occurs in a metal, which is why all metals are opaque. In order to make a modulator, one needs a way to control the carrier density, which is not realistic in a metal. In a semiconductor, the free-carrier population can be readily controlled by either injecting carriers into a diode (forward bias) or by lightly doping a region of the semiconductor and using a gate electrode or a reverse-biased junction to create an electrostatic field to deplete the carriers from the lightly doped region.

This section focuses on silicon-based plasma-effect modulators. Silicon is becoming a popular platform for manufacturing photonic components, largely because of the benefits of access to the huge electronic silicon manufacturing industry, its mature foundries and tools. Although silicon is not capable of generating light efficiently, i.e., making a laser, it is capable of absorbing light or modulating its phase by controlling the plasma effect in the optical path. Silicon optical modulators are now capable of reaching tens of gigahertz modulation speed, nearly rivalling the best $LiNbO_3$ modulators and in a much smaller volume, with much lower power requirements.

38.5 Si Electroabsorption versus Electrorefractive Effect

Silicon can act as an optical modulator either by a variation of the optical absorption in the silicon [15]—an "electroabsorptive modulator" or more popularly by modulation of the refractive index [16–19]—a so-called "electrorefractive modulator". In fact this distinction is somewhat arbitrary, since all changes in the real part of the refractive index will cause a change in the imaginary part, and vice versa. As discussed in Introduction, the real and imaginary changes do not generally happen at the same wavelength and the

relationship is described by the so Kramers–Kronig relationship. At a wavelength of 1.55 μm and for low doping densities, the real and imaginary changes are given approximately by [17]

$$\Delta n = -8.8x10^{-22}\Delta N - 8.6x10^{-18}(\Delta P)^{0.8}$$
$$\Delta\alpha = +8.5\times10^{-18}\Delta N + 6\times10^{-22}\Delta P$$

A big disadvantage of an electrorefractive modulator is that usually we want to modulate the amplitude of an optical signal and to create an amplitude modulation out of such a modulator generally requires interference effects in for example, a Mach–Zehnder interferometer [18–20] or a ring resonator [16,17, 21,22]. Both of these have a strong wavelength-dependent response, especially the ring resonator. Despite this, silicon electrorefractive modulators have arguably had the greater impact due to their lower power and higher modulation bandwidths.

Some of the earliest silicon modulator designs are based on carrier injection in a PIN diode [20–22]. Here a forward current is injected into the diode and the free charge carriers in the intrinsic region cause a change in refractive index. Figure 38.12 shows the cross section of a typical carrier-injection PIN diode modulator. The effect is only weakly dependent on wavelength, though once included in a Mach–Zehnder or resonant cavity structure, this advantage is lost. The effect is however only weakly dependent on temperature unlike QCSE or other bandgap-dependent effects.

An alternative Si modulator structure reverse biases a pn (or PIN) junction [18,19,23,24]. The reverse bias causes carrier depletion in the junction, and an associated alteration of the local refractive index. Figure 38.13 shows a possible geometry, one with a vertical junction [8]. In general, it is preferable to move the p-n junction to one side so most modulation is done by p-type carriers. This is because the p-carriers create less absorption at wavelengths of interest for a given modulation depth, since the absorption peak is moved further out.

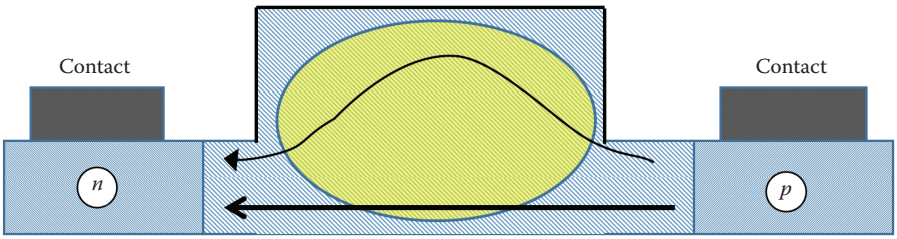

FIGURE 38.12 A PIN diode type silicon optical modulator. The silicon is etched to form a rib waveguide in the intrinsic region. Forward current interacts with the optical mode in the rib, changing its velocity or effective index.

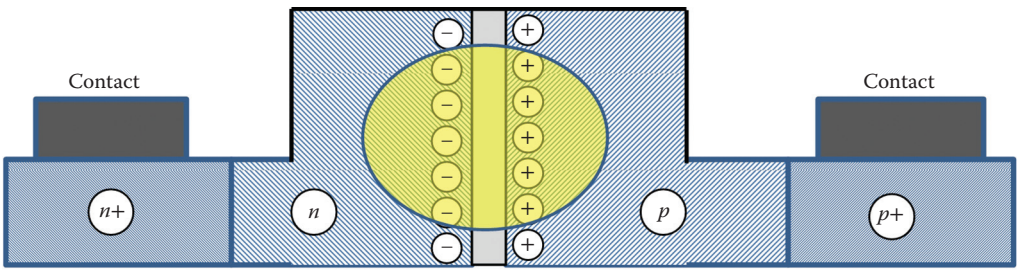

FIGURE 38.13 A reverse-bias type Si optical modulator. In these designs, the reverse bias caused electrostatic charge to build up at the p-n junction. These charges change the local refractive index.

Geometries similar to the former with asymmetric p-n junction have achieved a Vπ of 10 V in a 1-mm waveguide and bandwidths in excess of 50 Gb/s for a 1-mm long asymmetric Mach–Zehnder interferometer [19]. An experimentally measured eye diagram from this modulator is shown in Figure 38.14.

Figure 38.15 shows a recent design with a horizontal junction corresponding to a capacitor device [23]. The device is forward biased to cause carrier accumulation across the capacitor. This device achieved a modulation capability of 10 Gb/s at a length of 625 μm. Because the p and n carriers are close to each other, this design does not avoid the higher absorption of the n-type carriers.

Although most Si optical modulators are primarily electrorefractive, some electroabsorptive modulators have been demonstrated. In this case the free carriers are increased to a sufficient density that substantial FCA occurs. FCA is caused by carriers close to the band edge absorbing a photon and being excited to a state further up the conduction band (electrons) or down into the valence band (holes). The excited carriers then rapidly lose their energy by nonradiative transitions. An interesting design was recently proposed based on a Schottky diode [25]. The structure is predicted to reach a 3-dB bandwidth at 7 GHz for a 500-μm-long traveling-wave design. This is equivalent to a data rate of over 15 Gb/s using NRZ coding. Other groups have demonstrated EAMs using SiGe QWs—the QCSE effect (see Section 38.2) or by depositing graphene on the silicon [15,26,27].

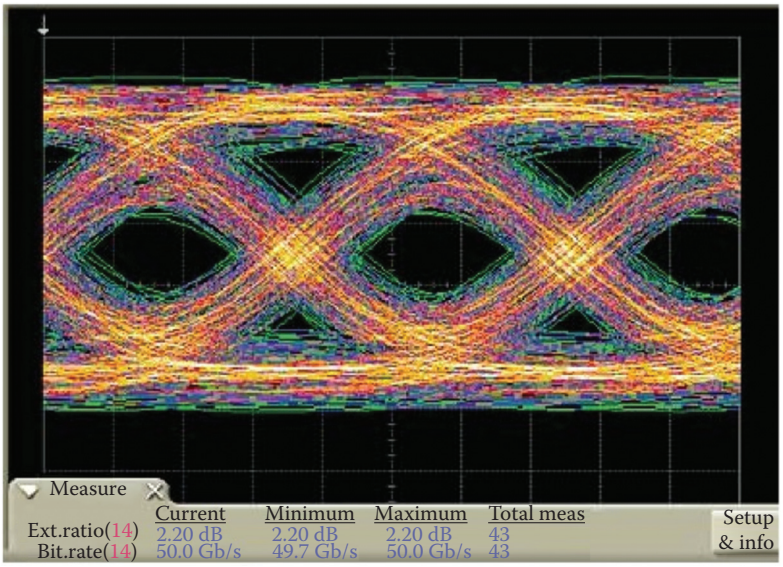

FIGURE 38.14 Experimental measurement of 50 Gb/s data stream in a depletion-type Silicon modulator in an asymmetric Mach–Zehnder interferometer configuration.

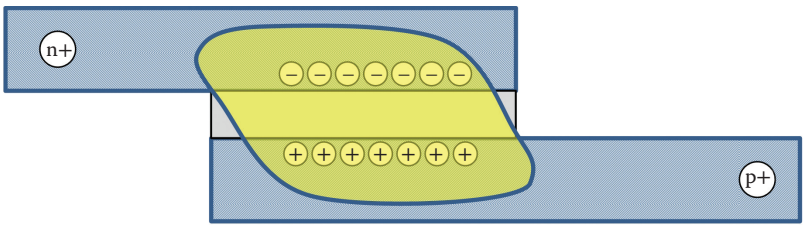

FIGURE 38.15 A capacitor-like structure also works by charge build up at the interface.

38.6 Resonant Modulators

In previous paragraphs, we outlined the use of traveling-wave designs such as the Mach–Zehnder interferometer to implement a modulator. Because the change in refractive index tends to be small for most modulator effects, this leads to long devices—hundreds of microns to tens of millimeter. This is not practical for many applications and so many have looked into resonant structures to reduce the modulator size. The principle is simple—in a resonant cavity, the light will remain trapped, bouncing or circulating around for a time much greater than L/c, where L is some characteristic dimension of the device. In this section, we focus on a popular structure—the ring resonator [16,17,21,22]—but many other designs are possible, for example, a photonic crystal cavity [28].

A resonator is popularly characterized by its "Q" factor, which can be defined as the photon lifetime divided by the optical period. A ring resonator has a roundtrip time of

$$t_{\text{round}} = \frac{2\pi R}{c}\mu \tag{38.19}$$

where μ is the characteristic refractive index, R is the ring radius. The photon lifetime of a cavity for optical frequency f is Q/f. This means that a ring resonator can increase the time spent in the ring by a factor of

$$\frac{Q\lambda_0}{2\pi R\mu} \tag{38.20}$$

This gives the modulator more time to change the phase of the signal, and allows the modulator to be shrunk be a corresponding factor. The disadvantage is that the ring resonator optical bandwidth is reduced to f/Q, making it sensitive to small changes in input wavelength.

A typical resonator configuration is shown in Figure 38.16 (left). This design has only one tangential waveguide and will act as a phase modulator. If a second tangential waveguide is added (Figure 38.16, right) then the modulator will direct the light into this when in resonance.

The high Q lowers the modulation speed of the device as well as its optical bandwidth. Figure 38.17 shows a PicWave [14] simulation of an eye diagram for a 20-μm-diameter silicon ring resonator modulator driven in carrier depletion mode with a 20 Gb/s NRZ bit pattern. Notice the overshoot caused by the high Q of the cavity, in this case over 6000. In practice, this ringing would be filtered out by a low-pass filter on the detector.

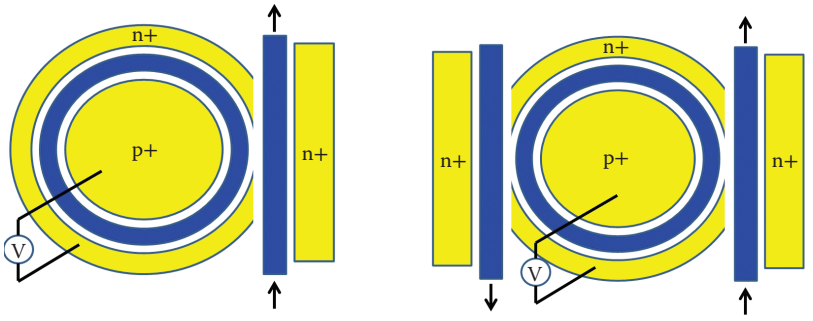

FIGURE 38.16 (Left) A silicon ring resonator modulator. The waveguides in blue (dark) are surrounded by electrodes imposing a reverse bias on the p-n junction (depletion-mode modulator). This design is a phase modulator. (Right) An amplitude modulator.

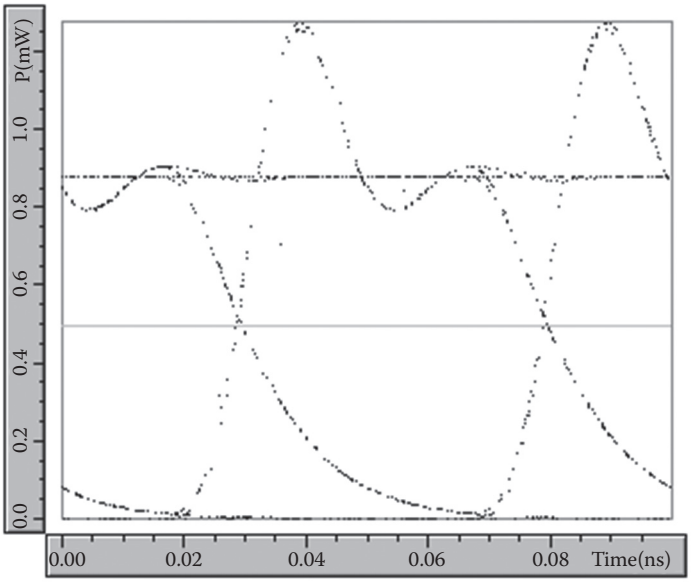

FIGURE 38.17 Eye diagram of a 20-μm-diameter silicon ring resonator modulator. The overshoot is caused by the high Q of the cavity.

References

1. Lefevre KR, Anwar AFM (1997) Electron escape time from single quantum wells. *IEEE J Quantum Electron*. 33: 187–191.
2. Hojfeldt S (2002) Modeling of carrier dynamics in electroabsorption modulators. PhD Thesis. Technical University of Denmark
3. Nikolaev V, Avrutin E (2003) Photocarrier escape time in quantum-well light-absorbing devices: Effects of electric field and well parameters. *IEEE J Quantum Electron*. 16: 24–26.
4. van Eisden J et al. (2007) Modulation properties of VCSEL with intracavity modulator. *Proc SPIE*. 6484: 64840A.
5. Piprek J (2003) *Semiconductor Optoelectronic Devices: Introduction to Physics and Simulations*. Academic Press London
6. Chao CYP, Chuang SL (1993) Momentum-space solution of exciton excited states and heavy-hole-light-hole mixing in quantum wells. *Phys Rev B*. 48: 8210–8221.
7. Mares PJ, Chuang SL (1993) Modeling of self-electro-optic-effect devices. *J Appl Phys*. 74: 1388–1397.
8. Alloatti L et al. (2011) 42.7 Gbit/s electro-optic modulator in silicon technology. *Opt Express*. 19: 11841–11851.
9. Cho HS, Prucnal PR (1989) Effect of parameter variation in the performance of GaAs/AlGaAs multiple-quantum-well electroabsorption modulator. *IEEE J Quantum Electron*. 25: 1682–1690.
10. Kuo YH et al. (2006) Quantum-confined stark effect in Ge–SiGe quantum wells on Si for optical modulators. *IEEE J Sel Topics Quantum Electron*. 12: 1503–1513.
11. Alferness RC, Joyner CH, Buhl LL, Korotky SK (1983) High-speed travelling-wave directional coupler switch/modulator for l = 1.32 μm. *IEEE J Quantum Electron*. 19: 1339–1341.
12. Kubota K, Noda J, Mikami O (1980) Traveling wave optical modulator using a directional coupler $LiNbO_3$ waveguide. *IEEE J Quantum Electron*. 16: 754–760.

13. Kondo J et al. (2002) 40-Gb/s X-cut $LiNbO_3$ optical modulator with two-step back-slot structure. *J Lightwave Technol*. 20: 2110–2114.
14. PicWave - optoelectronic simulator, ver. 5. www.photond.com/products/picwave.htm.
15. Hu YT et al. (2014) Broadband 10 Gb/s graphene electro-absorption modulator on silicon for chip-level optical interconnects. In Proceedings of 2014 IEEE Electron Devices Meeting (IEDM), San Francisco, CA.
16. Dong P et al. (2009) Low Vpp, ultralow-energy, compact, high-speed silicon electro-optic modulator. *Opt Express*. 17: 22484–22490.
17. Lipson M (2006) Compact electro-optic modulators on a silicon chip. *IEEE J Sel Topics Quantum Electron*. 12: 1520–1526.
18. Liu A et al. (2007) High-speed optical modulation based on carrier depletion in a silicon waveguide. *Opt Express*. 15: 660–668.
19. Thompson D et al. (2012) 50-Gb/s silicon optical modulator. *IEEE Phot Tech Lett*. 24: 234–236.
20. Zhou GR et al. (2008) Effect of carrier lifetime on forward-biased silicon Mach-Zehnder modulators. *Opt Express*. 16: 5218–5226.
21. Manipatruni S, Xu Q, Schmidt B, Shakya I, Lipson M (2007) High speed carrier injection 18 Gb/s silicon micro-ring electro-optic modulator. In Proceedings of 2007 Lasers and Electro-Optics Society, Lake Buena Vista, FL (LEOS 2007), 537–538.
22. Xu Q, Manipatruni S, Schmidt B, Shakya J, Lipson M (2007) 12.5 Gbit/s carrier-injection-based silicon microring silicon modulators. *Opt Express*. 15: 430–436.
23. Abraham A, Olivier S, Marris-Morini D, Vivien L (2014) Evaluation of the performances of a silicon optical modulator based on a silicon-oxide-silicon capacitor. In Proceedings of 2014 IEEE 11th International Conference on Group IV Photonics, 3–4.
24. Gardes FY, Tsakmakidis KL, Thomson D, Reed GT, Mashanovich GZ, Hess O (2007) Micrometer size polarisation independent depletion-type photonic modulator in Silicon on insulator. *Opt Express*. 15: 5879–5884.
25. Jeong U, Han D, Lee D, Lee K, Kim J, Park J (2014) A broadband silicon electro-absorption modulator (EAM) using a Schottky diode. *Proc SPIE*. 8988: 89881M.
26. Liu M, Yin X, Zhang X (2012) Double-layer graphene optical modulator. *Nano Lett*. 12(3): 1482–1485.
27. Phare CT, Lee YD, Cardenas J, Lipson M (2015) Graphene electro-optic modulator with 30 GHz bandwidth. *Nat Photonics*. 9: 511–514.
28. Tanabe T, Nishiguchi K, Kuramochi E, Notomi M (2009) Low power and fast electro-optic silicon modulator with lateral p-i-n embedded photonic crystal nanocavity. *Opt Express*. 17: 22505–22513.
29. Chao CYP, Chuang SL (1992) Spin-orbit-coupling effects on the valence-band structure of strained semiconductor quantum wells. *Phys Rev B*. 46: 4110–4122.
30. Lever L et al. (2010) Design of Ge-SiGe quantum-confined stark effect electroabsorption heterostructures for CMOS compatible photonics. *J Lightwave Technol*. 28: 3273–3281.
31. Weiner JS et al. (1985) Strong polarization-sensitive electroabsorption in GaAs/AlGaAs quantum well waveguides. *Appl Phys Lett*. 43: 1148–1150.

VIII

Solar Cells

39

Solar Cell Fundamentals

Matthias Müller

39.1 Introduction

Solar cells convert the energy of the sun light to electricity. It is an optical to electrical energy conversion. In principle, they make use of the photovoltaic (PV) effect, i.e., the generation of a potential difference between two electrodes under illumination. It was firstly observed by A. E. Becquerel (Becquerel, 1839) and practically implemented in a solar cell at Bell Laboratories in the United States (Chapin et al., 1954). Today it is believed that PV energy generation will be a part of the world's base of energy supply established in the twenty-first century. Conservative estimates by the International Energy Agency (IEA) show a share of the world electrical energy generation of about 1% nowadays to about 16% in 2050 (IEA, 2014). This would correspond to a yearly energy generation of about 6300 TWh compared to slightly more than 200 TWh today.

Looking back on the past decades, we can see that the development in solar research and industry was very fast and economically successful, which enabled such a bright perspective for solar electricity. The price for solar modules reduced with increasing manufactured PV module capacity following a power-law relation: the so called PV learning curve could be established which is shown in Figure 39.1.

For example, the price of solar modules decreased by about 80% between the years 2008 and 2015. About 1000 GW module shipments may be reached in 2024 for 10% market growth, or even in 2021 for 30% market growth, which corresponded to the market development in the past 15 years. Those cost reductions

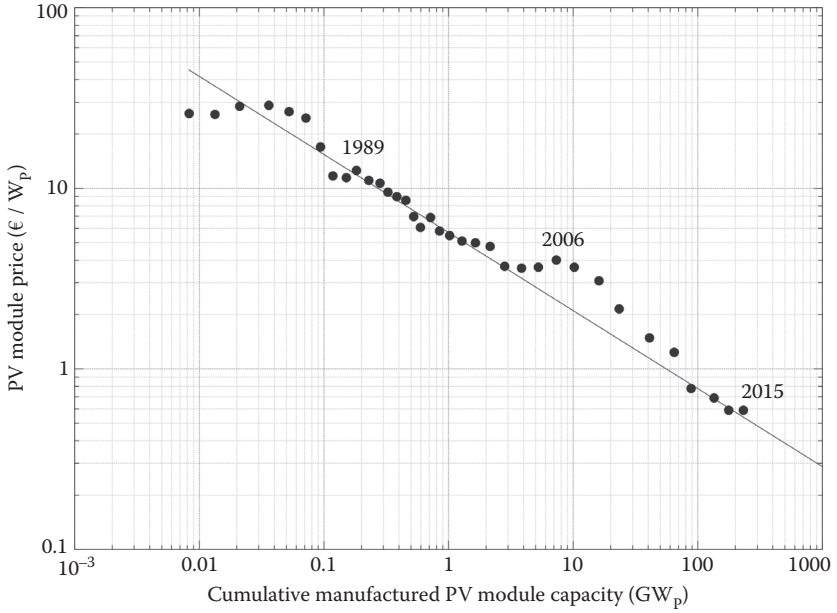

FIGURE 39.1 PV learning curve (Fraunhofer ISE, 2013; IEA, 2014; pvXchange, 2016) and a fit of a power law model to the years 1990–2000 and 2012–2015. In between the years 2000–2010, there was a shortage on raw silicon which slowed down the price reduction.

result largely from economies of scale and technological improvements. In 2015, the limit of 200 GW-shipped solar modules was exceeded, and the same volume is expected to be the future world production capacity to maintain the global solar energy supply in the twenty-first century starting from 2025 (IEA, 2014).

The current solar cell production is dominated by wafer-based crystalline silicon (c-Si) solar cells, which make about 90% of the market (ITRPV, 2015) and only 10% are thin-film technologies based mainly on cadmium telluride (CdTe) and copper indium gallium diselenide (CIGS). That is why the author focuses on the c-Si technology within this chapter.

Looking closely on c-Si, the shares of different wafer technologies that are applied nowadays are shown in the upper graph of Figure 39.2 (ITRPV, 2015), where mainly boron-doped p-type material is used (95%). The cast silicon materials multicrystalline (mc) and high-performance multicrystalline (HPmc-Si) currently make up to over 65%, whereas monocrystalline, mainly Czochralski (Cz), make 35%.

The share of different solar cell technologies is shown in the lower graph of Figure 39.2, where the main product currently is the full-area back-surface-field (BSF) solar cell (Mandelkorn and Lamneck, 1972), whereas the passivated emitter and rear cell (PERC) technology (Blakers et al., 1989) shows increasing market shares in the near future (ITRPV, 2015). The passivated emitter and rear totally diffused solar cell (PERT) (Zhao et al., 1999), the Si-heterojunction (SHJ) solar cell (Tanaka et al., 1992) and the back contacted solar cells—mainly interdigitated back contact solar cells (IBC) (Lammert and Schwartz, 1977)—are expected to play a minor role in the coming years but may be candidates for future high-efficiency solar cells with higher market shares (ITRPV, 2015). The structure of these different solar cells is shown in Figure 39.3.

This chapter gives a general overview about the working principle and basic models of solar cells. The view will be extended to the energy yield delivered by solar cells under real world conditions, i.e., the view will be extended to solar cells working in solar modules and the corresponding cell demands. Examples will

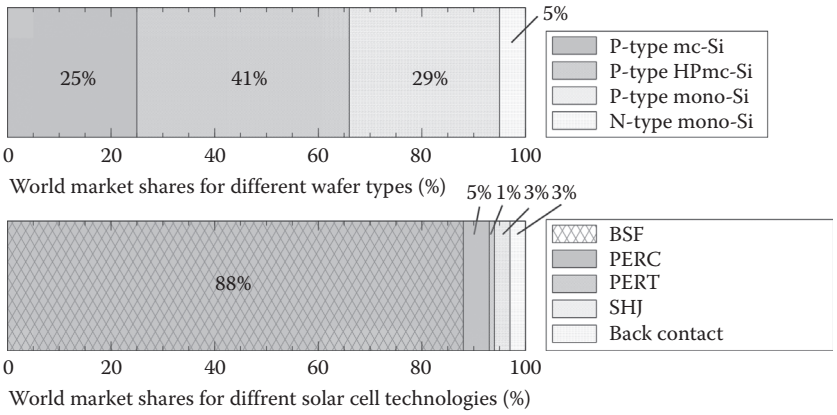

FIGURE 39.2 Share of different silicon wafer technologies for solar cell application and share of different solar cell technologies. [BSF, back-surface-field; PERT, passivated emitter and rear totally diffused; PERC, passivated emitter and rear cell; SHJ, Si-heterojunction. (ITRPV, 2015)]

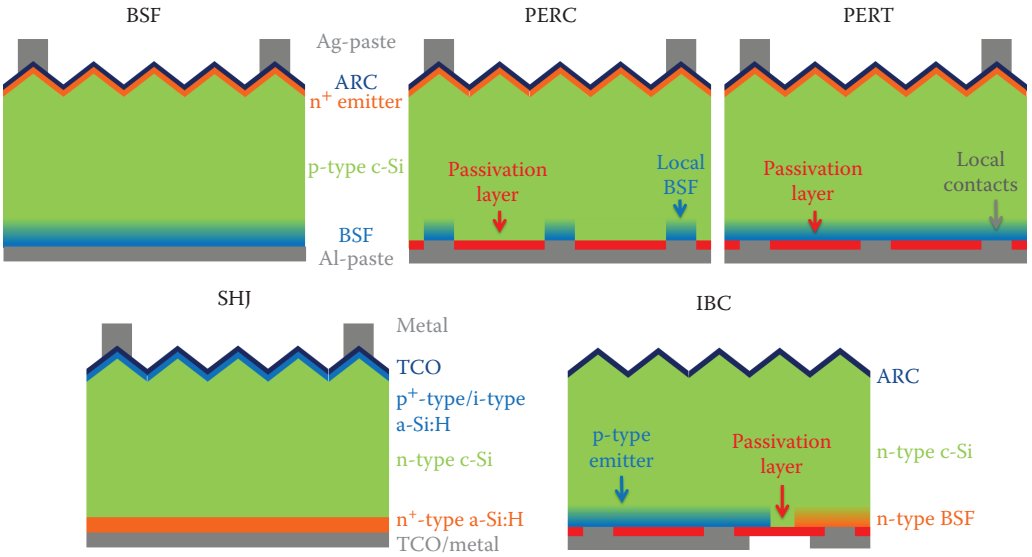

FIGURE 39.3 Schematic of typical solar cell structures in their two-dimensional cross section representation where antireflection coating (ARC) and transparent conductive oxide (TCO) are used. (BSF, back-surface-field; PERC, passivated emitter and rear cell; PERT, passivated emitter and rear totally diffused; SHJ, Si-heterojunction; IBC; interdigitated back contact solar cell.)

be presented considering only nonconcentrated crystalline silicon-based solar cells for 1-sun application. For additional information and explanations which are not given in this chapter, the author would like to refer to the notable literature by Wuerfel (2005), Green (1982), and Goetzberger et al. (1998). Additionally, the author would like to refer to the free online resources for PV engineers and scientists PVEducation (PVEDUCATION.ORG, 2016) and PV Lighthouse (PV Lighthouse, 2016).

39.2 The PV System

The solar cell is only a part, even so the functional device, of a full PV system. Such systems may be grid-connected as shown in Figure 39.4, or stand-alone systems, which means that all produced electricity is consumed locally. Those stand-alone PV systems are good options for rural areas of developing countries and special applications (e.g., parking meter). However, in comparison, grid-connected systems produce nowadays and will produce in the future a higher share of electricity. A grid-connected system consists of solar modules, which contain the solar cells. These modules are typically connected in series (called module string) to an inverter. Those inverters first do Maximum Power Point Tracking (MPPT) of the module strings and secondly convert the direct current produced by solar cells in alternating current. The inverter is connected to the load and to the electricity grid via an AC isolation switch and via an import/export electricity meter (compare Figure 39.4).

A current trend for grid-connected systems is the combination with storage systems. It increases energy self-consumption, as it is expected that during noon an excess of solar energy is produced. This excess energy can then be stored and used during times without irradiation. The trend will be enhanced, as the price for electricity is decreasing more and more at noon time when more solar power plants are connected to the grid. So the storage system will help for grid stabilization as PV electricity is a fluctuating energy challenging the electricity grid in the future. Those battery PV systems are controlled by an energy management system (EMS) which regulates production, load, storage, and grid connection (compare Figure 39.4).

Looking in more detail at the structure of a solar module as shown in Figure 39.5, one can see that solar cells are basically embedded between glass and plastics.

A typical solar module contains 60 solar cells on an area of approximately 1.7 m^2. They are connected in series by tabbing ribbons to strings of 20 cells while bus ribbons interconnect the three cell strings. Those cell strings are connected in parallel by a bypass diode which accounts for lower power losses during shading conditions. Finally, the bus ribbons are connected to the junction box, where cables with connectors go out.

Solar modules are made to produce electricity for at least 20–30 years; hence all used materials have to withstand changing environmental conditions. Solar glass, that in most cases is low-iron rolled glass or float glass, is situated on the side facing the sky and has to withstand rain, hail and snow. Besides, the front glass guarantees mechanical stability against wind and snow load. It may be manufactured with an antireflection coating (ARC) and/or textured, i.e., with a structured front surface for improved light trapping. The rear side is typically covered with polyethylene terephthalate (PET) polymer backsheet which features moisture resistance and electrical isolation. If replaced by another glass to produce a glass-glass module, the solar

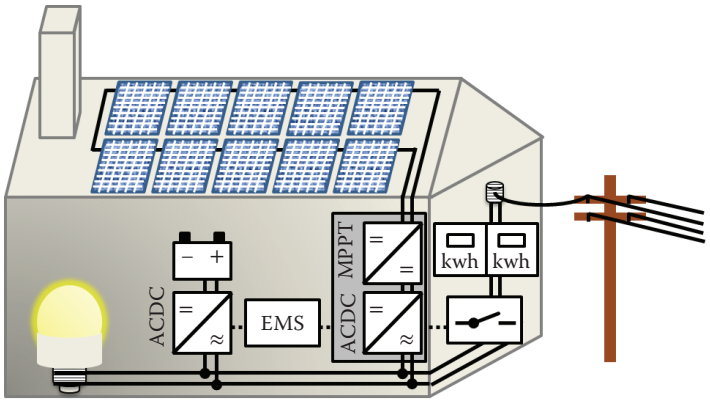

FIGURE 39.4 Schematic of a PV system with solar modules, cables, inverter, battery energy managements system (EMS), battery, load, AC isolation switch, and import/export electricity meter.

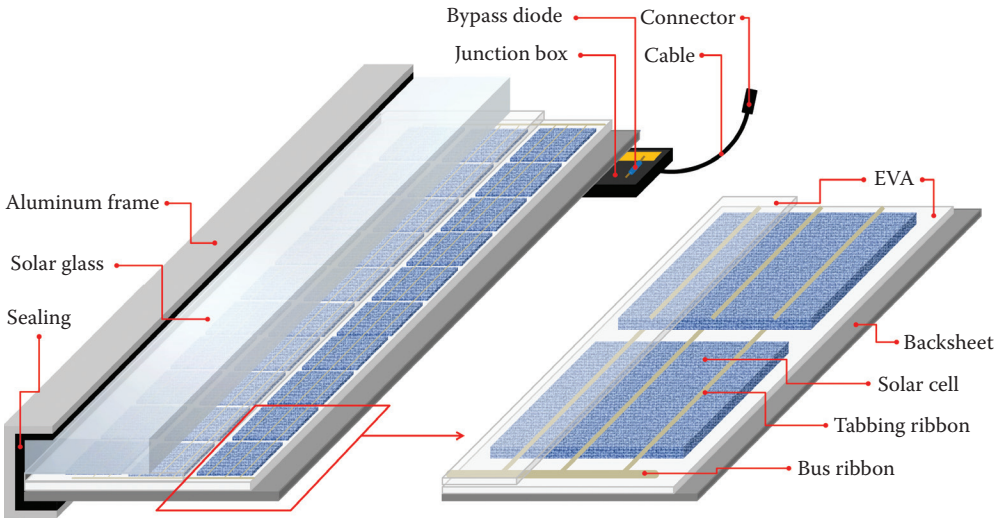

FIGURE 39.5 Structure of a solar module. EVA, ethylenvinylacetate.

module lifetime increases by about 5–10 years. Finally, the solar cells are embedded between the solar glass and the backsheet in Ethylenvinylacetate (EVA) as an encapsulant. It is a polymer which cross-links at about 150°C during manufacturing. Solar modules are sealed at the edges with silicone and an aluminum frame.

39.3 The Solar Spectrum

Our sun is basically a fusion power plant which emits electromagnetic radiation from its surface, called photosphere, into the solar system. It has a spectrum similar to a blackbody with a temperature of 5777 K, which is shown in Figure 39.6.

The light intensity reaching the planet earth is reduced by a factor of 2.17×10^{-5} due to geometrical considerations. The factor is the square of the suns radius divided by the square of the distance sun–earth. The resulting spectrum above the earth's atmosphere is shown in Figure 39.6 denoted as AM0, where AM stands for air mass index and the "zero" denotes that the light doesn't travel through the earth's atmosphere. This extraterrestrial solar spectrum AM0 has an irradiance of 1366.1 W/m^2 as defined within (ASTM E490-00a, 2014), which is called the solar constant. The latest measured value is 1360.8 W/m^2 (Kopp and Lean, 2011). This spectrum is relevant for solar cell power measurements for space applications.

Light traveling through the atmosphere faces absorption, scattering, and reflection effects which change the spectrum and the irradiance as shown schematically in Figure 39.7.

Absorption in the earth's atmosphere takes place in its molecules: ozone, oxygen, water vapor, carbon dioxide, nitrogen dioxide. Scattering in the atmosphere takes place at particles that are smaller than the wavelength of the light (Rayleigh scattering, e.g., N_2, O_2 molecules) and which are larger than the wavelength of the light (Mie scattering, e.g., aerosol particles, cloud droplets). The third mechanism is reflection of the light in the atmosphere, mainly at the clouds, and finally at the ground. This introduces a direct, a diffuse, and a ground-reflected component of irradiance on the module plane. The diffuse component describes all light coming from any region of the sky, mainly Rayleigh-scattered light which we see as blue sky on sunny days.

The standard spectrum AM1.5g is the result of light passing the earth's atmosphere approximately on a very sunny day and it is shown in Figure 39.6. The air mass index AM defines the direct optical path length through the atmosphere of the earth. Thus, AM1.5 is one and a half times the path length; "g" stands for

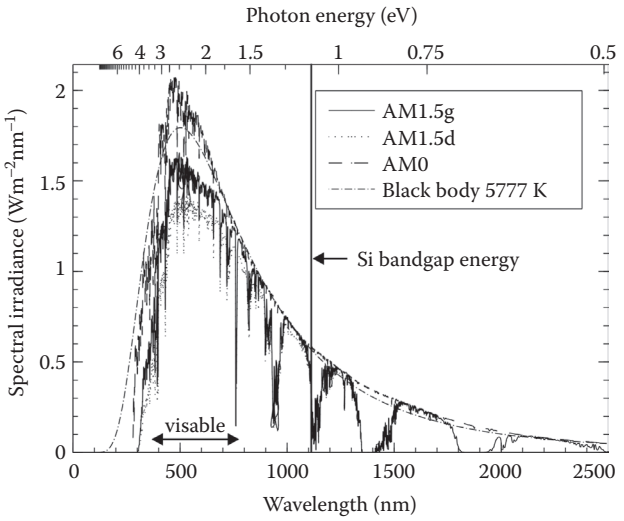

FIGURE 39.6 Comparison of different spectra: blackbody spectrum with a temperature of 5777 K, extraterrestrial spectrum AM0 (ASTM E490-00a, 2014), reference spectra AM1.5g and AM1.5d (ASTM G173-03, 2012) for global and direct solar spectral irradiance at AM1.5, respectively.

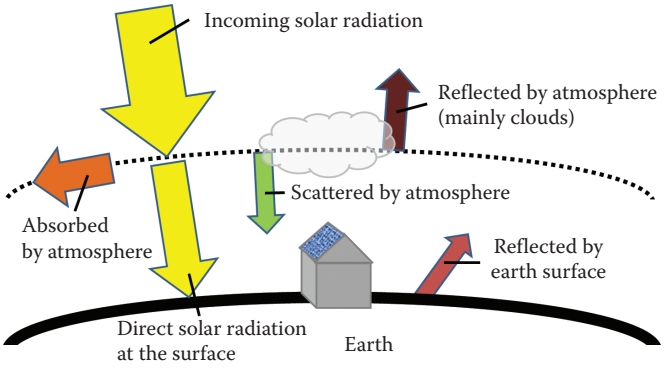

FIGURE 39.7 Schematic of solar radiation traveling through the earth's atmosphere.

global, i.e., with direct and diffuse light. The spectrum AM1.5d, which is also shown in Figure 39.6, is only the direct component of the AM1.5g and is used to measure solar cells for concentrated PV (CPV) applications.

39.4 Real Operating Conditions of Non-concentrated Solar Cells and Modules

The output power of a solar system, typically expressed in "Watt-peak" (W_P), and the efficiency of a solar system or solar components like solar cells or solar modules are determined by a standardized measurement of the current–voltage (I-V) characteristic under the "Standard Test Conditions" (STCs), i.e., under a light intensity of 1000 W/m^2 (one-sun), an ambient temperature of 25°C and the standardized spectrum AM1.5g (ASTM G173-03, 2012) which may altogether correspond approximately to a very sunny summer day in the morning on a 37° sun facing tilted surface with low ambient temperature.

Unfortunately, those environmental conditions do not occur very often. During the year and day-time deviations from the reference efficiency under STC are observed and the efficiency of the solar system changes due to a severe number of effects:

- Efficiency decrease due to module heat generation above 25°C
- Efficiency decrease due to low light/partial load condition (e.g., cloudy weather)
- Reflection losses due to incident light under increased angle of incidence (AOI) between the module and the sun
- Decreased or increased efficiency due to the change of the spectrum of the incident light
- Mismatch loss due to interconnection of cells in a module and solar modules in a PV system with different *I*-*V* characteristics
- Losses due to soiling and snow
- Losses due to shading
- Losses due to interconnection and (bypass) diode failures
- MPPT failure of the MPP tracker
- Conversion losses and self-consumption of the inverter (stand-by loss)
- Energy yield loss due to differences in real, measured and labeled module power
- Degradation, i.e., efficiency decrease over time due to environmental stress load

The effect of efficiency degradation of solar modules happens mainly due to mechanical stress and moisture resulting in corrosion, discoloration, delamination, breakage, and cracking cells (Jordan and Kurtz, 2013), but also due to solar cell degradation resulting from the formation of defects within the silicon wafers under illumination, e.g., the boron-oxygen-related defect (Schmidt et al., 1997). Another degradation mechanism may occur in solar modules under high potential (Pingel et al., 2010), which is called potential-induced degradation (PID). PID results in a low shunt resistance of the module and can decrease the output power down to zero. Nevertheless, the solar module lifetime exceeds 20–25 years maintaining typically at least 80% of its initial efficiency or for a glass-glass module even more than 30 years. Thus a solar module is one of the longest lasting products for electricity generation that does not need maintenance.

A parameter that considers all the real world influences on a PV system is the performance ratio *PR*, as defined by

$$PR = \frac{\text{real energy yield}}{\text{ideal energy yield}} = \frac{W_{\text{SP}}}{H_{\text{sun}} \times A_P \times \eta_{\text{STC}}} \tag{39.1}$$

PR is defined as the actual energy yield of a PV system W_{SP} related to the incoming sun irradiation H_{sun} on the array area A_{P} converted by the solar system efficiency under STC η_{STC} over a time interval of months/years. Also a wide spread parameter for PV system assessment is the specific or final system yield Y_{f} in kWh/kWp, as defined by Equation 39.2, which allows to compare the energy production for different systems at similar locations. Typical values are in the range of 500–2500 kWh/kWp.

$$Y_{\text{f}} = \frac{W_{\text{SP}}}{P_{\text{syst}}} \tag{39.2}$$

39.4.1 Irradiation

The frequency with which STC applies in real world is negligible. For a better understanding of which environmental conditions are more typical for a solar system, exemplarily three geographical locations are chosen for further visualization within this chapter, which are situated in middle-Europe (Halle, Germany), southern-Europe (Almeria, Spain) and in tropical climate close to the Equator (Brazzaville, Republic of the Congo).

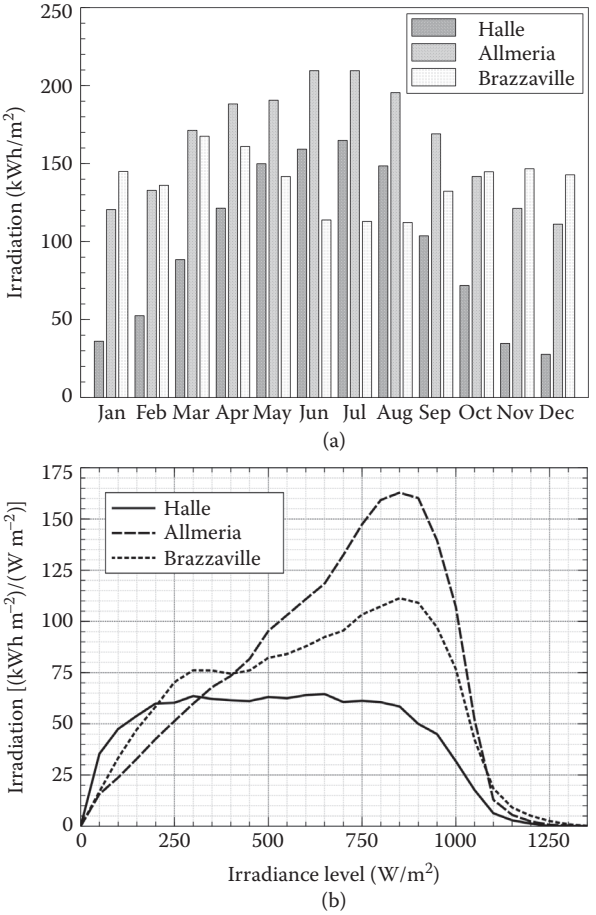

FIGURE 39.8 (a) Comparison of different irradiation conditions (Meteotest, 2008) per month, and lower graph versus in-plane irradiance level. (b) A histogram in steps of 50 W/m^2 per irradiance level bin where the frequency is multiplied by the mean irradiance of each bin.

An overview about the applying in-plane irradiation conditions, i.e., irradiation on the tilted surface, are shown in the upper graph of Figure 39.8 per month and in the lower graph versus irradiance level.

The geometric considerations concerning sun light hitting a tilted surface may be found in references (Duffie and Beckman, 2013; Quaschning, 2005). The yearly solar energy yield, i.e., in-plane irradiation, is 1159 kWh/m^2 for Halle, 1961 kWh/m^2 for Almeria and 1657 kWh/m^2 for Brazzaville for tilted surfaces having tilted angles of 30° for Halle and Almeria, and 10° for Brazzaville, respectively.

Meteorological data may be found worldwide at NASA Surface Meteorology and Solar Energy Data Set (NASA, 2016) and using the software Meteonorm (Meteotest, 2008), for EU at PVGIS EU Joint Research Centre (JRC, 2016), and for US at National Renewable Energy Laboratory (NREL, 2016).

39.4.2 Temperature

The illumination of solar modules does not only generate electricity, but also heat. The main reasons are the heat generation of the solar cells due to thermalization of charge carriers and parasitic absorption within the module of photons with higher wavelength than the bandgap of the semiconductor. A typical solar cell operating temperature increases linearly with the in-plane irradiance G_I above ambient temperature.

As commercial solar modules have a very similar structure as shown in Figure 39.5, they have comparable thermal properties and heat in a similar way. The thermal properties are measured in a standardized way determining the "Nominal Operating Cell Temperature" (NOCT) (IEC 61215, 2005) of the solar module under 800 W/m^2 light intensity, 20°C ambient temperature, AM1.5g spectrum, and 1 m/s wind velocity, which results in typical NOCT = 45°C ± 3°C.

For a better understanding of the system efficiency losses due to different operating temperatures deviating from 25°C (STC), an annual average operating temperature T_{ann} is calculated from the meteorological data set shown in Figure 39.8. The solar cell operating temperature T_{Cell} is calculated by (Ross and Smokler, 1986)

$$T_{\text{Cell}} = T_{\text{amb}} + (\text{NOCT} - 20°\text{C})\frac{G_I}{800\frac{\text{W}}{\text{m}^2}} \tag{39.3}$$

This model represents a standard roof top PV system installation. For an open-rack mounting condition lower T_{Cell} can be expected whereas for building integrated installations higher T_{Cell} is expected (Garcia and Balenzategui, 2004).

Assuming that a PV system only depends on irradiation and different ambient temperatures, the annual average (irradiance weighted) cell operating temperature T_{ann} is calculated by

$$T_{\text{ann}} = \frac{\int_{1a} P(t) \times T_{\text{Cell}}(t)\mathrm{d}t}{\int_{1a} P(t)\mathrm{d}t} = \frac{1}{W_{\text{SP}}} \int_{1a} P(t) \times T_{\text{Cell}}(t)\mathrm{d}t \tag{39.4}$$

Exemplarily, the monthly average operating temperature is calculated and shown in Figure 39.9, assuming a *NOCT* of 45°C and applying Equations 39.3 and 39.4 to the meteorological data sets.

The annual average cell operating temperature for the selected locations are those of 32.0°C for Halle, 42.9°C for Almeria, and 47.7°C for Brazzaville, respectively. Hence Almeria and Brazzaville are close to the *NOCT* = 45°C, which confirms that NOCT is a relevant environmental condition for many locations. Assuming that the relative temperature coefficient of the efficiency of a typical c-Si solar cell is $-0.4\%_{\text{rel}}/°\text{C}$, an energy yield loss compared to STC may be calculated based on the annual average cell operating temperatures. For Halle $(32.0°\text{C} - 25.0°\text{C}) \times -0.4\frac{\%_{\text{rel}}}{°\text{C}} = -2.8\%_{\text{rel}}$ may be expected, for Almeria $-7.2\%_{\text{rel}}$, and for Brazzaville $-9.1\%_{\text{rel}}$, respectively.

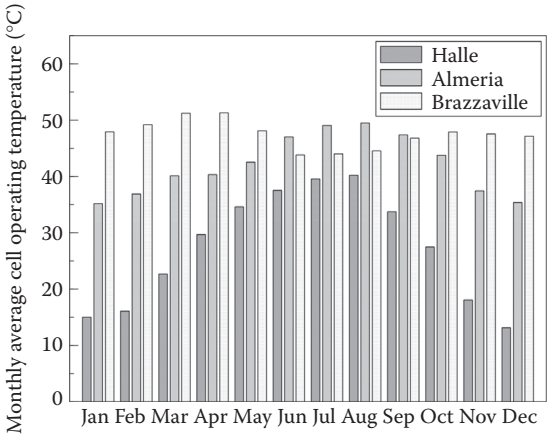

FIGURE 39.9 Monthly average (irradiance weighted) cell operating temperature per month for the locations Halle, Almeria, and Brazzaville.

39.5 Solar Cell Optics

In the field and under real operating conditions, solar cell's optical behavior is mainly determined by the optics of a solar module. In more detail, it strongly depends on the optics of the first medium and its surface structure facing the air (Winter et al., 2014), which is typically the solar glass. During year and day time, sun light appears under different oblique incidence to the module. An angle of incidence (AOI) θ can be defined, which is the angle between the sun's rays and the module normal. Light reaching the module's surface at an angle is spread out over a larger area, thus reduces the (direct) irradiance on the tilted plane following a cosine relation, as shown in Figure 39.10, described by:

$$G_{\text{direct}}(\theta) = G_{\text{direct}}\left(0^\circ\right) \times \cos(\theta) \tag{39.5}$$

The first interface air–glass may be simply described by the Fresnel equations and Beer–Lambert–Bouguer and Snell's laws (Duffie and Beckman, 2013), which are in general wavelength dependent. The angle of refraction θ_r is therefore

$$\theta_r = \arcsin\left(\frac{n_1}{n_2}\sin(\theta)\right) \tag{39.6}$$

where $n_1 = 1$ and $n_2 = 1.52$ are the refractive indices of air and solar glass, respectively. The angle-dependent transmission of such system is approximated by Equation 39.7 considering both reflective losses at the interface and absorption losses within the glass (Duffie and Beckman, 2013).

$$\left(A_{\text{opt}}T_{\text{opt}}\right)(\theta) = A_{\text{opt}}\left(1 - R_{\text{opt}}\right) = e^{-(KL/\cos\theta_r)}\left[1 - \frac{1}{2}\left(\frac{\left(\sin\left(\theta_r - \theta\right)\right)^2}{\left(\sin\left(\theta_r + \theta\right)\right)^2} + \frac{\left(\tan\left(\theta_r - \theta\right)\right)^2}{\left(\tan\left(\theta_r + \theta\right)\right)^2}\right)\right] \tag{39.7}$$

where $K = 4\ \text{m}^{-1}$ may be assumed for the extinction coefficient and $L = 2$ mm for the thickness of the glass (Duffie and Beckman, 2013).

The calculation of the product of optical absorption and transmission through the solar glass ($A_{\text{opt}}T_{\text{opt}}$) approximates here the angle-dependent losses of the solar module, which is defined as the incident angle modifier (IAM). However, *IAM* accounts for all angular and spectral losses due to increased absorption

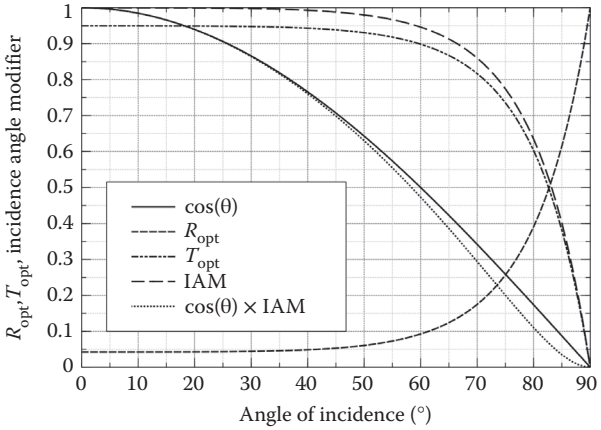

FIGURE 39.10 Cosine, optical reflection, and transmission ($n_1 = 1$ for air and $n_2 = 1.52$ for glass), incidence angle modifier (IAM), angle-dependent losses of a solar module versus angle of incidence.

within the EVA and for geometrically different, wavelength-dependent light propagation under oblique incident, i.e., multiple reflections at the cell or the backsheet. *IAM*, also shown in Figure 39.10, is used to correct the incoming direct irradiance G_{direct} on the tilted module plane as derived in Equation 39.8. The indirect irradiance component may be corrected in a similar approach.

$$G_{\text{direct}}(\theta) = G_{\text{direct}} \times \cos\theta \times IAM(\theta) = G_{\text{direct}} \times \cos\theta \times \frac{(A_{\text{opt}} T_{\text{opt}})(\theta)}{(A_{\text{opt}} T_{\text{opt}})(0°)} \tag{39.8}$$

Calculations for the angular-dependent irradiation losses over 1 year using Equations 39.5 through 39.8 and the meteorological data sets, as shown in Figure 39.8, result in $-1.3\%_{\text{rel}}$ for Halle, $-1.3\%_{\text{rel}}$ for Almeria, and $-0.9\%_{\text{rel}}$ for Brazzaville.

Latest developments of solar glass lead more and more to the introduction of ARCs and partially deeply structured (textured) glass, which improves the module optics.

The same concepts are also applied for solar cell optic optimization. ARCs for solar cells, as drawn in Figure 39.11, are typically applied as a single layer of silicon-nitride SiN_X films of about 75 nm thickness and a refractive index of about 2, whereas its optimum is slightly lower against air and slightly higher if optimized for the use in a module.

Those thin film layers may be well modeled by the transfer matrix method, which considers oblique incidences, absorbing media, coherent light, and even assemblies of thin films (e.g., double ARC) (Macleod, 2010). The second optimization strategy is to apply a texture on the front side of the solar cell, which increases the chance that reflected light reaches again the solar cell surface a second or even more times as shown in Figure 39.11. Most relevant textures for industrial application are random pyramids for monocrystalline (Cz) cells and isotexture, which forms bowl-like structures, for mc cells. Both are produced by wet chemical etching. Additional benefits of texturing arise because light is coupled into the silicon under an angle. First, this increases the chance of a photon to be absorbed close to the surface, second the path length for absorption is increased within the semiconductor and thirdly such photons may reach the rear side under the critical angle of total internal reflection. This internal light trapping may be even improved further if the rear side has properties of a back reflector, as reflected light of wavelengths close to the bandgap energy gets the chance to be absorbed on the way there and back (at least double length) within the semiconductor. Those back reflectors are, e.g., full area Aluminum layers for a BSF solar cell or dielectric layer stacks for PERC solar cells. Note that semiconductors having a direct bandgap (e.g., GaAs, CIGS, CdTe) may be fabricated with much smaller absorber thickness compared to, e.g., silicon with an indirect bandgap, and are called thin film solar cells.

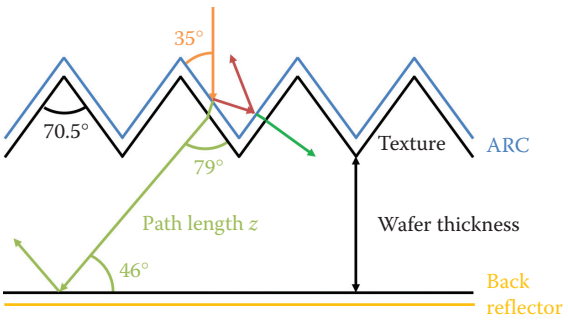

FIGURE 39.11 Schematic of light trapping within a solar cell with antireflection coating (ARC) and a random pyramid texture. Note that the schematic is not to scale, as a typical side length of such pyramid is about 5 μm and a wafer thickness is about 200μm.

39.5.1 Photogenerated Current

A way to model light-induced photocurrent of solar cells, as shown in Figure 39.11, is to use ray-tracing, which is a method to calculate the path of a photon through the solar cell considering its geometry. The light propagation may be calculated at surfaces and interfaces as necessary, e.g., using Fresnel equations. Raytracing software for solar cells are, e.g., Sunrays (Brendel, 1994), TCAD Sentaurus Device (Synopsys, 2016) and for solar modules, e.g., Daidalos (Holst et al., 2013). Also wide spread in the PV community is the usage of the simulation tool PC1D (Rover et al., 1985), which can also analytically calculate solar cell optics with texture and ARC.

The absorption of photons within the active semiconductor leads to the creation of electron–hole pairs, i.e., to a current generation. This photogenerated current density j_{gen} is not influenced by front contact shading and recombination losses and hence can be treated as the maximum current potentially extracted from the solar cell described by

$$j_{\mathrm{gen}} = -q \int_0^\infty A_{\mathrm{opt}}\left(E_{\mathrm{ph}}\right) \mathrm{d}j_{\mathrm{ph}}(E_{\mathrm{ph}}) \tag{39.9}$$

where q is the elementary charge, E_{ph} is the photon energy and j_{ph} is the photon flux. Assuming $A_{\mathrm{opt}}(E_{\mathrm{ph}} \geq E_{\mathrm{g}}) = 1$, Equation 39.9 solves to

$$j_{\mathrm{gen}} = -q \int_{E_{\mathrm{g}}}^\infty \mathrm{d}j_{\mathrm{ph}}(E_{\mathrm{ph}}) \tag{39.10}$$

The optical absorption in silicon is described by its absorption coefficient α which is shown in Figure 39.12.

The maximal achievable photogenerated current density for c-Si is about 46 mA/cm^2 for AM1.5g reference spectrum. Photons with more energy than the bandgap energy E_{g} quickly loose this "extra" energy to the lattice atoms in a process called thermalization. Photons with lower energy than E_{g} are not able to excite any electron from valence to conduction band and are not absorbed under the creation of an electron–hole pair.

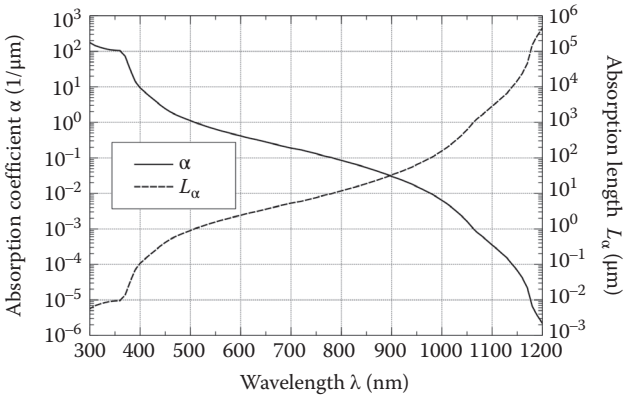

FIGURE 39.12 Absorption coefficient α and absorption length $L_\alpha = 1/\alpha$ for silicon at 300 K.

39.6 Conversion into Electrical Energy

The conversion of the optical energy of the sun light into electrical energy can be regarded in general in terms of particles in a two-level energy system. The particles are excited from their ground state to an excited state by illumination and carry the energy of the photon in the excited state. In general, the excited particles need to be extracted from the absorber into an external electrical circuit; otherwise recombination takes place which is the reverse reaction to excitation, also called generation. This is realized by the usage of selective contacts which only allow the excited particle to flow out at one side of the absorber, preventing the "overflow" of the absorber, which is the driving force. The excited particles flow into the external electrical circuit to perform work and then flow back into the absorber as relaxed particles through the other contact.

A possible realization for selective contacts is to create oppositely situated highly n- and p-type doped regions in the absorbing semiconductor for the respective contact, which is shown in the schematic of an illuminated semiconductor-based solar cell in Figure 39.13. Under illumination-free electrons and holes (excess charge carriers), introduced as excited particles, are generated in the conduction band and valence band, respectively. Thus the quasi-Fermi-levels of electrons E_{Fn} and holes E_{Fp} split in the semiconductor describing the energy distribution of electrons and holes by Fermi–Dirac statistics within the conduction and valence band, respectively. The Fermi levels of the majority carriers join at the surfaces at the respective metal contacts because the surface states exchange electrons very easily with the metal and maintain the same potential. The opposing doping at each contact results in three main effects. First, low minority carrier densities occur at the contacts due to the law of mass action and hence lead to almost zero contact recombination (selective contact). In consequence, only one kind of charge carrier is extracted. Second, it can result in ohmic contact behavior with sufficiently high doping close to the contact, which leads to low resistive losses at the contacts. Third, the creation of p-type and n-type regions in the semiconductor leads to the creation of a p-n junction with a SCR and a built-in potential, also called diffusion voltage, which is typically thought to be a necessity for solar cells, but it is not the "engine" of a solar cell (Wuerfel, 2005). Both carrier types are continuously created under illumination, and, e.g., under short-circuit condition electrons and holes flow out at their respective selective contacts as the reverse direction is blocked for them. Thus an excess of charge carriers is produced and the majority of excess charge carriers do not recombine before they diffuse along their concentration gradients. This creates a charge current through the solar cell. As charge current flows are the result of potential gradients, the behavior of solar cells is described below in this section in terms of potentials in more detail. Under an external forward bias, the difference of E_{Fn} in the n-type region and E_{Fp} in the p-type region correspond to the external voltage V, so that

$$E_{\mathrm{Fn}} - E_{\mathrm{Fp}} = qV \tag{39.11}$$

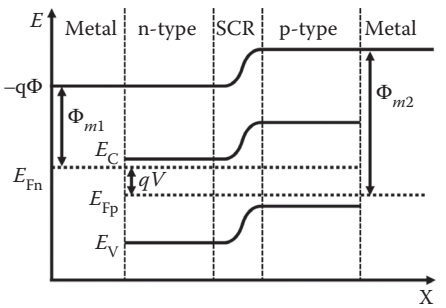

FIGURE 39.13 Schematic of the band diagram of an illuminated solar cell with two selective contacts with their respective metal work functions Φ_{m} separated by a p-n junction with a space-charge region (SCR) and the electrostatic potential Φ.

where q is the elementary charge. The forward bias reduces the built-in potential of the p-n junction. The driving forces of charge currents in a solar cell are the gradients of the quasi-Fermi-levels E_{Fn}/dx and E_{Fp}/dx, which are the gradients of the electrochemical potentials for electrons and holes, respectively. The electrochemical potential of electrons, which is constant in thermal equilibrium, is determined by the chemical potential $E_C - E_{Fn}$ and the electrostatic potential $-q\Phi$:

$$E_R - E_{Fn} = \left(E_C - E_{Fn}\right) - q\phi \tag{39.12}$$

where E_R is the reference potential. The absolute value of E_R is not relevant as only gradients of the potentials are of interest. The chemical potential describes the diffusion aspect of carrier transport due to a gradient in the carrier concentration whereas the electrostatic potential describes the force acting on charge carriers due to coulomb interaction and it is the drift aspect of charge carriers transport due to an electric field. Those two transport mechanisms have to be considered together, because diffusing mobile electrons and holes inherently carry a charge.

39.7 Modeling of Ideal *I–V* Characteristics

A PV solar cell can in first-order approximation be described in an equivalent circuit model as a photocurrent generator in parallel with a diode. It corresponds to the single diode equation which will be derived below, in a similar way to the reference (Green, 2003). In order to derive the model, the following assumptions are made: infinite thickness of the absorber, negligible contact resistances, zero contact recombination, zero surface recombination, low injection condition, i.e., the excess charge carrier density is much smaller than the doping density ($\Delta n \ll N_{dop}$). Under open-circuit condition, there is not a charge carrier flow (no current) into the external circuit resulting in the recombination of all generated charge carriers in the solar cell. Thus the diode equation can be determined from the net recombination rate. The continuity equation for electrons in one dimension under steady-state condition is

$$q\frac{\partial n}{\partial t} = 0 = \nabla \cdot j_n + q\left(G_n - U_n\right) = \nabla \cdot j_n + qG - q\frac{\Delta n}{\tau_n} \tag{39.13}$$

where n is the electron carrier density, j_n is the electron current density, G_n is the generation rate, U_n is the recombination rate and τ_n can be interpreted as the average time after which an excess minority carrier recombines. Solving Equation 39.13 for j_n leads to

$$j_n = q\int_0^{L_n}\left[G_n - q\frac{\Delta n}{\tau_n}\right]\mathrm{d}x = qG_nL_n + q\int_0^{L_n}\frac{\Delta n}{\tau_n}\mathrm{d}x \tag{39.14}$$

where L_n is the diffusion lengths for electrons. The relation between carrier lifetime and diffusion length for electrons is determined by the electron diffusion coefficient D_n:

$$\tau_n = \frac{L_n^2}{D_n} \tag{39.15}$$

Respective equations are derived for holes as minority carriers relevant for the n-type region. The excess charge carrier density Δn may be derived from the law of mass action at all dopant and injection conditions from

$$np = \left(N_D + \Delta n\right)\left(N_A + \Delta n\right) = n_{i,\,eff}^2 e^{\left(\frac{E_{Fn}-E_{Fp}}{k_B T}\right)} = n_{i,\,eff}^2 e^{\left(\frac{qV}{k_B T}\right)} \tag{39.16}$$

where p is the hole concentration, N_D is the donor concentration, N_A is the acceptor concentration, k_B is the Boltzmann constant, T is the temperature and $n_{i,eff}$ is the effective intrinsic carrier density (Altermatt et al., 2003). The latter depends on the intrinsic carrier concentration n_i and the bandgap narrowing (BGN) ΔE_g, which is temperature-, doping- and injection-dependent (Schenk, 1998) and is defined by

$$n_{i,eff}^2 = n_i^2 e^{\left(\frac{\Delta E_g}{k_B T}\right)} \tag{39.17}$$

For a p-type semiconductor under low-level injection, i.e., $\Delta n \ll N_A$, Equation 39.16 reduces to

$$\Delta n \approx \frac{n_{i,eff}^2}{N_A}\left(e^{\left(\frac{qV}{k_B T}\right)} - 1\right) \tag{39.18}$$

Substituting Equations 39.15 and 39.18 in Equation 39.14 for electrons and holes and integrating over x from L_n to L_p computes to the total current density:

$$j(V) = j_n + j_p = qG\left(L_n + L_p\right) + \left(\frac{qn_{i,eff}D_p}{N_D L_p} + \frac{qn_{i,eff}D_n}{N_A L_n}\right)\left(e^{\left(\frac{qV}{k_B T}\right)} - 1\right) \tag{39.19}$$

This can be simplified to the single diode equation

$$j(V) = j_{SC} + \left(j_{0e} + j_{0b}\right)\left(e^{\left(\frac{qV}{k_B T}\right)} - 1\right) = j_{SC} + j_0\left(e^{\left(\frac{qV}{k_B T}\right)} - 1\right) \tag{39.20}$$

where j_{SC} is the short-circuit current density, which differs from the photogenerated current density j_{gen}, as not all generated charge carriers are collected because of their finite diffusion lengths. The carriers recombine before they reach the p-n junction. The saturation current density j_0 can be separated in the contribution of the n-type region called emitter saturation current density j_{0e} and in the contribution of the p-type region called base saturation current density j_{0b}. A more realistic saturation current density of the base for a finite thickness of the solar cell and a non-zero surface recombination is defined by (Goetzberger et al., 1998)

$$j_{0b} = \frac{qn_{i,eff}D_n}{N_A L_n} G_{Fn} \tag{39.21}$$

where G_{Fn} is the geometry factor considering the effective surface recombination velocity of the electrons S_n and the thickness of the base (p-type region) W_p:

$$G_{Fn} = \frac{\cosh\left(\frac{W_P}{L_n}\right) + \frac{D_n}{S_n L_n}\sinh\left(\frac{W_P}{L_n}\right)}{\frac{D_n}{S_n L_n}\cosh\left(\frac{W_P}{L_n}\right) + \sinh\left(\frac{W_P}{L_n}\right)} \tag{39.22}$$

In Figure 39.14, an ideal solar cell I-V and power characteristic are shown, which are calculated using Equation 39.20 assuming $j_{SC} = -40$ mA/cm^2, $j_0 = 100$ fA/cm^2, and $T = 25°$C.

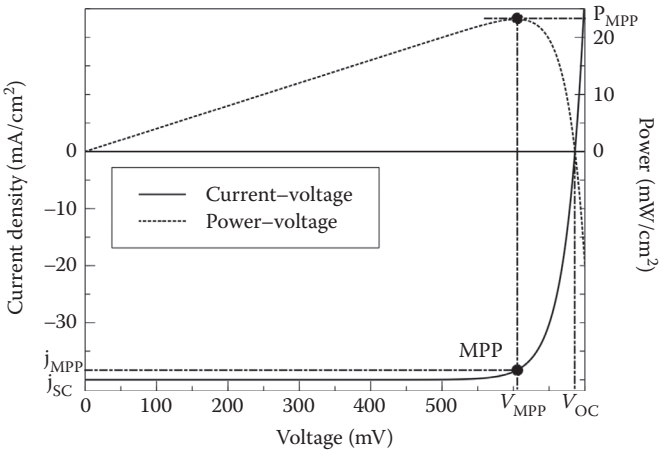

FIGURE 39.14 Example of solar cell current–voltage (I–V) and power characteristics. (MPP, maximum power point.)

39.7.1 *I–V* Parameters and Quantum Efficiency

In Figure 39.14, important output parameters of I–V and P–V characteristics can be extracted as there are: short-circuit current density $j_{SC}(V = 0\ \text{V})$, open-circuit voltage V_{OC} ($j(V_{OC}) = 0\ \text{mA/cm}^2$) and the maximum power point (MPP) ($P_{MPP}(V) = P_{max}$) with the current density and voltage at MPP j_{MPP} and V_{MPP}, respectively.

The short-circuit current density j_{SC} is the part of photogenerated current density j_{ph} which is collected by the p-n junction due to limited carrier diffusion length within the emitter and the base. A measure of this difference is the internal quantum efficiency (IQE). It is the relation between extracted charge carriers from the solar cell and generated charge carriers in the solar cell. The wavelength-dependent IQE(λ) can be derived from the measurable external quantum efficiency EQE(λ):

$$\text{IQE}(\lambda) = \frac{\text{extracted charge carriers}}{\text{generated charge carriers}} = \frac{\text{EQE}(\lambda)}{(1 - R_{opt}(\lambda) - A_{opt,par}(\lambda) - T_{opt}(\lambda))} \tag{39.23}$$

The EQE also considers optical losses due to reflection, transmission, and parasitic absorption $A_{opt,par}$ in, e.g., dielectric layers at front and rear side. It can be defined by the ratio:

$$\text{EQE}(\lambda) = \frac{\text{extracted charge carriers}}{\text{incident photons}} \tag{39.24}$$

The fill factor (FF) and the efficiency η are derived from the I–V characteristic. The FF is defined by

$$\text{FF} = \frac{j_{MPP} V_{MPP}}{j_{SC} V_{OC}} = \frac{P_{MPP}}{j_{SC} V_{OC}} \tag{39.25}$$

FF is always below 1 and is introduced to describe nonideal losses in the solar cell arising, e.g., due to series resistance losses as discussed in Section 39.8. The efficiency is calculated considering the irradiance G_I and the cell area A_{Cell}:

$$\eta = \frac{P_{MPP}}{G_I A_{Cell}} = \frac{\text{FF}\, j_{SC} V_{OC}}{G_I A_{Cell}} \tag{39.26}$$

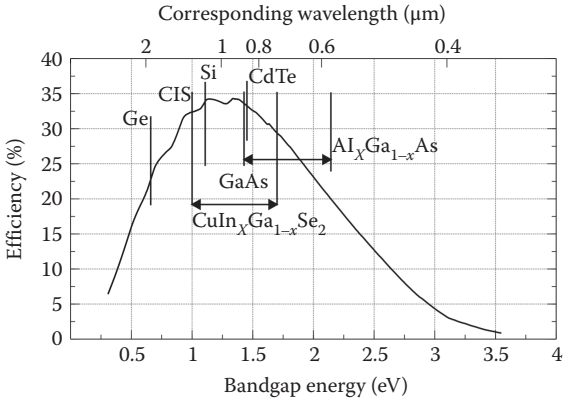

FIGURE 39.15 Dependence of a maximal achievable efficiency on the bandgap energy under an AM1.5g spectrum (1 sun) and a cell operating temperature of 300 K. Optimal bandgap energies of semiconductors are in the range of 1.0–1.5 eV. Established semiconductor absorber materials for solar cells are silicon (Si), Gallium-Arsenide (GaAs), Aluminum-Gallium-Arsenide ($Al_xGa_{1-x}As$), Germanium (Ge), CdTe, Copper-Indium-Diselenide (CIS), Copper-Indium-Gallium-Diselenide (CIGS).

39.7.2 Maximum Efficiency of Ideal Single Junction Solar Cells

For ideal single-junction solar cells, a maximal reachable efficiency may be calculated depending on their bandgap energy E_g following the reference (Wuerfel, 2005) which is very similar to the so-called Shockley–Queisser limit calculation (Shockley and Queisser, 1961). Two loss mechanisms are solely considered. First, nonabsorption losses due to photons with energy lower than the bandgap energy of the semiconductor are considered by Equation 39.10. Second, radiative recombination losses of the semiconductor occur assuming a one-sided emission of only photons of a blackbody spectrum with a temperature of 300 K higher than the bandgap energy, which is described by

$$j_{0,\mathrm{rad}} = q j_{\mathrm{ph,emission}}^{300\mathrm{K}} = q \int_{E_g}^{\infty} \mathrm{d}j_{\mathrm{ph}}^{300\mathrm{K}}(E_{\mathrm{ph}}) \tag{39.27}$$

Equation 39.27 is derived in Wuerfel (2005) assuming a constant distribution of charge carriers over the absorber thickness, large diffusion lengths, and perfected optics where all photons with the bandgap energy or lower are reabsorbed. Using the single diode model described by Equation 39.20 and substituting 39.26 and 39.27 computes a bandgap energy-dependent maximal efficiency $\eta(E_g)$ for a single p-n junction which is shown in Figure 39.15 for an AM1.5g reference spectrum.

39.8 Nonideal *I–V* Characteristics

39.8.1 Series and Parallel Resistance

Solar cells are large diodes. They suffer under real operation from resistance losses. One differentiates two resistances in solar cells: series resistance r_S and parallel resistance r_P. Note that the small letters indicate area-weighted quantities in $\Omega \cdot \mathrm{cm}^2$ similar to the current density j in $\mathrm{mA/cm}^2$. Series resistance losses arise from current transport in the front metal busbars and metal fingers as well as in front contacts, emitter, base, rear contact, and rear metallization which is visualized by a sketch of a BSF solar cell in Figure 39.16.

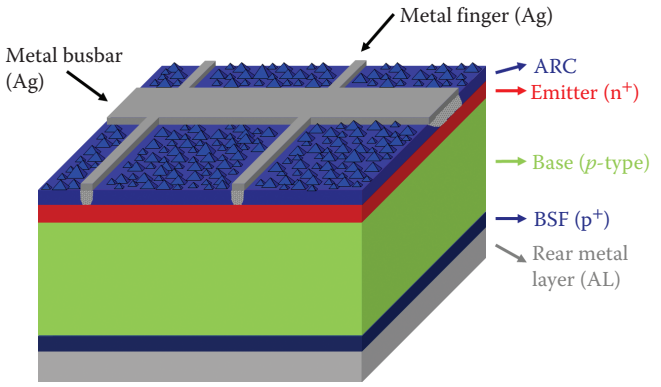

FIGURE 39.16 Sketch of back-surface-field (BSF) solar cell with an indicated random pyramid texture. The screen-printed front and rear metallization consist mainly of silver and aluminum, respectively. During solar cell fabrication, the p-type boron-doped wafer is phosphorus-diffused creating a highly n-type doped emitter. The BSF is formed by an alloying process during a fast firing step while creating the front and rear contact at the same time.

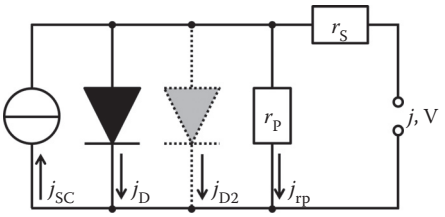

FIGURE 39.17 Equivalent circuit diagram of the single diode and double diode (dotted second diode) model of a solar cell.

As all these parts are included, r_s is called lumped series resistance. Parallel resistance losses, also called shunt resistance losses, arise when the p-n junction is short-circuited, e.g., by grown-in defects of the material (Breitenstein et al., 2004). A corresponding single diode equivalent circuit model is set up as shown in Figure 39.17.

It is described by the implicit Equation 39.28 and examples for analytically calculated *I-V* characteristics are shown in Figure 39.18.

$$j(V) = j_{SC} + j_D + j_{rp} = j_{SC} + j_0 \left(e^{\left(\frac{q(V + j(V) r_S)}{k_B T} \right)} - 1 \right) + \frac{V + j(V)\, r_S}{r_P} \tag{39.28}$$

39.8.2 Recombination

As r_S cannot be zero and r_P cannot be infinite, both reduce the FF and thus also the efficiency compared to the ideal case. The consideration of series and parallel resistance is in most cases not sufficient to fully describe *I–V* characteristics. Additional influences which are the root causes for deviations from the ideal *I–V* curve are the following:

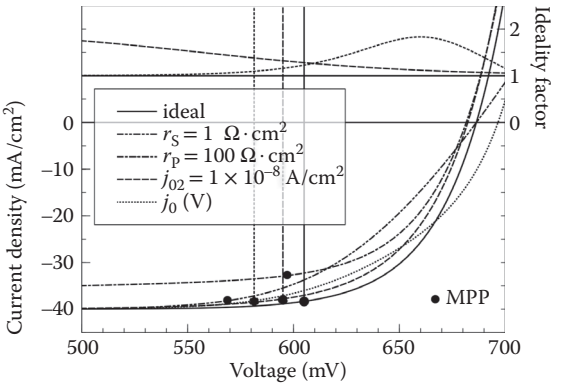

FIGURE 39.18 Solar cell I–V characteristics: ideal, suffering r_S, suffering r_P, suffering nonidealities modeled by the double diode model (compare Equation 39.30) and by a voltage-dependent saturation current density $j_0(V)$ in the voltage range from 500 mV to 700 mV, which is the relevant range for maximum power point (MPP). For details on $j_0(V)$ please refer to the text in Section 39.8.2.

- Distributed series resistance losses (Araújo et al., 1986; Fong et al., 2011) leading to a voltage-dependent change of current flow pattern in solar cells (Altermatt et al., 1996) which results in a voltage-dependent series resistance $r_S(V)$
- Injection-dependent lifetimes, i.e., voltage-dependent recombination within the volume, e.g., due to the boron-oxygen-related defect (Schmidt et al., 2001), or injection-dependent surface recombination velocities, e.g., at thermal oxides (Aberle et al., 1993)
- Edge recombination (Grove and Fitzgerald, 1966; Kuehn et al., 2000)
- Recombination in the space-charge region (SCR) (Wuerfel, 2005)

An injection-dependent lifetime within the absorber, i.e., voltage-dependent recombination, may be modeled by a voltage-dependent saturation current density $j_0(V)$ (Müller, 2016) which is typically experienced for Shockley–Read–Hall (SRH) recombination with a single defect with asymmetric capture cross sections, as, e.g., for the boron-oxygen-related defect (Schmidt et al., 1997). A $j_0(V)$ may be derived by Equation 39.29 due to a change of the injection density Δn in the solar cell with voltage as described by Equation 39.29. Assuming an excess carrier lifetime for a SRH midgap defect with the electron lifetime parameter $\tau_{n0} = 100$ µs and the hole lifetime parameter $\tau_{p0} = \tau_{n0} \times 10 = 1000$ µs for the parameters $j_{0e} = 10$ fA/cm², $N_A = 7.2 \times 10^{15}$cm^{-3}, $W_A = 160$ µm, a corresponding I–V curve is calculated with the single diode Equation 39.20 and is shown in Figure 39.18.

$$j_0(V) = j_0(\Delta n) = j_{0e} + j_{0b}(\Delta n) = j_{0e} + \frac{qW_A n_{i,eff}^2(N_A)}{N_A \tau_{eff}(\Delta n)} \tag{39.29}$$

Recombination in the SCR can be described by the double diode model which is given by

$$j(V) = j_{SC} + j_D + j_{D2} + j_{rp} = j_{SC} + j_0\left(e^{\left(\frac{q(V+jr_S)}{k_B T}\right)} - 1\right) + j_{02}\left(e^{\left(\frac{q(V+jr_S)}{2k_B T}\right)} - 1\right) + \frac{V + jr_S}{r_P} \tag{39.30}$$

Its corresponding equivalent circuit model is also shown in Figure 39.17 and an example for an I–V characteristic is shown in Figure 39.18. The second diode is analytically derived with an ideality factor $n = 2$ (Wuerfel, 2005) for recombination via defects in the space-charge region, compared to $n = 1$ for the first diode. However, SCR recombination is typically not the main contribution to nonideal behavior, i.e., the saturation current density of the second diode j_{02} is mainly a fit parameter, when applying the double diode

model to fit dark I-V characteristics of solar cells. For crystalline wafer-based silicon solar cells, the root cause is typically different and a fitted j_{02} accounts for all effects listed earlier.

Another possibility for the description of nonideal I–V characteristics is the usage of a voltage-dependent ideality factor $n(V)$ (Rhoderick and Williams, 1988), which can be calculated from dark or light I–V characteristics by Equation 39.31 if $V > 3k_BT/q$:

$$\frac{1}{n} = \frac{k_BT}{q}\frac{\mathrm{d}\left(\ln j\right)}{\mathrm{d}V} \tag{39.31}$$

Exemplarily deriving the voltage-dependent ideality factor $n(V)$, the double diode model is now compared to the voltage-dependent saturation current density $j_0(V)$ using Equation 39.29, which is shown in in the upper part of Figure 39.18. The second diode model has as expected an ideality factor of $n = 2$ for voltages below 400 mV and it decreases toward $n = 1$ almost at V_{OC}. At MPP it is in-between at $n = 1.33$ for the shown example. For $j_0(V)$ the ideality factor increases in this example from $n = 1$ to $n = 1.1$ at MPP to a maximum of about $n = 1.83$ and then decreases again toward $n = 1$ while it is $n = 1.25$ at V_{OC}.

Thus, in general, nonideal I–V behavior can be expected for any real solar cell and the $n(V)$ evaluation may give hints for its origin (McIntosh, 2001). When analyzing I–V characteristics, it is necessary to firstly derive a voltage-dependent series resistance with preferable a multiple-light-level determination method (Fong et al., 2011) because current flow patterns within a solar cell remain similar and injection-dependent effects are reduced when changing the illumination only by 10%–20%. In a second step, a series resistance free I-V curve is derived for the $n(V)$ evaluation (Steingrube et al., 2011).

As an alternative and quicker approach to analyze I–V characteristics, the *FF* can be evaluated in terms of series resistance induced and recombination-induced losses. The *FF* is compared to the pseudo fill factor (*pFF*) described by Equation 39.32 which accounts only for recombination driven *FF* changes if r_P is sufficiently high ($r_P > 10\ \mathrm{k\Omega \cdot cm^2}$).

$$\mathrm{pFF} = \mathrm{FF} - \Delta\mathrm{FF}_{rs} = \mathrm{FF} - \frac{j_{MPP}^2 r_S(V_{MPP})}{j_{SC}V_{OC}} \tag{39.32}$$

39.9 Numerical Simulation and Models for Crystalline Silicon Solar Cells

The diode model-based approaches are very often not sufficient to describe solar cells; therefore numerical device simulations are carried out. The advantage of numerical simulations compared to analytical solutions is an increased accuracy and provides a deeper insight into device physics and solar cell losses. Analytical solutions are restricted to particular assumptions such as constant photogeneration, constant doping or low injection conditions, which often only approximate real device behavior. Typical problems that demand numerical solar cell simulation are the following:

- Local line front contacts desire a two-dimensional simulation
- Local point rear contacts desire a three-dimensional simulation
- Inhomogeneous distributed wafer quality in case of mc silicon
- Injection-dependent SRH recombination in the bulk
- Injection-dependent SRH recombination at surfaces
- Consideration of Auger recombination in the emitter or BSF with their doping profiles

39.9.1 Numerical Simulation

The physics for the numerical simulation of solar cells are implemented in various simulators. Most developed and widely used simulators are PC1D (Rover et al., 1985), PC1Dmod (Haug et al., 2015)

for one-dimensional problems and TCAD Sentaurus Device (Synopsys, 2016), Atlas (Silvaco, 2016), Quokka (Fell, 2013), and AFORS-HET (Varache et al., 2015) for two and three-dimensional simulations. An in-depth introduction into numerical device simulation of crystalline silicon solar cells can be found in Altermatt (2011) which is also used as the source of the most important facts and models discussed below in this section. An update will be given below for the intrinsic recombination model and the incomplete ionization model.

For numerical solar cell simulation a coupled set of differential equations, which are the basic equations of semiconductor device physics, have to be solved by an iterative procedure. This set consists of the Poisson Equation 39.33, which describes the relationship between the electrostatic potential Φ and the charge distribution, the continuity equations for electrons and holes (Equations 39.13, 39.34, 39.35), which conserve the quantity charge carrier, and the current density equations in the drift-diffusion approach for electrons (Equation 39.36) and holes (Equation 39.37) in one dimension:

$$\nabla \cdot (\varepsilon \nabla \Phi) = -q\left(p - n + N^{+}_{\mathrm{don}} - N^{-}_{\mathrm{acc}}\right) \tag{39.33}$$

$$q\frac{\partial n}{\partial t} = \nabla \cdot j_{\mathrm{n}} + q\,(G - U) \tag{39.34}$$

$$q\frac{\partial p}{\partial t} = \nabla \cdot j_{\mathrm{p}} + q\,(G - U) \tag{39.35}$$

$$j_{\mathrm{n}} = -q\mu_{\mathrm{n}} n \nabla \Phi + qD_{\mathrm{n}} \nabla n \tag{39.36}$$

$$j_{\mathrm{p}} = -q\mu_{\mathrm{p}} p \nabla \Phi - qD_{\mathrm{p}} \nabla p \tag{39.37}$$

In Equations 39.33 through 39.37, ε is the permittivity of the material, p is the hole carrier density, N^{+}_{D} is the ionized donor concentration and N^{-}_{A} is the ionized acceptor concentration, μ_{n} and μ_{p} are the mobilities for electrons and holes, respectively and D_{p} is the hole diffusion coefficient. The mobilities and diffusion coefficients are related through the Einstein relationships $D_{\mathrm{n}} = \frac{k_{\mathrm{B}}T}{q}\mu_{\mathrm{n}}$ and $D_{\mathrm{p}} = \frac{k_{\mathrm{B}}T}{q}\mu_{\mathrm{p}}$ for a nondegenerated semiconductor. For a degenerated semiconductor (e.g., doping densities higher than 1×10^{19} cm^{-3}), Fermi–Dirac statistics are valid and the electron transport equation is extended by the additional term $-nk_{\mathrm{B}}T\nabla\left(\ln\left(\frac{n}{N_{\mathrm{C}}}e^{(E_{\mathrm{C}}-E_{\mathrm{Fn}})/(k_{\mathrm{B}}T)}\right)\right)$ and the hole transport equation is extended by $-pk_{\mathrm{B}}T\nabla\left(\ln\left(\frac{p}{N_{\mathrm{V}}}e^{\left(E_{\mathrm{Fp}}-E_{\mathrm{V}}\right)/(k_{\mathrm{B}}T)}\right)\right)$, respectively. N_{C} and N_{V} are the effective density of states for electrons in the conduction band and for holes in the valence band, respectively. At those doping densities higher than 1×10^{19} cm^{-3}, Boltzmann statistics overestimates the product of electron and hole density and may lead to a distorted loss analysis between Auger recombination losses in the highly doped region and surface recombination. Thus Fermi–Dirac statistics should be used, especially as the widely used PC1D software is updated using Fermi–Dirac statistics nowadays (Haug et al., 2015).

39.9.2 Models for Crystalline Silicon Solar Cells

A doping and carrier concentration-dependent as well as temperature-dependent mobility model (Klaassen, 1992a,b) is used to calculate Figure 39.19.

For highly doped semiconductors, the bandgap decreases, which is called bandgap narrowing (BGN). It occurs due to carrier-dopant and carrier–carrier interactions, which may be considered for silicon using a BGN model (Schenk, 1998) which is used to calculate Figure 39.20.

The effective intrinsic carrier concentration $n_{\mathrm{i,eff}}$ changes correspondingly with the doping density (Altermatt et al., 2003) using Equation 39.17 and $n_{\mathrm{i,eff}}$ is shown in Figure 39.21.

For medium-doped semiconductors in the range 1×10^{18}cm^{-3} to 1×10^{19}cm^{-3} the effect of incomplete ionization occurs, which is caused by the Fermi-level approaching the donor or acceptor level, resulting

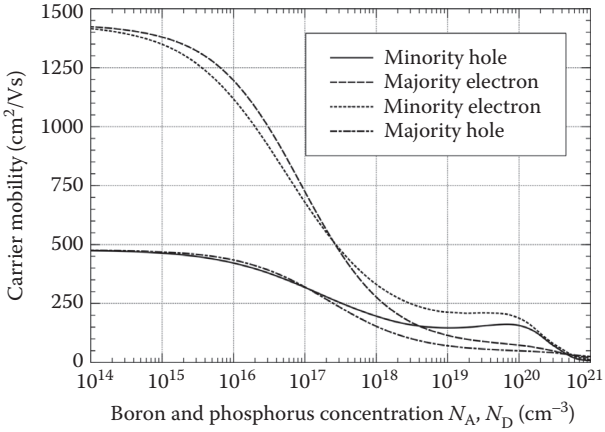

FIGURE 39.19 Klaassen's parametrization of the minority and majority carrier mobility in boron- and phosphorus-doped silicon at 300 K.

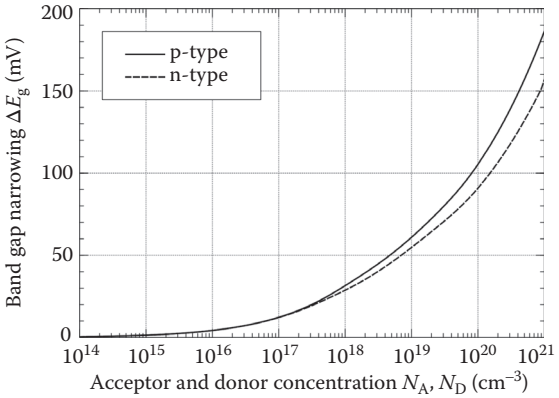

FIGURE 39.20 Bandgap narrowing in p- and n-type silicon calculated with the Schenk model at 300 K.

in the restriction of electrons or holes to localized donor or acceptor states, which prevents free electron or hole movement, respectively. The consideration of incomplete ionization is recommended for the BSF solar cell and a parametrization may be found in reference (Steinkemper et al., 2015).

In order to include the geometrical properties of the solar cell and to decrease the simulation time geometrical considerations have to be performed to derive a well-suited geometry of the device for simulations. Most solar cells have an area of about 15.6×15.6 cm^2 and are about 150–200 μm thick. Smallest features are, e.g., front contacts with a width down to 10 μm. However, it is not necessary to simulate the complete solar cell because its structure is highly symmetric. The simulation domain can be reduced to a geometrically irreducible *standard domain* as shown in Figure 39.22.

This domain is very often two-dimensional where its width is half the spacing between the front metal fingers. Note that the front surface is typically textured, which it is not accounted for in the flat device domain due to the fact that, e.g., random pyramids would result in a huge three-dimensional domain. For simulation, the device domain is discretized on a mesh as shown in Figure 39.22. The mesh should be fine

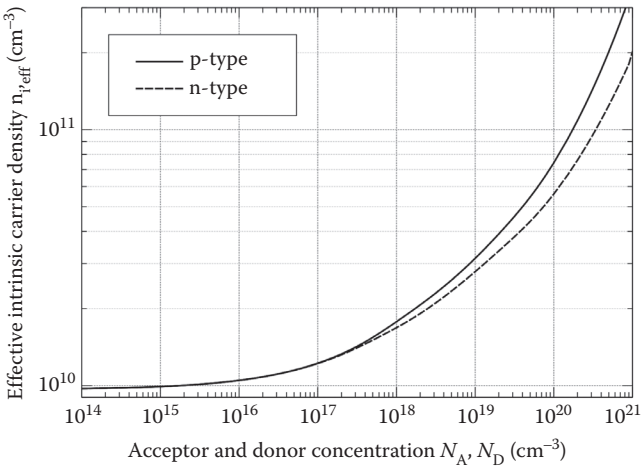

FIGURE 39.21 Influence of the doping density on the effective intrinsic carrier concentration in p- and n-type silicon at 300 K.

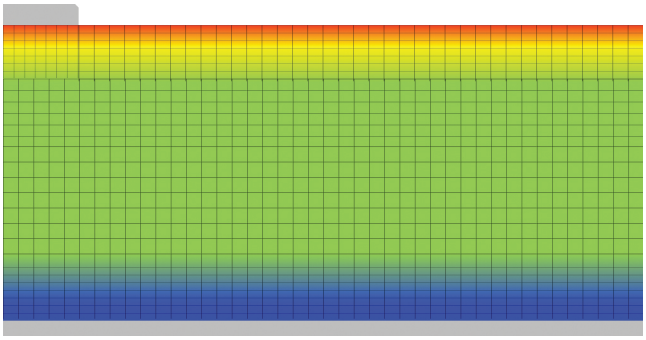

FIGURE 39.22 Geometrically irreducible two-dimensional *standard domain* for a BSF solar cell. Mesh refinements near the front surface, in diffused layers and close to contacts are indicated.

enough where any input or output parameter varies strongly with distance, which is for solar cells close to the front surface where the blue part of the sunlight is absorbed within nanometers, and where a front diffusion layer is typically situated. Besides near metal contacts the mesh should be refined. The generation of charge carriers as described in Section 39.4 is typically considered by an one-dimensional generation profile $G(x)$. The reverse process of intrinsic recombination may be included by a combined model for the radiative recombination and the Auger recombination (Richter et al., 2012), which is used to calculate Figure 39.23.

At surfaces, surface recombination occurs via surface states within the bandgap and additionally fixed charges are present which lead to local band bending (Kingston and Neustadter, 1955).

The most interesting result of a simulation is the sweep of an I–V curve which is performed changing the applied voltage at the contacts as boundary condition. For a full solar cell I–V characteristic, the standard domain I-V curves can be interconnected in a circuit by ohmic resistances, in, e.g., a SPICE network simulation, which represent the front metal fingers. A sketch of such a network is shown in Figure 39.24.

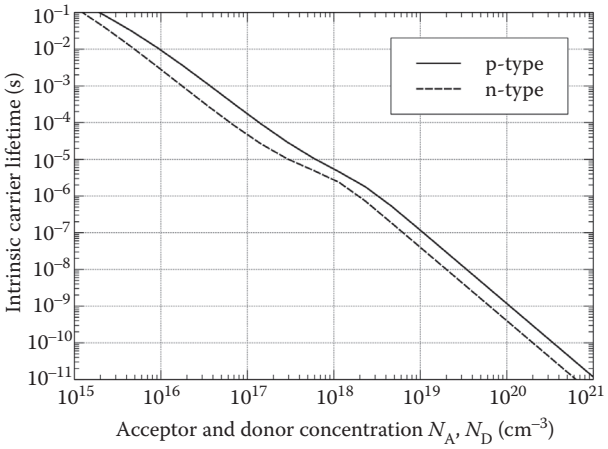

FIGURE 39.23 Influence of the doping density on radiative and Auger recombination in silicon (intrinsic recombination) in p- and n-type silicon at 300 K.

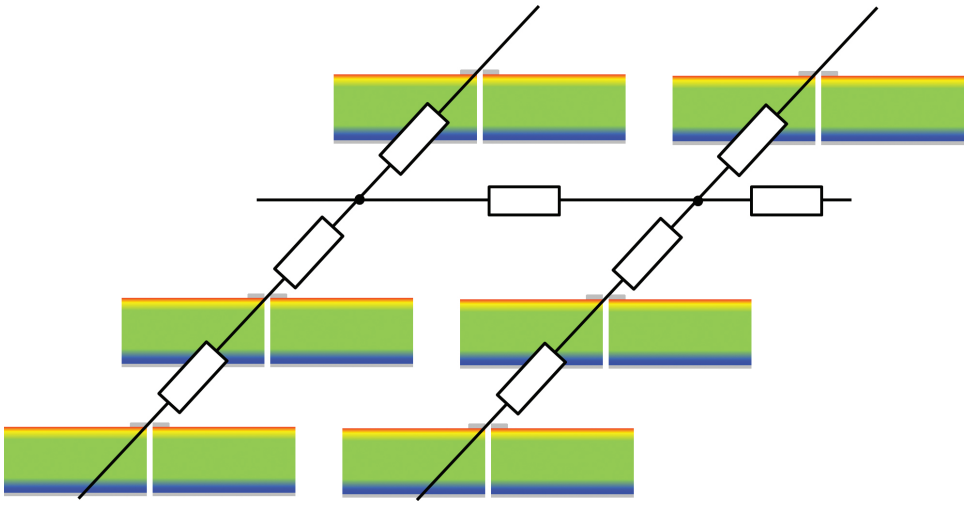

FIGURE 39.24 Sketch of a network of resistances and standard domains which allow simulating a full solar cell I–V characteristic by device and circuit simulation. Compare Figure 39.16 to the sketch of the same solar cell part.

39.10 Real Working Behavior of Solar Cells

Under real operating conditions, main changes of solar cell performance arise due to its intensity and temperature behavior.

39.10.1 Light Intensity Behavior

The intensity dependence is mainly governed by the decrease of irradiance G which results in a decrease of the short-circuit current density j_{SC}. For typical crystalline silicon solar cells under nonconcentrated sunlight conditions, the output power is mainly determined by the linear relationship $G \propto j_{SC}$. Deviations

may occur for irradiance conditions below 0.2 suns, where injection-dependent effects, e.g., injection-dependent carrier lifetime, become apparent. The general change of the open-circuit voltage V_{OC} with irradiance G may be derived from the ideal-behavior solving the single diode Equation 39.20 to V_{OC} for $j(V_{OC}) = 0$ mA/cm^2:

$$V_{OC} = \frac{k_B T}{q} \ln\left(1 + \frac{j_{SC}}{j_0}\right) \tag{39.38}$$

The change of *FF* with irradiance G can be approximated by Equation 39.39 (Green, 1982) for the ideal-behavior, which depends on V_{OC}, and is called *ideal fill factor* FF_0.

$$FF_0\left(v_{OC}\right) = \frac{v_{OC} - \ln\left(v_{OC} + 0.72\right)}{v_{OC} + 1}, \text{with} v_{OC} = \frac{q}{k_B T} V_{OC} \tag{39.39}$$

An example of a calculated *I-V* characteristic under low illumination using the single diode Equation 39.20 is shown in Figure 39.25, assuming $j_{SC} = 20$ mA/cm^2 and $j_0 = 100$ fA/cm^2. The almost halved j_{SC} results in nearly half of the maximum power.

The so-called low light behavior of a solar cell is shown in Figure 39.26 using the single diode equivalent circuit model (Equation 39.28) with j_{SC} (STC) $= -40$ mA/cm^2 and $j_0 = 100$ fA/cm^2. Additionally, the relative efficiency normalized to its STC value is calculated and compared:

$$\eta_{rel}(G, T) = \frac{\eta(G, T)}{\eta(1000\frac{W}{m^2}, 25°C)} \tag{39.40}$$

For the ideal-behavior ($r_S = 0\ \Omega \cdot$ cm^2) in Figure 39.26, it is deduced that the efficiency decreases with decreasing irradiance, which is mainly due to the V_{OC} decrease and only slightly due to the *FF* decrease. The variation of the lumped series resistance in Figure 39.26 shows that its influence governs

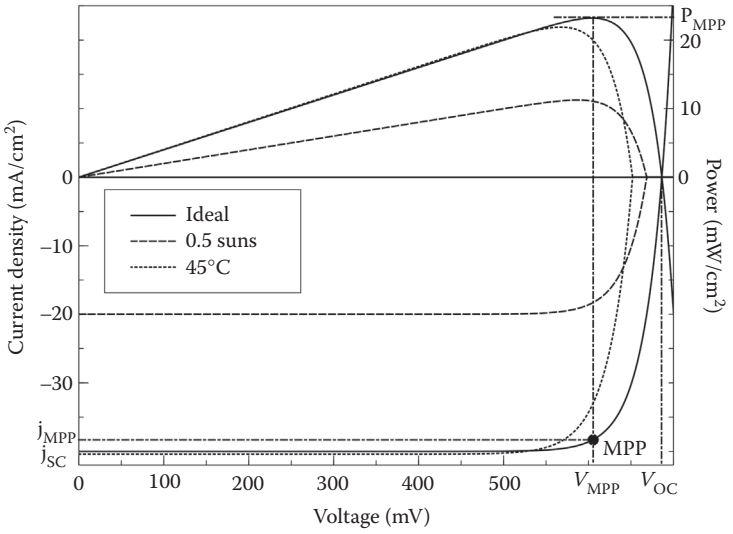

FIGURE 39.25 Comparison of solar cell I-V and power characteristics under STC, under low illumination of 0.5 suns, and under increased temperature of $T = 45°$C. (MPP, maximum power point.)

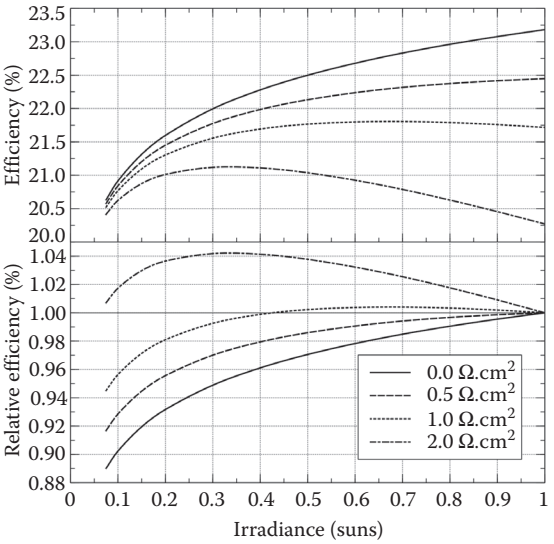

FIGURE 39.26 Relative efficiency versus irradiance for different lumped series resistances r_s, where $r_S = 0\ \Omega \cdot \text{cm}^2$ represents the ideal behavior, whereas $r_S = 0.5\ \Omega \cdot \text{cm}^2$ is a typical value for c-Si solar cells and r_S between 1 and $2\ \Omega \cdot \text{cm}^2$ is typical for solar modules due to the cell interconnection, e.g., the tabbing ribbons cause additional series resistances.

within the irradiance range from 200 W/m^2 to 1000 W/m^2. In this range, the current in the solar cell is sufficiently high and η saturates or even reduces toward STC with increasing r_S. This results in an increasing relative efficiency under low light condition for solar cells and modules with high r_S (compare lower part of Figure 39.26), which seems beneficial under real operating condition. However, the series resistance solely limits the STC efficiency (compare upper part of Figure 39.26) and the solar module will stay behind its potential. In conclusion, r_S should be low for maximum energy yield.

39.10.2 Temperature Behavior

The temperature behavior is determined by several effects (Dupré et al., 2015). It increases the effective intrinsic carrier density $n_{i,\text{eff}}$ (Misiakos and Tsamakis, 1993), which decreases the quasi-Fermi-level splitting and thus at higher temperature the voltage at the same injection density is reduced (compare Equations 39.17 and 39.18). A detrimental effect, but smaller, is the increase of the bandgap with increasing temperature, which results in a slight increase in absorbed photon current density. Thus j_{SC} is slightly increased. The change of *FF* with temperature depends on a variety of influences, e.g., temperature-dependent conductivity of the front and rear metallization, as well as front and rear contacts. For silicon, the *FF* typically decreases with higher temperature. Surprisingly, the temperature behavior of j_{SC}, V_{OC} and η is typically linearly under relevant operating conditions in the range of 0°C–75°C. Therefore, it is expressed in terms of their absolute temperature coefficients $TC_{abs}(j_{SC})$, $TC_{abs}(V_{OC})$ and $TC_{abs}(\eta)$ or more often in terms of their relative counterparts $TC_{rel}(j_{SC})$, $TC_{rel}(V_{OC})$ and $TC_{rel}(\eta)$ normalized to STC similar to Equation 39.40. The relevant temperature coefficient of the efficiency $TC_{rel}(\eta)$ is in the range of −0.5%/K to −0.25%/K for crystalline silicon solar cells. An example of a calculated *I-V* curve at 45°C using the single diode model is shown in Figure 39.25 assuming a $j_{SC}(45°\text{C}) = 40.4\ \text{mA/cm}^2$ and $j_0\ (45°\text{C}) = 191\ \text{fA/cm}^2$. This corresponds to a $TK_{rel}(j_{SC}) = 0.05\%/\text{K}$, i.e., $\Delta j_{SC}\,(\Delta T) = j_{SC}\,(25°\text{C}) \times TC_{rel}\left(j_{SC}\right) \times \Delta T$, and a j_0 increase which is approximated by the proportionality to $T^3 \times e^{-\frac{E_g(T)}{k_B T}}$ (Wolf et al., 1977). The increased temperature leads to decreased V_{OC} and thus to slightly lower P_{MPP}.

39.10.3 Energy Yield

The real working behavior of solar cells and solar modules, respectively, is regarded in terms of their delivered energy yield below to complete the information given in this chapter.

The main influences on solar cell performance under real operating conditions are changing irradiance, cell temperature, angle of incidence, and spectral variations compared to the STC condition. The spectral variation is below 1% for c-Si solar cells (Huld et al., 2011) and hence will not be considered in the further discussion. For a typical solar module, the energy yield is calculated for the three locations introduced in Section 39.3 using the Equations 39.3, 39.5 through 39.8, 39.26 and 39.28 assuming c-Si solar module relevant parameters $j_0 = 100$ fA/cm^2, $r_S = 1\Omega \cdot cm^2$, $r_P = \infty$, $TC_{rel}(\eta) = 0.4\%/°C$, $K = 4$ m^{-1} and $L = 2$ mm and $NOCT = 45°$C. The specific yields per month are shown in the lower part of Figure 39.27 using Equation 39.2 and the specific yields over 1 year are $Y_f = 1120$ kWh/kWp for Halle, $Y_f = 1823$ kWh/kWp for Almeria, and $Y_f = 1504$ kWh/kWp for Brazzaville.

In a first-order approximation, the specific yield follows the available solar irradiation as shown in Figure 39.8 which results, e.g., in low Y_f for winter months in Halle or similar Y_f over the whole year for Brazzaville near the equator. To separate the irradiation influence, the performance ratio has been introduced in Equation 39.1 and here *PR* is shown in the upper part of Figure 39.27. It is mostly below 100% due to increased cell operating temperature above 25°C in summer months for all locations and due to the warmer climate in Almeria and Brazzaville. Performance ratios over 1 year are $PR = 96.6\%$ for Halle, $PR = 92.9\%$ for Almeria and $PR = 90.8\%$ for Brazzaville, which are very similar to the estimated temperature losses derived from the annual average cell operating temperature in Section 39.3.

The calculations are carried out assuming no additional cell-to-module efficiency changes while module series resistance and optics under oblique incidence are included. Additional *PR* losses arise in reality due to losses in the inverter (up to 5%), due to cell-to-cell and module-to-module mismatches (about 1%), due to shading (depends on location), thus good *PR* are in the range of 85%–90% in the first years. Additionally, efficiency degradation occurs over time of about 0.5%$_{rel}$/year (Jordan and Kurtz, 2013), which further decreases the *PR* of about 6.25% over the system lifetime of about 25 years.

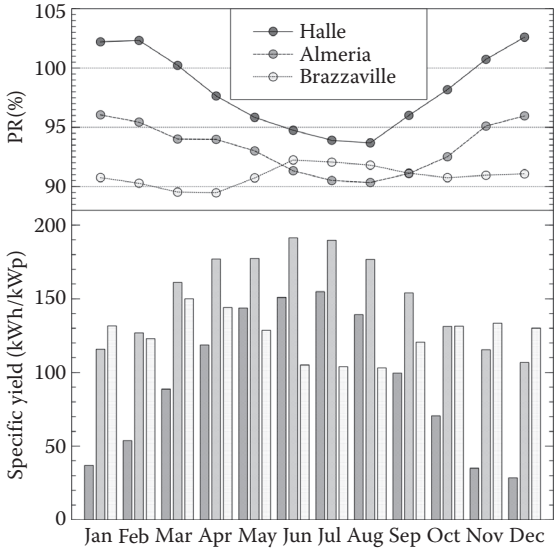

FIGURE 39.27 Specific yield (lower graph) and performance ratio (upper graph) per month for the three locations Halle, Almeria, and Brazzaville.

39.11 Conclusion and Outlook

PV systems became already an integral part of our world in terms of energy supply and also in our landscapes. The levelized cost of electricity (LCOE) of PV is approaching that of conventional electricity power plants (IEA, 2014). Thus, a major share of electricity is expected to be PV electricity in the next decades. However, it is a fluctuating energy resource and storage is getting more and more important.

Wafer-based silicon solar cells became the market leading technology and at the moment a change is unlikely to happen. The competitive thin-film solar cells on CIGS and CdTe absorber material benefit from reduced material consumption, but are still built in solar module with front glass for long lifetime which nowadays lowers significantly the inherent cost advantage. Organic and dye-sensitized (inorganic) thin-film absorbers are not showing a high efficiency potential yet (Green et al., 2015).

New solar cell structures have to be designed right away in cost competitive solar modules with a long lifetime. For example, the recently emerging perovskite-based solar cells (Green et al., 2014) have shown an incredible efficiency increases in the past 5 years, but they still lack stability against UV light and will face similar module cost disadvantage as thin-films solar cells do. Efficiency matters, but costs and stability have to be competitive in order to gain economic success.

Besides the single-junction solar cells for 1 sun illumination, there are several other developing technologies and approaches. Concentrating photovoltaics (CPV) is close to enter the market. CPV uses optical elements to increase the irradiance on a smaller solar cell size for increased efficiency and reduced material consumption. However, for higher concentration (>20 suns) the PV system needs to track the sun, which results in higher complexity and higher maintenance requirements. Increased efficiency of solar cells may be reached by so-called third generation PVs (Green, 2003). Examples are multijunction solar cells which are stacks of semiconductor absorber materials of different bandgap energies reducing thermalization losses which have shown great potential (Green et al., 2015), whereas hot carrier solar cells, multiband solar cells, and photon up- and down-converting materials are still in focus of basic research.

Acknowledgments

I would like to thank Eric Schneiderlöchner, Pietro Altermatt, Byungsul Min, and Julia Müller for their great support.

References

Aberle A G, Robinson S J, Wang A, Zhao J, Wenham S R and Green M A (1993) High-efficiency silicon solar cells: Fill factor limitations and non-ideal diode behaviour due to voltage-dependent rear surface recombination velocity. *Prog. Photovolt Res. Appl.* 1(2):133–143.

Altermatt P P (2011) Models for numerical device simulations of crystalline silicon solar cells—A review. *J. Comput. Electron.* 10(3):314–330.

Altermatt P P, Heiser G, Aberle A G, Wang A, Zhao J, Robinson S J, Bowden S and Green M A (1996) Spatially resolved analysis and minimization of resistive losses in high-efficiency Si solar cells. *Prog. Photovolt Res. Appl.* 4:399–414.

Altermatt P P, Schenk A, Geelhaar F and Heiser G. (2003) Reassessment of the intrinsic carrier density in crystalline silicon in view of band-gap narrowing. *J. Appl. Phys.* 93(3):1598–1604.

Araújo G L, Cuevas A and Ruiz J M (1986) The effect of distributed series resistance on the dark and illuminated current-voltage characteristics of solar cells. *IEEE Trans. Electron Dev*. 33(3):391–401.

ASTM E490-00a (2014) *Standard Solar Constant and Zero Air Mass Solar Spectral Irradiance Tables*. West Conshohocken, PA: ASTM International. www.astm.org.

ASTM G173-03 (2012) *Standard Tables for Reference Solar Spectral Irradiances: Direct Normal and Hemispherical on* 37° *Tilted Surface*. West Conshohocken, PA: ASTM International. www.astm.org.

Becquerel A E (1839) Recherches sur les effets de la radiation chimique de la lumière solair au moyen des courants électriques. *C. R. Acad. Sci.* 9:561–567.

Blakers A, Wang A, Milne A, Zhao J and Green M A (1989) 22.8% efficient silicon solar cell. *Appl. Phys. Lett.* 55:1363–1365.

Breitenstein O, Rakotoniaina J P, Al Rifai M H and Werner M (2004) Shunt types in crystalline silicon solar cells. *Prog. Photovolt Res. Appl.* 12(7):529–538.

Brendel R (1994) Sunrays: A versatile ray tracing program for the photovoltaic community. In: *Proceedings of the 12th European Photovoltaic Solar Energy Conference*, Amsterdam, Netherlands, pp. 1339–1342.

Chapin D M, Fuller C S and Pearson G O (1954) A new silicon p–n junction photocell for converting solar radiation into electrical power. *J. Appl. Phys.* 25:676–677.

Duffie J A and Beckman W A (2013) *Solar Engineering of Thermal Processes* (4th edition). Hoboken, NJ: John Wiley & Sons Inc.

Dupré O, Vaillon R and Green M A (2015) Physics of the temperature coefficients of solar cells. *Sol. Energ. Mat. Sol. Cells*. 140:92–100.

Fell A (2013) A free and fast three-dimensional/two-dimensional solar cell simulator featuring conductive boundary and quasi-neutrality approximations. *IEEE Trans. Electron Dev*. 60(2):733–738.

Fong K C, McIntosh K R and Blakers A W (2011) Accurate series resistance measurement of solar cells. *Prog. Photovolt Res. Appl.* 21(4):490–499.

Fraunhofer ISE (2013) *Photovoltaics Report*, 7 November 2013. http://www.ise.fraunhofer.de/en/downloads-englisch/pdf-files-englisch/photovoltaics-report-slides.pdf

Garcia M C A and Balenzategui J L (2004). Estimation of photovoltaic module yearly temperature and performance based on Nominal Operation Cell Temperature calculations. *Renew. Energy* 29:1997–2010.

Goetzberger A, Knobloch J and Voß B (1998) *Crystalline Silicon Solar Cells*. Chichester: John Wiley & Sons Ltd.

Green M (1982) *Solar Cells: Operating Principles, Technology, and System Applications*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Green M A (2003) *Third Generation Photovoltaics–Advanced Solar Energy Conversion.* Berlin Heidelberg: Springer-Verlag.

Green M A, Emery K, Hishikawa Y, Warta W and Dunlop E D (2015). Solar cell efficiency tables (Version 45). *Prog. Photovolt Res. Appl.* 23(1):1–9.

Green M A, Ho-Baillie A and Snaith H J (2014) The emergence of perovskite solar cells. *Nat. Photonics.* 8(7):506–514.

Grove A S and Fitzgerald D J (1966) Surface effects on p-n junctions: Characteristics of surface space-charge regions under non-equilibrium conditions. *Solid State Electron.* 9:783–806.

Haug H, Greulich J, Kimmerle A and Marstein E S (2015) PC1Dmod 6.1—State-of-the-art models in a well-known interface for improved simulation of Si solar cells. *Sol. Energ. Mat. Sol. Cells*. 142:47–53.

Holst H, Winter M, Vogt M R. Bothe K, Kontges M, Brendel R and Altermatt P (2013) Application of a new ray tracing framework to the analysis of extended regions in Si solar cell modules. *Energy Procedia* 38:86–93.

Huld T, Friesen G, Skoczek A, Kenny R P, Sample T, Field M and Dunlop E D (2011) A power-rating model for crystalline silicon PV modules. *Sol. Energ. Mat. Sol. Cells*. 95(12):3359–3369.

IEC 61215 (2005) *Crystalline Silicon Terrestrial Photovoltaic (PV) Modules—Design Qualification and Type Approval*. Geneva, Switzerland: International Electrotechnical Commission (IEC). http://www.iec.ch.

International Energy Agency (IEA) (2014) *Technology Roadmap: Solar Photovoltaic Energy* (2014 edition). Accessed January 12, 2016. https://www.iea.org/media/freepublications/technologyroadmaps/solar/TechnologyRoadmapSolarPhotovoltaicEnergy_2014edition.pdf

ITRPV (July 2015) *International Technology Roadmap for Photovoltaic: Results 2014*, 6th edition. Accessed January 2016. http://www.itrpv.net/.cm4all/iproc.php/Reports%20downloads/ITRPV_Roadmap_2015_Rev1_July_150722.pdf?cdp=a

Joint Research Center (JRC) (2016) Photovoltaic Geographical Information System (PVGIS). Accessed 18 January 2016. http://re.jrc.ec.europa.eu/pvgis/

Jordan D C and Kurtz S R (2013) Photovoltaic degradation rates—An analytical review. *Prog. Photovolt Res. Appl.* 21(1):12–29.

Kingston R H and Neustadter S F (1955) Calculation of the space charge, electric field, and free carrier concentration at the surface of a semiconductor. *J. Appl. Phys.* 26(6):718–720.

Klaassen D B M (1992a) A unified mobility model for device simulation—I. Model equations and concentration dependence. *Solid State Electron.* 35(7):953–959.

Klaassen D B M (1992b) A unified mobility model for device simulation—II. Temperature dependence of carrier mobility and lifetime. *Solid State Electron.* 35(7):961–967.

Kopp G and Lean J L (2011) A new, lower value of the total solar irradiance: Evidence and climate significance. *Geophys. Res. Lett.* 38:L01706.

Kuehn R, Fath P and Bucher E (2000) Effects of pn junctions bordering on surfaces investigated by means of 2D-modeling. In: *Proceedings of the 28th IEEE Photovoltaic Specialists Conference*, Anchorage, AK, pp. 116–119.

Lammert M and Schwartz R (1977) The interdigitated back contact solar cell: A silicon solar cell for use in concentrated sunlight. *IEEE Trans. Electron Dev.* 24(4):337–342.

Macleod H A (2010) *Thin-Film Optical Filters* (4th edition). Boca Raton, FL: CRC Press, Taylor & Francis.

Mandelkorn J and Lamneck J (1972) Simplified fabrication of back surface electric field silicon cell and novel characteristics of such cells. In *Proceeding 9th IEEE Photovoltaic Specialists Conference*, Silver Springs, MD.

McIntosh K R (2001) *Lumps, Humps and Bumps: Three Detrimental Effects in the Current-Voltage Curve of Silicon Solar Cell.* PhD thesis, University of New South Wales, Sydney, Australia.

Meteotest (2008) *Meteonorm Version 6.1.0.0.* Accessed January 12, 2016. www.meteotest.com.

Misiakos K and Tsamakis D (1993) Accurate measurements of the silicon intrinsic carrier density from 78 to 340 K. *J. Appl. Phys.* 74(5):3293–3297.

Müller M (2016) Reporting effective lifetimes at solar cell relevant injection densities. *Energy Proc*. 92:138—144.

National Aeronautics and Space Administration (NASA) (2016) Surface meteorology and Solar Energy Accessed January 18, 2016. https://eosweb.larc.nasa.gov/sse/.

National Renewable Energy Laboratory (NREL) (2016) *Renewable Resource Data Center (RReDC).* Accessed January 18, 2016. http://www.nrel.gov/rredc/.

Pingel S, Frank O, Winkler M, Daryan S, Geipel T, Hoehne H and Berghold J (2010) Potential induced degradation of solar cells and panels. In: *Proceedings 35th IEEE Photovoltaic Specialists Conference*, Honolulu, HI, 2010, pp. 2817–2822.

PV Lighthouse (2016) PV Lighthouse. Accessed 12 January 2016. https://www.pvlighthouse.com.au

PV*EDUCATION.ORG* (2016) PVEducation Accessed January 12, 2016. http://pveducation.org

pvXchange Trading GmbH (2016) Market price data. Accessed January 12, 2016. http://www.pvxchange.com/priceindex/Default.aspx?langTag=de-DE

Quaschning V (2005) *Understanding Renewable Energy Systems*. London, UK: Earthscan.

Rhoderick E H and Williams R H (1988) *Metal-Semiconductor Contacts* (2nd edition). New York, NY: Oxford University Press.

Richter A, Glunz S W, Werner F, Schmidt J and Cuevas A (2012) Improved quantitative description of Auger recombination in crystalline silicon. *Phys. Rev. B*. 86(16):165202.

Ross R G Jr and Smokler M I (1986). Flat-Plate Solar Array Project–Final Report, Vol. VI Engineering Sciences and Reliability, JPL Pub. No. 86–31, http://ntrs.nasa.gov/search.jsp?R=19870011218.

Rover D T, Basore P A, Thorson G M (1985) PC-1D version 2: Enhanced numerical solar cell modelling. In: *Proceedings of the 18th IEEE Photovoltaic Specialists Conference*, Las Vegas, NV, p. 703.

Schenk A (1998) Finite-temperature full random-phase approximation model of band gap narrowing for silicon device simulation. *J. Appl. Phys.* 84(7):3684–3695.

Schmidt J, Aberle A G and Hezel R (1997) Investigation of carrier lifetime instabilities in Cz-grown silicon. In: *Proceedings of the 26th IEEE Photovoltaic Specialists Conference*, New York, NY, pp. 13–18.

Schmidt J, Cuevas A, Rein S and Glunz S W (2001) Impact of light-induced recombination centres on the current-voltage characteristic of czochralski silicon solar cells. *Prog. Photovolt Res. Appl.* 9(4):249–255.

Shockley W and Queisser H J (1961) Detailed balance limit of efficiency of p-n junction solar cells. *J. Appl. Phys*. 32(3):510–519.

Silvaco (2016) *Atlas*. Santa Clara, CA: Silvaco. http://www.silvaco.com.

Steingrube S, Wagner H, Hannebauer H, Gatz S, Chen R, Dunham S T, Dullweber T, Altermatt P P, Brendel R (2011) Loss analysis and improvements of industrially fabricated Cz-Si solar cells by means of process and device simulations. *Energy Procedia* 8:263–268.

Steinkemper H, Rauer M, Altermatt P P, Heinz F D, Schmiga C and Hermle M (2015) Adapted parameterization of incomplete ionization in aluminum-doped silicon and impact on numerical device simulation. *J. Appl. Phys.* 117(7):074504.

Synopsys (2016) *TCAD Sentaurus Device version 2015.06*. Mountain View, CA: Synopsys. http://www.synopsys.com.

Tanaka M, Taguchi M, Matsuyama T, Sawada T, Tsuda S, Nakano S, Hanafusa H and Kuwano Y (1992) Development of new a-Si/c-Si heterojunction solar cells: ACJ-HIT (artificially constructed junction-heterojunction with intrinsic thin-layer). *Jpn. J. Appl. Phys.* 31:3518–3522.

Varache R, Leendertz C, Gueunier-Farret M E, Haschke J, Muñoz D and Korte L (2015) Investigation of selective junctions using a newly developed tunnel current model for solar cell applications. *Sol. Energ. Mat. Sol. Cells*. 141:14–23.

Winter M, Vogt M R, Holst H, Altermatt P P (2014) Combining structures on different length scales in ray tracing: Analysis of optical losses in solar cell modules. In: *Proceedings of the 14th International Conference on Numerical Simulation of Optoelectronic Devices,* Palma de Mallorca, Spain.

Wolf M, Noel G T and Stirn R J (1977) Investigation of the double exponential in the current-voltage characteristics of silicon solar cells. *IEEE Trans. Electron Dev.* 24(4):419–428.

Wuerfel P (2005) *Physics of Solar Cells*. Weinheim, Germany: Wiley-VCH Verlag GmbH. KGaA.

Zhao J, Wang A and Green M A (1999) 24.5% efficiency silicon PERT cells on MCZ substrates and 24.7% efficiency PERL cells on FZ substrates. *Prog. Photovolt Res. Appl.* 7:474–474.

40

Multijunction Solar Cells

Matthew Wilkins
and
Karin Hinzer

40.1 Introduction

Multijunction solar cells (MJSCs) currently hold the records for greatest power conversion efficiency of any solar cell, under one-sun conditions (Spectrolab, 38.8% [1]) and under concentration (Soitec/Fraunhofer, 46.0% at 508 suns). MJSCs based on III–V materials such as GaAs and GaInP find applications in aerospace applications where size and weight take precedence over cost, and in terrestrial concentrating photovoltaic (CPV) systems where concentrating optics focus the incident sunlight onto very small ($< 5 \times 5$ mm) cells. MJSCs based on material systems, such as amorphous silicon/silicon and II-VI compounds, and those using III–Vs grown on alternative substrates, have potential to combine high efficiency with lower costs suitable for nonconcentrating terrestrial applications.

Single-junction solar cells have two major loss mechanisms: thermalization and transmission. Photons of energy less than the semiconductor bandgap are transmitted through the material without being absorbed (transmission), while photons of energy in excess of the bandgap convert that excess energy to heat. Single junctions, then, must be designed to minimize transmission (requiring a low bandgap E_g), and minimizing thermalization (requiring a large E_g). For typical solar spectra, this leads to an optimal bandgap of ~1.3–1.4 eV [2,3].

MJSCs consist of a stack of *n-p* semiconductor diodes of different materials, each chosen to absorb a different part of the solar spectrum. As light passes through the stack, the short wavelengths are absorbed in the top junction which has a large E_g, and longer wavelengths are absorbed in each of the subsequent junctions. This arrangement minimizes the impact of transmission and thermalization losses by providing materials that are capable of absorbing the vast majority of photons within the solar spectrum,

and by ensuring that each photon is absorbed in a junction with E_g similar to the photon energy. Typically, the junctions are all grown epitaxially in a monolithic device, with all junctions connected in series; this implies that the current through all junctions must be the same. Less commonly, some devices are designed such that separate contacts are made to each junction so that they can each be operated independently.

Since MJSCs are more expensive to produce than single-junction solar cells, terrestrial applications of MJSCs typically involve concentrating optics that focus light from a large aperture onto a small area of solar cell. In these systems, the sun's intensity is concentrated by a factor of 10 to 2000. Besides the cost savings from using smaller cells, there is also a significant increase in cell efficiency under concentration. We can write the efficiency η as

$$\eta = \frac{P_{\max}}{XE_{1-\text{sun}}} = \frac{J_{SC}\, V_{OC}\, FF}{XE_{1-\text{sun}}}, \tag{40.1}$$

where $E_{1-\text{sun}}$ is the incident light intensity under standard 1-sun conditions, X is the concentration factor, $P_{\max}$ is the output electrical power density at the optimal voltage bias, J_{SC} is the short-circuit current density, V_{OC} is the open-circuit voltage, and FF is the fill factor.

A simple equivalent circuit model for the series-connected triple-junction solar cell is shown in Figure 40.1b; each junction i is modeled with a current source providing the 1-sun photocurrent for that

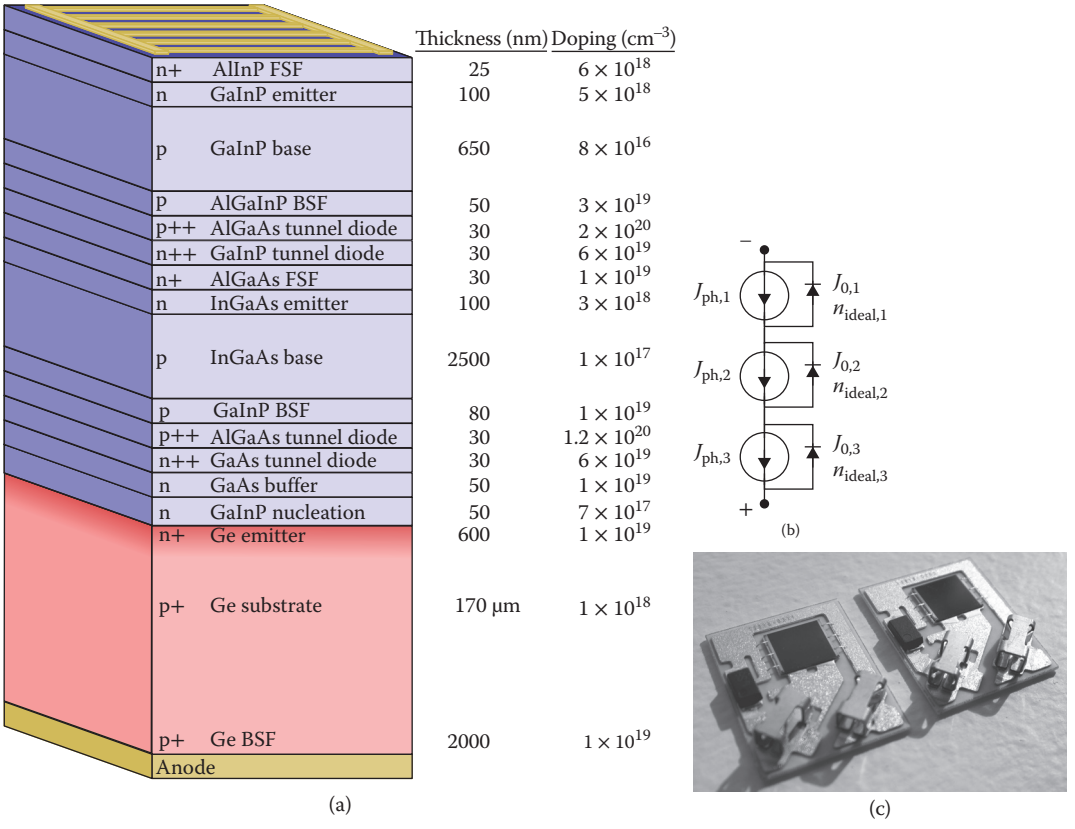

FIGURE 40.1 (a) Layer structure of a triple-junction GaInP/(In)GaAs/Ge solar cell. (b) Simple circuit model for the triple-junction cell. (c) Two 5 × 5-mm MJSCs mounted on ceramic carriers.

subcell, $J_{\mathrm{ph},i}$. A parallel diode provides the exponential diode behavior of each cell, so the device current density

$$J = XJ_{\mathrm{ph},i} - J_{0,i}\left(\exp\frac{qV_i}{n_{\mathrm{ideal},i}kT} - 1\right) \quad \forall\, i, \tag{40.2}$$

where kT/q is the thermal voltage, $n_{\mathrm{ideal},i}$ is the diode ideality factor of each subcell and $J_{0,i}$ is the saturation current density. It is useful to model each junction with two parallel diodes with different ideality factors, and also add parasitic series and shunt resistances. The diodes effectively model the Shockley–Read–Hall (SRH), radiative, and Auger recombination mechanisms, which are active in each junction.

If all three photocurrents $J_{\mathrm{ph},i}$ are equal, then the device's short-circuit current density $J_{\mathrm{SC}} = J_{\mathrm{ph},i}$ and the current density through the diodes is zero. If they are not equal, then $J_{\mathrm{SC}} = \min(J_{\mathrm{ph},i})$, and any excess photocurrent is dissipated through the diodes.

Under open-circuit conditions, the current at the device terminals, and at the interconnections between subcells, is zero. Therefore, all photogenerated current recombines via the diodes.

The short-circuit current density J_{SC} increases linearly with concentration, but there is also an increase in open-circuit voltage V_{OC}, which comes about due to the higher concentration of carriers in the semiconductor and larger separation of the quasi-Fermi levels. Under a concentration factor X, the open-circuit voltage is [4]

$$V_{\mathrm{OC}} = \frac{kT}{q}\left[\sum_i n_{\mathrm{ideal},i}\ln\left(\frac{J_{\mathrm{ph},i}}{J_{0,i}}\right) + \ln X\sum_i n_{\mathrm{ideal},i}\right]. \tag{40.3}$$

If we neglect effects due to parasitic resistances, the fill factor also increases with concentration; consequently, MJSCs generally gain in efficiency as concentration is increased. Of course, this relation does not extend to infinitely high concentration; eventually the assumptions used to derive the diode-like behavior of Equation 40.2 break down, while resistive terms become increasingly important. In such cases, a more complete physics-based device simulation can be useful to study these limits.

The most common commercially available MJSC is a lattice-matched design with triple junctions consisting of $Ga_{0.49}In_{0.51}P$, $In_{0.01}Ga_{0.99}As$, and Ge, respectively [5]. The GaInP and (In)GaAs junctions are clad with a front surface field and a back surface field of a larger bandgap, which helps to confine minority carriers within the junction region. The triple *n-p* junctions (or subcells) are connected in series with *p*++/*n*++ tunnel diodes, which provide an optically transparent, ohmic connection between the *p*-type base of one subcell and the *n*-type emitter of the next subcell. The structure is grown by metal-organic chemical vapor deposition (MOCVD) on a germanium substrate, with the germanium junction being formed by diffusion of group V dopants into the top surface of the substrate during the initial growth of epitaxial material [6,7]. The remaining subcell and tunnel-diode layers are formed by conventional MOCVD growth, requiring ∼4 μm of epitaxial material in ∼20 layers. A typical layer structure is shown in Figure 40.1a, and some typical 5 × 5-mm cells are shown in Figure 40.1c. This basic structure will be used as an example throughout this chapter.

MJSCs have been simulated with good results using a variety of approaches including lumped-parameter models [2], distributed parameter models [8], and drift-diffusion-based simulators in 1D, 2D, and 3D. In this chapter, we focus primarily on drift-diffusion-based methods. The simulation results shown here were generated using the Sentaurus TCAD package from Synopsys, Inc.; example simulation projects for MJSC devices are available from most of the major simulation software vendors [9–11] and should provide a good starting point.

One of the greatest challenges in simulating these devices is managing the large number of material parameters that are required for each of the materials; this is particularly difficult for ternary and quaternary materials such as AlGaAs and InGaAs for example, where the material composition may be subject to optimization and so material parameters are needed for a wide variety of compositions.

40.2 Device Model and Simulation Workflow

The process of setting up and running a simulation begins by defining each of the material regions (Figure 40.2). In the case of MJSC, this will be a simple structure of uniform layers for the most part; the layers are defined in Figure 40.1a. The exception in our example structure is the germanium subcell, which is formed by diffusion of dopants rather than by epitaxial growth of the emitter. For the germanium subcell, we model the *p*-type substrate as a single region, with a smoothly varying "error function" profile of *n*-type doping added to the emitter area. The back surface field at the bottom of the substrate is also modeled in the same way.

Each of the regions is discretized with a mesh, material parameters are assigned to each region, and ohmic contact boundary conditions are defined at the cathode and anode.

From this point, the process differs for QE or for *J*–*V* curve calculations. For a QE calculation, additional "virtual contacts" are be defined between each junction, enabling the current from each junction to be extracted separately. An optical calculation is done to determine the optical generation rate as a function of position within the device at the starting wavelength. The Poisson equation is solved to find the potential profile at equilibrium, and then the coupled Poisson and electron and hole continuity equations

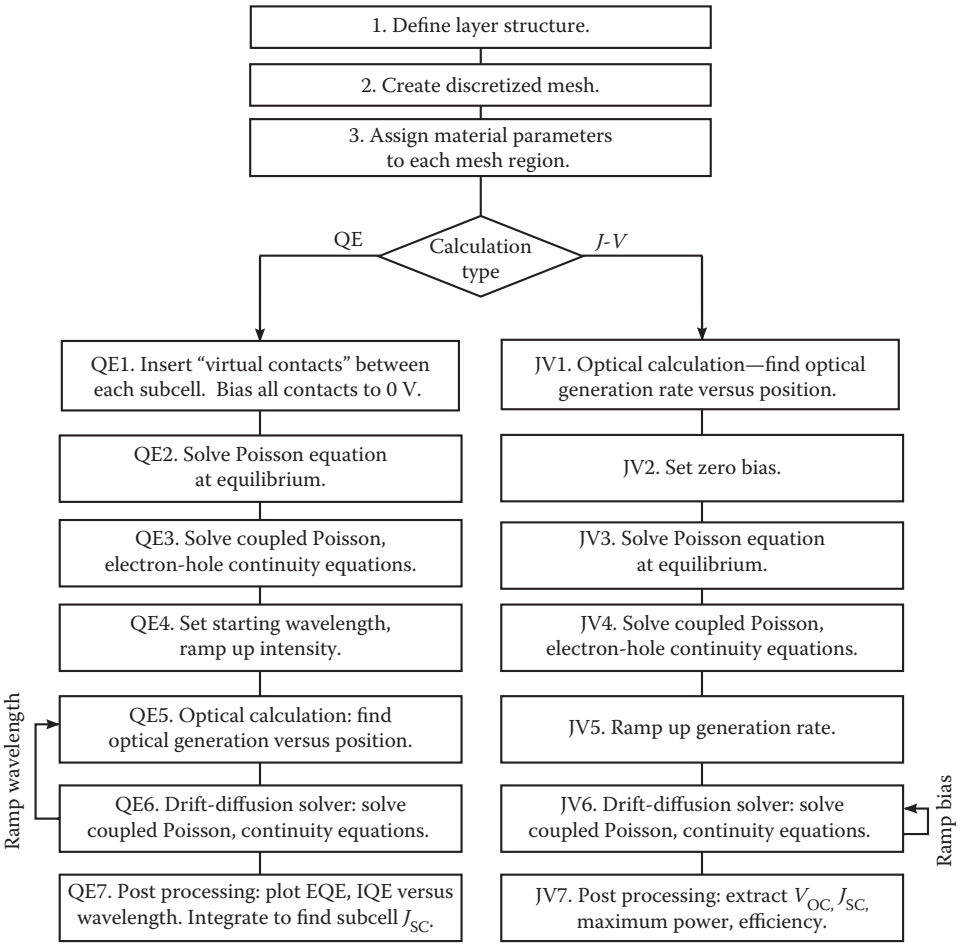

FIGURE 40.2 Typical simulation workflow for calculation of quantum efficiency (QE) and current–voltage (*J*–*V*) characteristics of multijunciton solar cells.

are solved as the incident optical intensity (and the resulting optical generation rate) is ramped up to the desired value. Finally, the incident wavelength is ramped over the range of interest, with the optical problem being recalculated at each wavelength, and the coupled Poisson-continuity problem is also solved at each wavelength to find the short-circuit current of each junction. This wavelength and short-circuit current data are postprocessed to produce plots of QE.

In the case of a J–V curve calculation, the optical problem is only solved once, but with an appropriate broadband spectrum. The simulation then starts with a Poisson solution at zero bias. Intensity is gradually ramped up as was the case for the QE calculation, and then the voltage bias is ramped to produce a J–V curve. Postprocessing is done to identify the maximum power point and extract key parameters, such as open-circuit voltage (V_{OC}), short-circuit current (J_{SC}), fill factor (FF), and power conversion efficiency (η).

40.3 Mesh Discretization

Depending on the physical phenomena which are being studied and the symmetry of the device, the region of interest can be discretized with a one-dimensional mesh [12], a 2D mesh representing a small symmetry element such as a half-gridline pitch [13], an entire half of the device, or a 3D model of a portion of the device [14]. The mesh must resolve steps in potential at material heterointerfaces, so ideally mesh vertices along the interfaces should be duplicated with a boundary condition relating the solution variables on each side of the interface (Figure 40.3).

In the direction normal to the interfaces, the minimum required mesh density is typically driven by a need to resolve potential gradients near the interfaces and p–n junctions. Tunnel diodes in particular have depletions regions of perhaps 10 nm in thickness, and these must be resolved in order for a nonlocal tunneling model to work properly. Far from the interfaces and junctions, the minimum mesh density is driven by the need to resolve standing waves in the optical intensity profile. In the lateral direction, mesh elements can usually be allowed to grow quite large, except at the edges of gridlines and at the device perimeter.

For many studies, a 1D model or a 2D half-gridline pitch option is sufficient to study the phenomena of interest; however, certain phenomena will not be accurately represented. Table 40.1 summarizes the

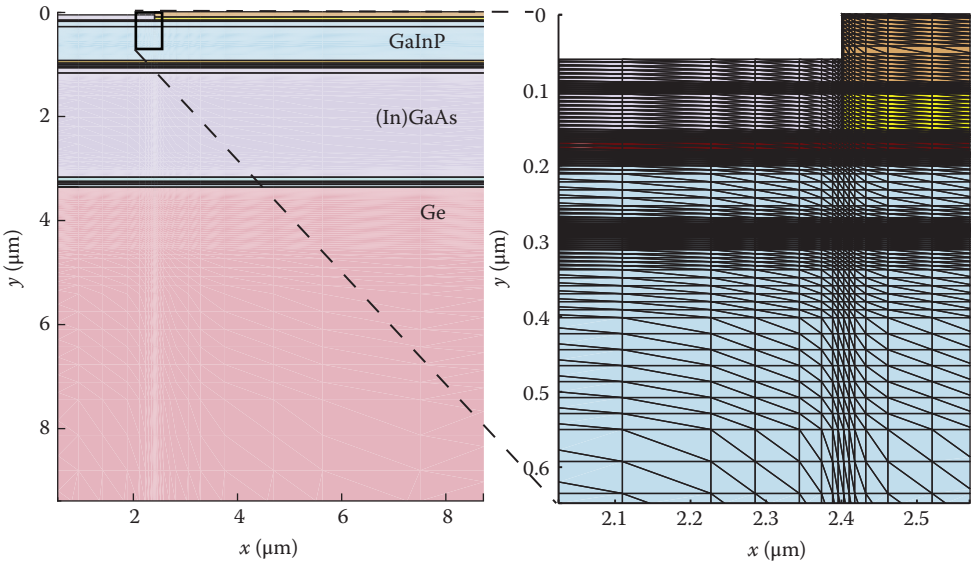

FIGURE 40.3 (a) A 2D, half-grid pitch mesh for the triple-junction solar cell. The full simulation region extends to 60 × 170 μm. (b) Detail of the area near the top contact gridline.

TABLE 40.1 Limitations of Reduced Model Geometries.

	1D	2D Half-Grid Pitch	2D Half-Device	3D Distributed Parameter Model Calibrated Using 2D Simulations	3D
p-n junctions and tunnel diodes	•	•	•	•	•
ARC/window interface		•	•	•	•
Emitter sheet resistance		•	•	•	•
Shading due to gridlines		•	•	•	•
Perimeter recombination			Partial	•	•
Grid ohmic losses				•	•
Nonuniform bias due to grid resistance				•	•
Nonuniform illumination			•	•	•
Busbar regions			Partial	•	•

ARC, anti-reflection coating

limitations of various reduced-dimensional representations. The full 2D and 3D options may be very computationally intensive and will likely be impractical using standard desktop computer hardware; however Létay et al. have demonstrated a method of addressing this by modeling small portions of the device representing the perimeter and the center of the device, linked through an external circuit simulator [13]. Alternatively, a 2D half-grid pitch model can be used to characterize the illuminated I–V performance of a small cell element, and that information can be used to calibrate the parameters of a distributed circuit model [15].

40.4 Optics

Several types of solvers exist which can calculate the carrier generation rate as a function of position within the device layer stack for a given illumination. Raytracing and finite-difference time domain methods [14] can be used, but here we focus on the transfer matrix method (TMM), which is widely used in simulating MJSCs because they have a planar structure and minimal surface texture.

40.4.1 Transfer Matrix Method

In a 2D or 3D simulation including gridlines, we can apply the TMM to calculate the optical generation rate as a function of depth within the layer stack. The propagation of electromagnetic waves through the layer stack is represented by a series of matrices. For each layer there is a matrix describing the reflections from the top interface, and a second matrix representing the phase change and attenuation through the thickness of the layer.

For a stack of N layers, we define phasors representing the amplitude and phase of the y-component of the electric field for downward- and upward-propagating waves at the top and bottom of each layer. The vector

$$\begin{bmatrix} E^{+}_{m,\text{top}} \\ E^{-}_{m,\text{top}} \end{bmatrix} \tag{40.4}$$

contains the phasors describing the electric field at the top of layer m. The symbol + refers to downward propagation (into the cell) and − refers to upward propagation (reflected waves). Layer 0 is the medium

above the cell, and layer $N + 1$ is the medium below the cell, which represents the opaque back contact, or air in the case of a bifacial solar cell.

Note that for a plane wave of wavelength λ_0 with an angle of incidence θ_0 with respect to the surface normal, the angle of propagation within any layer m can be found from Snell's equation, $n_0 \sin\theta_0 = n_m \sin\theta_m$. The y-component of the wave vector is then

$$k_{y,m} = \frac{2\pi}{\lambda_0}\frac{n_m}{n_0}\cos\theta_m. \tag{40.5}$$

The phasors at the top and bottom interfaces of a layer are related by a propagation matrix $\mathbf{P}_m$,

$$\begin{bmatrix} E^+_{m,\text{top}} \\ E^-_{m,\text{top}} \end{bmatrix} = \begin{bmatrix} e^{-jk_{y,m}d_m} & 0 \\ 0 & e^{jk_{y,m}d_m} \end{bmatrix} \begin{bmatrix} E^+_{m,\text{bot}} \\ E^-_{m,\text{bot}} \end{bmatrix}, \tag{40.6}$$

where j is $\sqrt{-1}$, and d_m is the thickness of layer m.

The phasors at bottom of one layer and the top of the next are are related by

$$\begin{bmatrix} E^+_{m-1,\text{bot}} \\ E^-_{m-1,\text{bot}} \end{bmatrix} = \frac{1}{t_{m-1,m}} \begin{bmatrix} 1 & r_{m-1,m} \\ r_{m-1,m} & 1 \end{bmatrix} \begin{bmatrix} E^+_{m,\text{top}} \\ E^-_{m,\text{top}} \end{bmatrix}, \tag{40.7}$$

where $t_{m-1,m}$ and $r_{m-1,m}$ are the Fresnel coefficients for reflection and transmission at the interface. We denote this transmission matrix $\mathbf{T}_{m-1,m}$.

A transfer matrix for the entire layer stack can then be assembled as

$$M = \prod_{m=1}^{N} \left(\mathbf{T}_{m-1,m}\mathbf{P}_m\right)\mathbf{T}_{N,N+1}. \tag{40.8}$$

And finally, we set the electric field incident on the top of the cell, $E^+_{0,\text{bot}}$, to any convenient value and write

$$\begin{bmatrix} E^+_{0,\text{bot}} \\ E^-_{0,\text{bot}} \end{bmatrix} = M \begin{bmatrix} E^+_{N+1,\text{top}} \\ 0 \end{bmatrix}. \tag{40.9}$$

This can be solved for the reflected field, $E^-_{0,\text{bot}}$, and from that point the total electric field at any point in the structure can be found straightforwardly by assembling a transfer matrix to relate the fields at the top surface with the fields at the point of interest.

This type of calculation is often done assuming that unpolarized light is incident perpendicular to the device, but is also valid for other angles of incidence. The calculation is done as a 1D problem and the resulting carrier generation rate profile is applied to all regions that are not shaded by the grid (Figure 40.4b). Where a broad spectrum illumination is being studied, the TMM is solved repeatedly at each wavelength and the carrier generation rate is summed, weighted according to the incident spectral irradiance [16]. Similarly, illumination from a wide range of angles can be modeled by summing the optical generation calculated from many TMM calculations.

40.4.2 Complex Refractive Index

Solar cell simulations require data on the complex refractive index (refractive index n and extinction coefficient k) over a wide range of wavelengths. This can be a challenge as published data are often limited to a certain range of wavelengths [17]; it may be necessary to measure properties of some materials

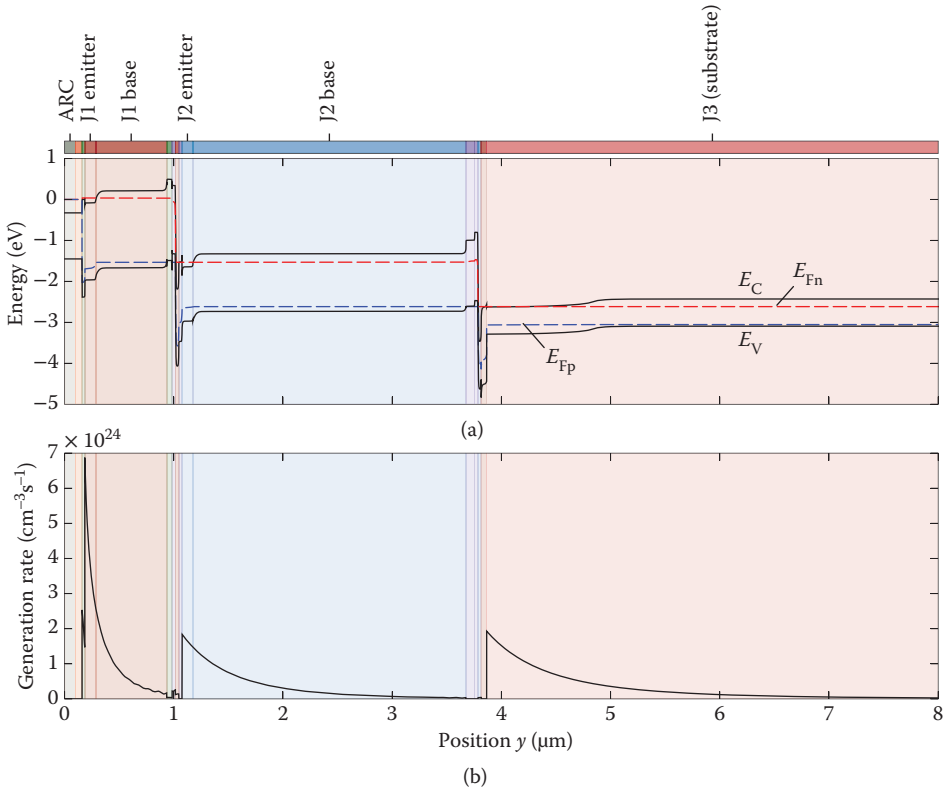

FIGURE 40.4 (a) Simulated band diagram at maximum power point for the triple-junction solar cell. (b) Profile of optical generation rate through the depth of the triple-junction solar cell, as calculated using the TMM. Illumination is broadband (AM1.5D spectrum). (ARC, antireflection coating.)

using ellipsometry or to combine data from several studies to yield a useable dataset. There can be further difficulties due to the use of ternary and quaternary alloys where data are only available for specific compositions.

In some cases, research has been done to model the dielectric function of a material in terms of a model dielectric function (MDF) consisting of a series of Lorentzian oscillators and other terms representing the various band-to-band transitions [18–20]. If the oscillator parameters are estimated as a function of alloy composition, this can be an effective method to generate *n*,*k* data for arbitrary alloy compositions. For use in simulating solar cells, care must be taken because these MDF formulations usually yield excessively large *k* values at photon energies below the bandgap.

40.4.3 Absorption and Quantum Yield

Once the electric field and optical intensity as a function of position within the structure is calculated using the TMM (or any other method), the local absorbed power density is the negative derivative of the optical intensity, $A = -\nabla(\frac{1}{2}\epsilon|\mathbf{E}|^2)$, where ϵ is the dielectric permittivity. If we assume that the quantum yield is unity (i.e., all absorbed photons result in generated carriers), then the local carrier generation rate is

$$G = -\frac{\lambda}{hc}\nabla(\frac{1}{2}\epsilon|\mathbf{E}|^2). \tag{40.10}$$

The assumption of unity quantum yield generally works very well in III–V heterostuctures, but in some materials such as highly doped silicon, free-carrier absorption is significant and will result in a quantum yield less than unity.

The resulting optical generation profile is plotted in Figure 40.4b, for the case of broadband illumination with the AM1.5D spectrum as defined by ASTM G173-03. We can note that within each subcell, there is strong carrier generation near the top of the junction, which then decays quickly through the thickness of the junction. This is to be expected since each subcell is designed to absorb nearly all of the incident light of energy greater than its bandgap, so the generation decays nearly to zero at the bottom of each junction.

40.5 Drift-Diffusion Calculation

Having determined the optical generation rate as a function of position, we then proceed to simulation of the electrical device performance. The electrostatic potential within the device follows the Poisson equation,

$$\nabla \cdot (\epsilon \nabla \phi + \vec{P}) = -q(p - n - N_A + N_D), \tag{40.11}$$

where ϕ is the electrostatic potential, $\vec{P}$ is the polarization which is zero in semiconductors with a zincblende crystal structure, p and n are the free hole and electron concentrations respectively, and N_A and N_D are the acceptor and donor concentrations. ϵ is the permittivity of the material.

The electron and hole concentrations also follow the continuity equations,

$$\begin{aligned} \nabla \cdot \vec{J}_n &= qR_{net} + q\frac{\partial n}{\partial t}, \\ -\nabla \cdot \vec{J}_p &= qR_{net} + q\frac{\partial p}{\partial t}. \end{aligned} \tag{40.12}$$

$\vec{J}_n$ and $\vec{J}_p$ are the electron and hole current densities; R_{net} is the net recombination rate, i.e., the sum of contributions from all recombination and generation processes. At a minimum this will include SRH recombination, Auger recombination, and radiative recombination, as well as the optical generation calculated using the TMM. It can also include additional terms such as band-to-band tunneling and coupled generation, as we discuss below. For studies of solar cells, we are generally interested in steady-state solutions so the time derivaties are zero. The current densities $\vec{J}_n$ and $\vec{J}_p$ can be further expanded into drift and diffusion terms, which depend on the gradient of the local electrical potential and the gradients of the carrier concentrations, the same set of unknowns as in Equation 40.11. Using the Einstein relation, the current densities are [21]

$$\begin{aligned} \vec{J}_n &= -nq\mu_n \nabla E_{Fn}, \\ \vec{J}_p &= -pq\mu_p \nabla E_{Fp}, \end{aligned} \tag{40.13}$$

where μ_n and μ_p are the electron and hole mobilities and E_{Fn} and E_{Fn} are the electron and hole quasi-Fermi levels. Relating the conduction band energy and valence band energy E_C and E_V to the material properties and the electrostatic potential, we have

$$\begin{aligned} E_C &= -\chi - q(\phi - \phi_{ref}), \\ E_V &= -\chi - E_{g,eff} - q(\phi - \phi_{ref}), \end{aligned} \tag{40.14}$$

where χ is the electron affinity of the material, $E_{g,eff}$ is the effective bandgap (including the effect of bandgap narrowing), and ϕ_{ref} is an arbitrarily chosen reference potential.

Using Fermi–Dirac statistics for the electron and hole concentrations and making the assumptions that the carrier populations and the crystal are at a constant temperature throughout the device and that the

conduction and valence bands have parabolic band structures, the electron and hole quasi-Fermi levels are [21]

$$E_{\mathrm{F}n} = kT\mathcal{F}_{1/2}^{-1}\left(\frac{n}{N_{\mathrm{C}}}\right) + E_{\mathrm{C}},$$
$$E_{\mathrm{F}p} = -kT\mathcal{F}_{1/2}^{-1}\left(\frac{p}{N_{\mathrm{V}}}\right) + E_{\mathrm{V}}, \tag{40.15}$$

where k is the Boltzmann constant, T is the device temperature, $\mathcal{F}_{1/2}^{-1}$ is the inverse Fermi-Dirac integral of order 1/2, and N_{C} and N_{V} are the effective densities of states in the conduction and valence bands, respectively. By substituting Equations 40.13 through 40.15 into Equation 40.12, we arrive at a formulation of the three coupled differential equations for steady-state problems, stated solely in terms of the material properties and functions of the three variables n, p, and ϕ:

$$\nabla \cdot \epsilon \nabla \phi = -q(p - n - N_{\mathrm{D}} - N_{\mathrm{A}}), \tag{40.16}$$

$$\nabla \cdot \left[n\mu_n \nabla \left(kT\mathcal{F}_{1/2}^{-1}\left(\frac{n}{N_{\mathrm{C}}}\right) - \chi - q(\phi - \phi_{\mathrm{ref}}) \right) \right] = -R_{\mathrm{net}},$$
$$\nabla \cdot \left[p\mu_p \nabla \left(kT\mathcal{F}_{1/2}^{-1}\left(\frac{p}{N_{\mathrm{V}}}\right) + \chi + E_{\mathrm{g,eff}} + q(\phi - \phi_{\mathrm{ref}}) \right) \right] = -R_{\mathrm{net}}. \tag{40.17}$$

Typically, we need to solve the coupled set of Equations 40.16 and 40.17 to find the local potential and carrier concentrations throughout the device. Once those quantities are known, then the local electron and hole current densities and the quasi-Fermi levels are also known, and the voltage and current at each electrode can also be found. We are dealing with heterostructures with abrupt changes in the valence and conduction band energies and other quantities at each material interface. To handle these discontinuities, the coupled equations are solved separately in each material region with a set of boundary conditions relating the solution variables on each side of the interfaces. The boundary conditions allow for thermionic emission across or tunneling through small potential barriers at the interfaces [22,23].

It can be difficult to achieve convergence of the coupled Poisson and continuity equations, so we typically begin first with a solution of the Poisson equation under equilibrium conditions (i.e., with all currents and net recombination/generation equal to zero). This provides an initial guess for the solution of the coupled Poisson and continuity equations, again under equilibrium. Next, we gradually ramp up the illumination intensity, leading to a solution at nonequilibrium, and then sweep either the bias at the contacts (to produce a J–V curve) or the illumination wavelength (to produce a QE plot). In the case of QE, the optical generation must be recalculated at each wavelength.

In order to provide physically accurate results, the set of physical models and material parameters must be carefully chosen. Given the wide range of materials and compositions that are present in a single device, managing this database of material parameters is probably the single most difficult part of a MJSC device simulation. Altermatt provides an excellent guide to the parameters needed to do good simulations of silicon solar cells [24]; similar studies are needed for each of the materials in the MJSC structure. Wherever possible, it is advisable to validate simulations of single junctions of each material before attempting to simulate a multijunction device.

40.5.1 Band Structure and Intrinsic Carrier Concentration

While material bandgaps are often well known to a high degree of accuracy, Altermatt makes the point that it is more important to have a self-consistent set of parameters (i.e., bandgap, effective masses, and

doping-induced bandgap narrowing) that lead to an accurate value of the effective intrinsic carrier concentration, than to have any one of these individually correct.

If Fermi–Dirac statistics are being used in the simulation, then the experiments used to determine these parameters must also have been based on the same statistics. For the III–V materials commonly used in MJSCs, [25] is a good starting point.

Ternary materials in the $(Al,Ga)_{0.5}In_{0.5}P$ system can have varying degrees of Cu:Pt ordering, where Al or Ga atoms and In atoms are found on separate planes of the crystal structure [26]. This means that a range of bandgap values are possible for the same material composition. For example, $Ga_{0.5}In_{0.5}P$ can have a bandgap in the range between 1.8 and 1.9 eV.

In Tables 40.2 through 40.4, we provide the material parameters used for the drift-diffusion-based simulation of the triple-junction solar cell as presented in this chapter. The effective bandgap, $E_{g,eff}$ at a temperature T is described using a Varshni model [25] for the temperature dependence,

$$E_{g,eff} = E_{g,ref} - \frac{\alpha T^2}{T+\beta} + \frac{\alpha T_{ref}^2}{T_{ref}+\beta} - \Delta E_g, \tag{40.18}$$

where $E_{g,ref}$ is the bandgap at a reference temperature T_{ref}, and α and β are the Varshni parameters. ΔE_g is a doping-dependent bandgap narrowing term, which we specify using a table of data points. Similarly, the electron affinity χ (difference between the vacuum energy level and the conduction band energy) is described by

$$\chi = \chi_{ref} + \frac{\alpha T^2}{T+\beta} - \frac{\alpha T_{ref}^2}{T_{ref}+\beta} + \Delta E_g \cdot (\Delta\chi/\Delta E_g), \tag{40.19}$$

where the ratio $\Delta\chi$ / ΔE_g is specified as a constant.

We calculate the conduction band and valence band densities of states, N_C and N_V, from the electron and hole effective masses m_n and m_p [21],

$$N_C = 2.5094 \times 10^{19} \left(\frac{m_n}{m_0} \cdot \frac{T}{300\ \mathrm{K}}\right)^{3/2} \mathrm{cm}^{-3}, \tag{40.20}$$

$$N_V = 2.5094 \times 10^{19} \left(\frac{m_p}{m_0} \cdot \frac{T}{300\ \mathrm{K}}\right)^{3/2} \mathrm{cm}^{-3}. \tag{40.21}$$

The effective intrinsic carrier density is then calculated from the band edge densities of states and the effective bandgap,

$$n_{i,eff} = \sqrt{N_C N_V}\, \exp\frac{-E_{g,eff}}{2kT}. \tag{40.22}$$

Carrier mobilities μ_n and μ_p are adjusted for temperature- and doping-dependent effects. One of two parameterizations, the "Masetti model" [27] or the "Arora model" [28] are used depending on what data is available and/or which model provides the best fit to data. The Masetti model takes the form

$$\mu = \mu_{min1}\, \exp\left(\frac{-P_c}{N_A+N_D}\right) + \frac{\mu_{max}(T/300\ \mathrm{K})^{\zeta} - \mu_{min2}}{1+\left[(N_A+N_D)/C_r\right]^{\alpha}} - \frac{\mu_1}{1+\left[C_s/(N_A+N_D)\right]^{\beta}}, \tag{40.23}$$

where N_D and N_A are the concentrations of donors and acceptors respectively. The remaining parameters are listed in the parameter tables.

TABLE 40.2 Table of Material Parameters for the GaInP/GaInAs/Ge Solar Cell Example.

	Symbol	Units	$Ga_{0.49}In_{0.51}P$	$In_{0.01}Ga_{0.99}As$	Germanium
	ϵ/ϵ_0		12.005	12.91	16.0
			[29]	[29]	[30,31]
	n, k		[20]	[32]	[33]
Band gap	$E_{g,0}$	eV	1.88	1.40	0.744
	α	eV/K	1.82×10^{-4}	5.405×10^{-4}	4.77×10^{-4}
	β	K	81	204	235
	T_{ref}	K	300	300	0
	χ_0	eV	3.924	4.07	3.960
	ΔE_g	eV	[29]	[34]	[29]
	$\Delta\chi/\Delta E_g$		0.5	0.5	0.5
			[35]	[25]	[31]
Density of states	m_n/m_0		0.08515	0.06553	0.5438
	$N_{\text{C, 300}}$	cm^{-3}	6.235×10^{17}	4.210×10^{17}	1.0063×10^{19}
	m_p/m_0		0.7125	0.5236	0.3406
	$N_{\text{V, 300}}$	cm^{-3}	1.5093×10^{19}	9.509×10^{18}	4.9883×10^{18}
			[29,35]	[25,35]	[29,35]
Mobility (Masetti model)	μ_{max}	$\text{cm}^2/\text{V.s}$		(9400, 491.5)	
	ζ			(2.1, 2.2)	
	μ_{min1}	$\text{cm}^2/\text{V.s}$		(500.0, 20.0)	
	μ_{min2}	$\text{cm}^2/\text{V.s}$		(500.0, 20.0)	
	μ_1	$\text{cm}^2/\text{V.s}$		(0, 0)	
	P_c	cm^{-3}		(0, 0)	
	C_r	cm^{-3}		$(6 \times 10^{16}, 1.48 \times 10^{17})$	
	C_s	cm^{-3}		(0, 0)	
	α			(0.394, 0.38)	
	β			(0, 0)	
				[36]	
Mobility (Arora model)	A_{min}	$\text{cm}^2/\text{V.s}$	(400.0, 15.0)		(850.0, 300.0)
	α_{m}		(0, 0)		(0, 0)
	A_{d}	$\text{cm}^2/\text{V.s}$	(3900, 135)		(2950, 1500)
	α_{d}		(0, 0)		(0, 0)
	A_N	cm^{-3}	$(2 \times 10^{16}, 1.5 \times 10^{17})$		$(2.6 \times 10^{17}, 1 \times 10^{17})$
	α_N		(1.955, 1.47)		(0, 0)
	A_{a}		(0.7, 0.8)		(0.56, 1.0)
	α_{a}		(0, 0)		(0, 0)
			[36]		[31]
Recombination	B_{rad}	cm^3/s	1×10^{-10}	2×10^{-10}	6.4×10^{-14}
	τ_{max}	s	$(2 \times 10^{-8}, 2 \times 10^{-8})$	$(1 \times 10^{-7}, 2 \times 10^{-8})$	(0.001, 0.001)
	τ_{min}	s	(0, 0)	(0, 0)	(0, 0)
	N_{ref}	cm^{-3}	$(1 \times 10^{19}, 1 \times 10^{19})$	$(1 \times 10^{16}, 2 \times 10^{18})$	$(1 \times 10^{16}, 1 \times 10^{16})$
	γ		(1.0, 1.0)	(1.0, 3.0)	(1.0, 1.0)
	T_α	K	(0, 0)	(0, 0)	(0, 0)
	T_{coeff}		(0, 0)	(0, 0)	(0, 0)
	A_{A}	cm^6s^{-1}	$(3 \times 10^{-30}, 3 \times 10^{-30})$	$(5 \times 10^{-30}, 5 \times 10^{-30})$	$(1 \times 10^{-30}, 1 \times 10^{-30})$
	B_{A}	cm^6s^{-1}	(0, 0)	(0, 0)	(0, 0)
	C_{A}	cm^6s^{-1}	(0, 0)	(0, 0)	(0, 0)
	H		(0, 0)	(0, 0)	(0, 0)
	N_0	cm^{-3}	$(1 \times 10^{18}, 1 \times 10^{18})$	$(1 \times 10^{18}, 1 \times 10^{18})$	$(1 \times 10^{18}, 1 \times 10^{18})$
			[35]	[35]	[35]
Nonlocal tunneling	g		(0.21, 0.4)	(0.21, 0.4)	
	m_{t}/m_0		(0.24, 0.48)	(0.085, 0.34)	
			[37]	[38]	

Note: Paired values in parentheses are specified for electrons and for holes, respectively.

TABLE 40.3 Table of Material Parameters for the GaInP/GaInAs/Ge Solar Cell Example

	Symbol	Units	$Al_{0.5}In_{0.5}P$	$Al_{0.3}Ga_{0.7}As$	$Al_{0.2}Ga_{0.8}As$
	ϵ/ϵ_0		11.355	12.055	12.34
			[29]	[29]	[29]
	n, k		[20]	[19]	[19]
Band gap	$E_{g,0}$	eV	2.382	1.7976	1.6726
	α	eV/K	1.76×10^{-4}	4.98×10^{-4}	5.12×10^{-4}
	β	K	134.2	142.8	163.2
	T_{ref}	K	300	300	300
	χ_0	eV	3.75	3.756	3.877
	ΔE_g	eV	[29]	[34]	[34]
	$\Delta\chi/\Delta E_g$		0.5	0.5	0.5
			[25]	[25,35]	[25,35]
Density of states	m_n/m_0		0.23	0.0879	0.0796
	$N_{C,300}$	cm^{-3}	2.768×10^{18}	6.540×10^{17}	5.636×10^{17}
	m_p/m_0		0.36	0.5842	0.5593
	$N_{V,300}$	cm^{-3}	5.4203×10^{18}	1.1206×10^{19}	1.0496×10^{19}
			[25]	[25,35]	[25,35]
Mobility (Masetti model)	μ_{max}	$cm^2/V.s$	(150.0, 180.0)	(3721.1, 240)	(5896.6, 308.21)
	ζ		(1.0, 1.0)	(1.74, 1.83)	(1.81, 1.9)
	μ_{min1}	$cm^2/V.s$	(0, 0)	(195.09, 5.0)	(313.6, 8.8571)
	μ_{min2}	$cm^2/V.s$	(0, 0)	(195.09, 5.0)	(313.6, 8.8571)
	μ_1	$cm^2/V.s$	(0, 0)	(0, 0)	(0, 0)
	P_c	cm^{-3}	(0, 0)	(0, 0)	(0, 0)
	C_r	cm^{-3}	$(5 \times 10^{17}, 2.75 \times 10^{17})$	$(1.16 \times 10^{17}, 1.00 \times 10^{17})$	$(9.33 \times 10^{16}, 1.07 \times 10^{17})$
	C_s	cm^{-3}	(0, 0)	(0, 0)	(0, 0)
	α		(0.436, 0.397)	(0.5758, 0.324)	(0.5152, 0.33425)
	β		(0, 0)	(0, 0)	(0, 0)
			[36]	[36]	[36]
Mobility (Arora model)	A_{min}	$cm^2/V.s$			
	α_m				
	A_d	$cm^2/V.s$			
	α_d				
	A_N	cm^{-3}			
	α_N				
	A_a				
	α_a				
Recombination	B_{rad}	cm^3/s	1×10^{-10}	1×10^{-10}	1×10^{-10}
	τ_{max}	s	$(2.5 \times 10^{-8}, 2.5 \times 10^{-8})$	$(1 \times 10^{-6}, 1 \times 10^{-8})$	$(1 \times 10^{-6}, 1 \times 10^{-8})$
	τ_{min}	s	$(5 \times 10^{-10}, 5 \times 10^{-10})$	(0, 0)	(0, 0)
	N_{ref}	cm^{-3}	$(1 \times 10^{16}, 1 \times 10^{16})$	$(1 \times 10^{16}, 2 \times 10^{18})$	$(1 \times 10^{16}, 2 \times 10^{18})$
	γ		(1.0, 1.0)	(1.0, 3.0)	(1.0, 3.0)
	T_α	K	(0, 0)	(0, 0)	(0, 0)
	T_{coeff}		(0, 0)	(0, 0)	(0, 0)
	A_A	cm^6s^{-1}	$(3 \times 10^{-30}, 3 \times 10^{-30})$	$(1 \times 10^{-30}, 1 \times 10^{-30})$	$(1 \times 10^{-30}\ 1 \times 10^{-30})$
	B_A	cm^6s^{-1}	(0, 0)	(0, 0)	(0, 0)
	C_A	cm^6s^{-1}	(0, 0)	(0, 0)	(0, 0)
	H		(0, 0)	(0, 0)	(0, 0)
	N_0	cm^{-3}	$(1 \times 10^{18}, 1 \times 10^{18})$	$(1 \times 10^{18}, 1 \times 10^{18})$	$(1 \times 10^{18}, 1 \times 10^{18})$
			[35]	[35]	[35]
Nonlocal tunneling	g			(0.21, 0.4)	
	m_t/m_0			(0.09, 0.37)	
				[38]	

Note: Paired values in parentheses are specified for electrons and for holes, respectively.

TABLE 40.4 Table of Material Parameters for the GaInP/GaInAs/Ge Solar Cell Example.

	Symbol	**Units**	**$Al_{0.25}Ga_{0.25}In_{0.5}P$**	**MgF**	**TiOx**
	ϵ/ϵ_0		11.84	5.3	6
			[29]	[39]	[40]
	n, k		[20]	[41]	[42,43]
Band gap	$E_{g,0}$	eV	1.984		
	α	eV/K	1.76×10^{-4}		
	β	K	134.2		
	T_{ref}	K	300		
	χ_0	eV	3.997		
	ΔE_g	eV	[29]		
	$\Delta\chi/\Delta E_g$		0.5		
			[25]		
Density of states	m_n/m_0		0.1214		
	$N_{\text{C 300}}$	cm^{-3}	1.061×10^{18}		
	m_p/m_0		0.6244		
	$N_{\text{V 300}}$	cm^{-3}	1.24×10^{19}		
			[29,35]		
Mobility (Masetti model)	μ_{max}	$\text{cm}^2/\text{V.s}$			
	ζ				
	μ_{min1}	$\text{cm}^2/\text{V.s}$			
	μ_{min2}	$\text{cm}^2/\text{V.s}$			
	μ_1	$\text{cm}^2/\text{V.s}$			
	P_c	cm^{-3}			
	C_r	cm^{-3}			
	C_s	cm^{-3}			
	α				
	β				
Mobility (Arora model)	A_{min}	$\text{cm}^2/\text{V.s}$	(400.0, 15.0)		
	α_{m}		(0, 0)		
	A_{d}	$\text{cm}^2/\text{V.s}$	(3900, 135)		
	α_{d}		(0, 0)		
	A_N	cm^{-3}	$(2 \times 10^{16}, 1.5 \times 10^{17})$		
	α_N		(1.955, 1.47)		
	A_{a}		(0.7, 0.8)		
	α_{a}		(0, 0)		
			[36]		
Recombination	B_{rad}	cm^3/s	1×10^{-10}		
	τ_{max}	s	$(1 \times 10^{-9}, 1 \times 10^{-9})$		
	τ_{min}	s	(0, 0)		
	N_{ref}	cm^{-3}	$(1 \times 10^{19}, 1 \times 10^{19})$		
	γ		(1.0, 1.0)		
	T_α	K	(0, 0)		
	T_{coeff}		(0, 0)		
	A_{A}	cm^6s^{-1}	$(3 \times 10^{-30}, 3 \times 10^{-30})$		
	B_{A}	cm^6s^{-1}	(0, 0)		
	C_{A}	cm^6s^{-1}	(0, 0)		
	H		(0, 0)		
	N_0	cm^{-3}	$(1 \times 10^{18}, 1 \times 10^{18})$		
			[35]		
Nonlocal tunneling	g				
	m_{t}/m_0				

Note: Paired values in parentheses are specified for electrons and for holes, respectively.

TABLE 40.5 Table of Interface Parameters for the GaInP/GaInAs/Ge Solar Cell Example.

	Symbol	Units	TiOx/AlInP	AlInP/GaInP	GaAs/AlGaAs	GaInP/Ge
Interface recombination	S_0	cm/s	0 fitted	$(2\times10^5, 2\times10^5)$ [44]/fit to data	(200, 200) [44]	$(1\times10^5,\ 1\times10^5)$ [6]/fit to data

Note: Paired values in parentheses are specified for electrons and for holes respectively.

The Arora model for mobility uses a set of temperature-corrected constants,

$$N_0 = A_N \left(\frac{T}{300\ \mathrm{K}}\right)^{\alpha_N}; A^* = A_\mathrm{a} \left(\frac{T}{300\ \mathrm{K}}\right)^{\alpha_\mathrm{a}} \tag{40.24}$$

$$\mu_\mathrm{min} = A_\mathrm{min} \left(\frac{T}{300\ \mathrm{K}}\right)^{\alpha_\mathrm{m}}; \mu_\mathrm{d} = A_\mathrm{d} \left(\frac{T}{300\ \mathrm{K}}\right)^{\alpha_\mathrm{d}}. \tag{40.25}$$

Mobility is then calculated with

$$\mu = \mu_\mathrm{min} + \frac{\mu_\mathrm{d}}{1 + \left[(N_\mathrm{A} + N_\mathrm{D})/N_0\right]^{A^*}}. \tag{40.26}$$

40.5.2 Recombination, Lifetimes, and Surface/Interface Recombination Velocities

Having defined some of the basic material properties, we can now look at the recombination rates, each of which have their own material parameters. Here we present the rates as they are implemented in the Sentaurus software [21]; similar formulas can be found in texts on device simulation such as Palankovski [31]. The radiative recombination rate can be written as

$$R_\mathrm{rad} = B_\mathrm{rad}(np - \gamma_n \gamma_p n_\mathrm{i,eff}^2), \tag{40.27}$$

where γ_n and γ_p are correction factors which are needed for simulations using Fermi–Dirac statistics, and are defined by

$$\gamma_n = \frac{n}{N_\mathrm{C}} \exp\frac{E_\mathrm{C} - E_{\mathrm{F}n}}{kT}, \tag{40.28}$$

$$\gamma_p = \frac{p}{N_\mathrm{V}} \exp\frac{E_{\mathrm{F}p} - E_\mathrm{V}}{kT}. \tag{40.29}$$

The SRH recombination rate is somewhat more complicated. We begin by finding the doping-dependent minority carrier lifetimes for the SRH process,

$$\tau_\mathrm{SRH} = \tau_\mathrm{min} + \frac{\tau_\mathrm{max} - \tau_\mathrm{min}}{1 + \left(\frac{N_\mathrm{A}+N_\mathrm{D}}{N_\mathrm{ref}}\right)^\gamma}. \tag{40.30}$$

This calculation will be done separately using different parameters for the electron and hole minority carrier lifetimes. Then we can write the expression for the SRH recombination rate,

$$R_\mathrm{SRH} = \frac{np - \gamma_n \gamma_p n_\mathrm{i,eff}^2}{\tau_{\mathrm{SRH},p}(n + \gamma_n n_1) + \tau_{\mathrm{SRH},n}(p + \gamma_p p_1)}, \tag{40.31}$$

where n_1 and p_1 are

$$n_1 = n_{\text{i,eff}} \exp\left(\frac{E_{\text{trap}}}{kT}\right); \quad p_1 = n_{\text{i,eff}} \exp\left(\frac{-E_{\text{trap}}}{kT}\right). \tag{40.32}$$

Note that we have not included a temperature dependence of the lifetimes, or any enhancement of the SRH process due to electric field, but these could be significant in some devices. E_{trap} is the position of the trap level relative to the intrinsic level; we generally take it to be zero.

Finally, we write the Auger recombination rate,

$$R_{\text{Auger}} = (C_n n + C_p p)(np - n_{\text{i,eff}}^2). \tag{40.33}$$

Due to the long nonradiative lifetimes in the bulk of many III–V materials and the relatively thin layers, interface recombination can be a significant factor in the performance of these devices. The values of surface recombination velocity can vary widely due to various growing conditions, and so this along with the SRH lifetime can be treated as a "fitting parameter" rather than a known material property [44]. The surface recombination rate is treated with an SRH model,

$$R_{\text{surf}} = \frac{np - \gamma_n \gamma_p n_{\text{i,eff}}^2}{(n + \gamma_n n_1)/S_p + (p + \gamma_p p_1)/S_n}, \tag{40.34}$$

where S_n and S_p are the electron and hole surface recombination velocities.

40.5.3 Tunnel Diodes

The tunnel diodes that allow for the flow of carriers between subcells in MJSCs operate via band-to-band tunneling across a degenerately doped $p{+}{+}/n{+}{+}$ junction [45]. The depletion width in this case will be extremely narrow (~5 nm), with carrier populations at similar energy levels on both sides of the junction (Figure 40.5a). When a small bias is applied, the free electrons become aligned in energy with holes on the opposite side, and can tunnel across the junction, creating a net current.

A typical J–V curve of a tunnel diode is shown in Figure 40.5b. Under normal conditions, the tunnel diode should remain in the ohmic region, where large currents can be conducted with minimal potential difference across the tunnel diode. This requires that the peak tunneling current density should be significantly greater than the solar cell's short-circuit current density J_{SC}. If the bias voltage is increased beyond the ohmic region, the free electrons and holes are no longer aligned in energy and the tunneling current decreases. Eventually, the diode will be forward biased and will conduct via normal mechanisms.

Many drift-diffusion-based simulators allow for modeling band-to-band tunneling using approaches derived from the work of Kane et al. [46] and others [47,48]. These models can reproduce all of the behavior of a tunnel junction [37,38,49] including the ohmic region, tunneling peak and excess current region of the tunnel diode J–V curve [30].

Tunnel diodes typically have minimal impact on the performance of MJSCs, except in pathological cases or where the device is operated at very high current density. For many studies, a simplistic band-to-band tunneling model based only on local quantities will be sufficient. For example, in Synopsys Sentaurus, there is a model that adds a band-to-band recombination term to the continuity equation following the work of Hurkx [47],

$$R^{\text{bb}} = A \frac{np - n_{\text{i,eff}}^2}{(n + n_{\text{i,eff}})(p + n_{\text{i,eff}})} \left(\frac{|F|}{1\text{V/cm}}\right)^{\sigma} \exp\left(-\frac{F_0}{|F|}\right), \tag{40.35}$$

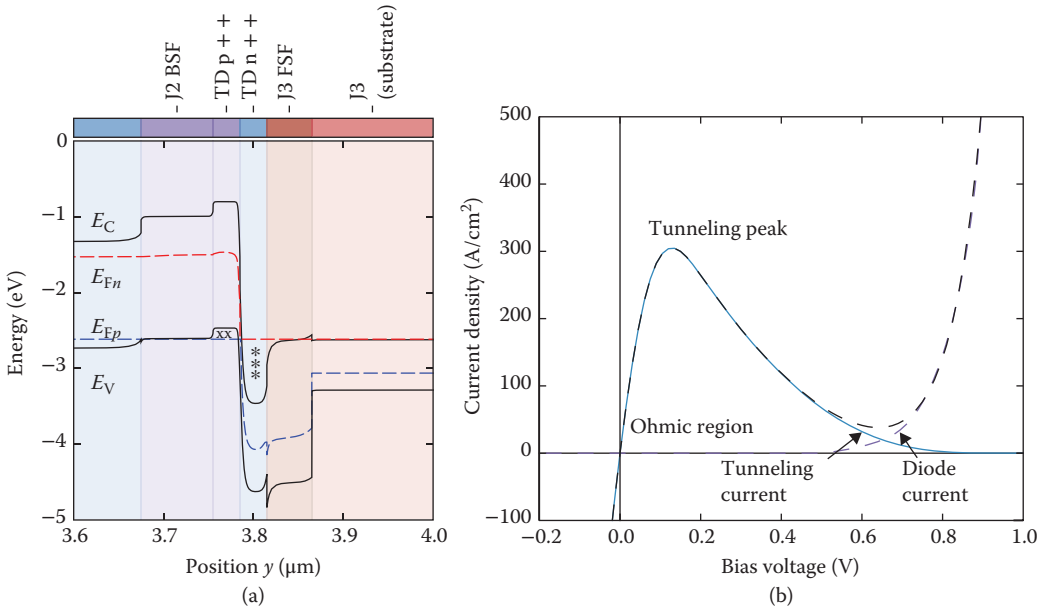

FIGURE 40.5 (a) Detail of the bottom tunnel-diode portion of the cross-sectional band diagram. The tunnel diode is short circuited. "xx" symbols represent holes and "∗" symbols represent electrons that are available to tunnel across the junction into the opposite band. (b) J–V curve of a tunnel diode. Contributions from band-to-band tunneling and the diode diffusion current are shown.

where n and p are the electron and hole concentrations, $n_{i,eff}$ is the effective intrinsic carrier concentration, and F is the local electric field strength. σ has a value of 2 for direct transitions and 5/2 for indirect transitions. A and F_0 are material-dependent parameters that must be determined experimentally. This type of model should be adequate to give good results in cases where tunnel diodes are not dominating the behavior of an MJSC, but because it does not consider the variation of the band profile and material properties across the tunneling path, it will not be physically accurate.

A more accurate, but also more computationally intensive, approach uses a nonlocal band-to-band tunneling model combined with trap-assisted tunneling [50]. A "nonlocal" model is one where the tunneling current at a given position is calculated based on quantities at other locations, such as the potential profile across the tunneling path, and carriers are effectively removed from one location in the structure and reinserted at a different location. The use of this type of nonlocal model is described in detail in [38] and [51]. Implementation in Sentaurus is shown in Reference [52]. Figure 40.5b was simulated using nonlocal models for band-to-band tunneling as well as SRH recombination enhanced by trap-assisted tunneling. The individual contributions from each of these two mechanisms is shown in the figure. Some typical values of the tunneling masses m_t and effective Richardson constants g are included in Tables 40.2 and 40.3. Readers should refer to references [38,50–52] for interpretation of these parameters.

It is our experience that, while it it very possible to fit tunnel-diode parameters to a measured J–V curve, the resulting model tends not to be predictive for varying tunnel junction designs (i.e., varying doping levels or material compositions). The peak tunneling current of devices can vary by several orders of magnitude for relatively small changes in doping or material composition.

40.6 Current–Voltage Calculation

A J–V curve calculation is performed by first calculating the optical generation as a function of position for the appropriate illumination conditions, then solving the drift-diffusion problem while ramping bias

at the contacts. It can also be useful to generate J–V curves for individual subcells. Additional "virtual" ohmic contact boundary conditions are defined at locations in between the subcells in order to bias a subcell individually. The current through a given subcell is then the sum of currents at all contacts above it (Figure 40.6).

It's very informative to study the subcell dark J–V curves on a log-current axis, as the contributions from various physical processes can be distinguished [2]. In Figure 40.6c, the dark J–V curve of the (In)GaAs subcell is plotted along with the recombination currents due to SRH, radiative, and Auger mechanisms. The recombination currents are calculated as

$$J_i = \frac{q}{A}\int_V R_i \, \mathrm{d}V, \tag{40.36}$$

where q is the electron charge, R is the recombination rate per unit volume, A is the cell area and i refers to the recombination process in question. The integral is evaluated over the volume of a subcell. Comparing simulated dark-current curves against measured single-junction devices can be a good first step to validating multijunction simulations. When plotted on a log axis, the slope of the curve is directly related to the diode ideality factor. In particular, the SRH mechanism usually produces a region where the ideality factor $n_{\text{ideal}} \simeq 2$; this gives an excellent method to verify that the SRH current is modeled accurately independent of other factors. In contrast, fitting the SRH lifetimes based on V_{OC} measured at high intensity (1 sun or greater) would not be effective as the cell is not strongly dominated by the SRH process at that level of bias.

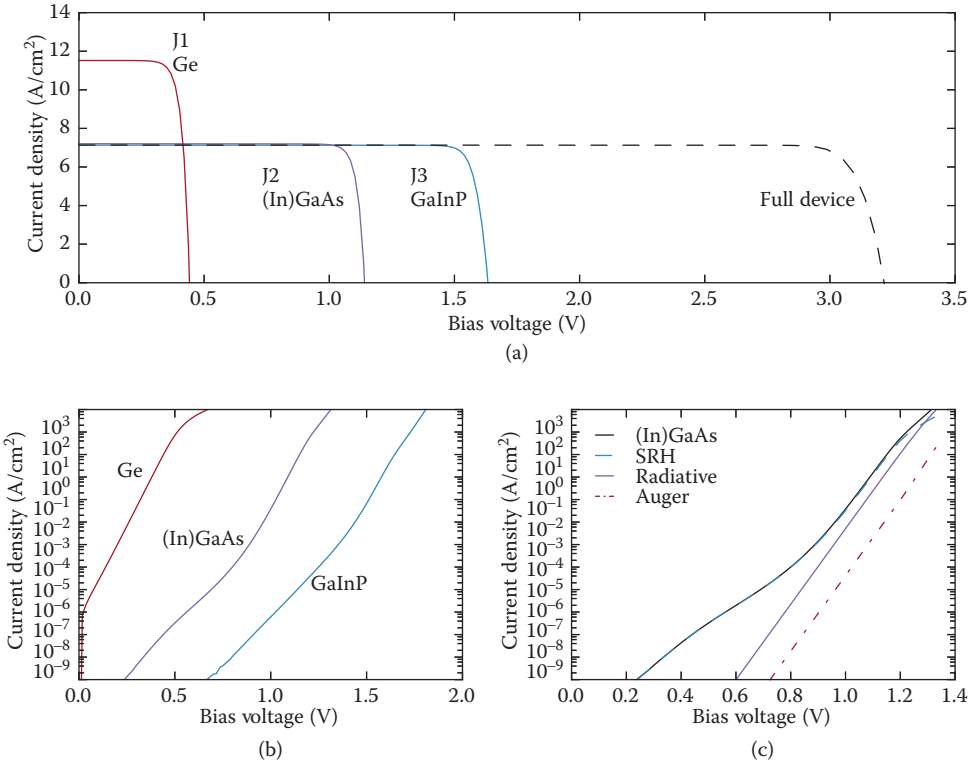

FIGURE 40.6 (a) J–V curves of the subcells and the triple-junction device under 500 suns, AM1.5D illumination. (b) Dark current curves of each of the subcells. (c) Dark J–V curve of (In)GaAs with the contributions from individual processes.

40.7 Quantum Efficiency Calculation

Experimentally, QE of multijunction cells is measured by using bias lights to ensure that the subcell of interest is limiting the overall device current, as described in Reference [53]. In this way, it is possible to probe each of the subcells individually. This experiment can also be replicated in simulation by adding bias illumination sources, or simply adding a fixed amount of carrier generation to the non-limiting subcells, in addition to a monochromatic source, which is scanned in wavelength over the range of interest. The bias light intensity must be adjusted such that the short-circuit currents of the non-probed junctions are greater than the current produced by the probed junction at any wavelength; this requires some iteration to find the correct bias intensities.

An alternative approach that can be implemented in simulation is to simply define ohmic "virtual contacts" to the front surface field and back surface field layers of each subcell. This provides a more direct control over the voltage bias being applied to each subcell than the light-biasing method, but should generally yield identical results. It also has advantages over the light-biasing method in that there is no iteration required to find the correct setup, and where the light-biasing method requires a separate sweep over wavelength for each junction, with the "virtual contact" method the short-circuit currents and QE plots of all triple junctions can be collected in one wavelength sweep. The simulated QE for the triple-junction solar cell, calculated using this method, is presented in Figure 40.7. The simulation will output short-circuit current density as a function of wavelength λ for subcell i, $J_{SC,i}$. From the optical calculation, we also know the reflectivity from the top of the layer stack, $R(\lambda)$ and transmissivity through the layer stack, $T(\lambda)$. The external quantum efficiency (EQE) of a given subcell is defined as the ratio of short-circuit current density to incident photon flux density $\Phi(\lambda)$,

$$EQE_i(\lambda) = \frac{J_{SC,i}(\lambda)}{q\Phi(\lambda)}, \tag{40.37}$$

or in terms of the optical intensity $E(\lambda)$,

$$EQE_i(\lambda) = \frac{hc}{q\lambda}\frac{J_{SC,i}(\lambda)}{E(\lambda)}, \tag{40.38}$$

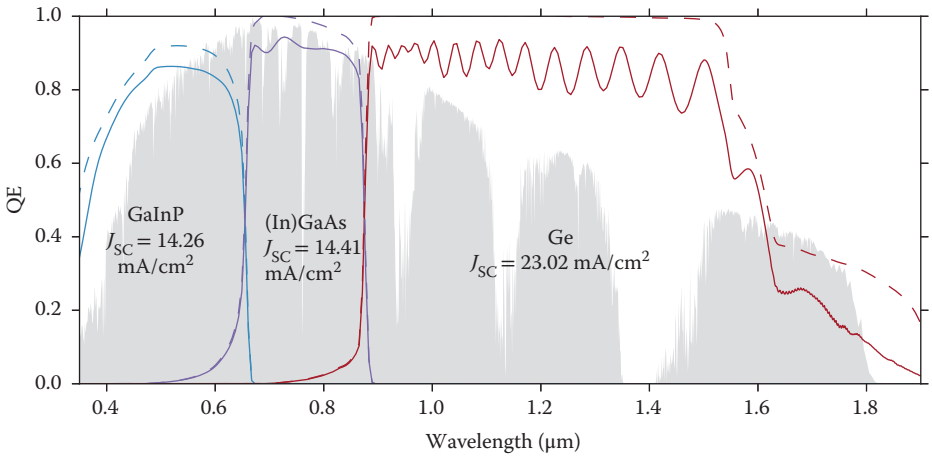

FIGURE 40.7 External (solid) and internal (dashed) quantum efficiency (EQE and IQE) of the triple-junction solar cell. The normalized AM1.5D solar spectral irradiance is shown filled in gray.

where h is Planck's constant, c is the speed of light and q is the electron charge. Internal quantum efficiency (IQE) is given by

$$IQE_i(\lambda) = \frac{EQE_i(\lambda)}{1 - R(\lambda) - T(\lambda)}. \tag{40.39}$$

QE curves as calculated in software often show Fabry–Pérot type oscillations at long wavelengths which do not appear in physical measurements. Frequently these oscillations are physically correct for a perfectly collimated beam, but in an experimental setup with a finite beam divergence and monochromator resolution these fine details cannot be resolved. In order to make meaningful comparisons between measured and simulated data, it can therefore be useful to convolve the simulated QE with a spectral function approximating the monochromator lineshape, thus giving a representation of the simulated QE as it would be measured by an instrument. We can write this effective QE as

$$EQE_{\text{eff}}(\lambda) = \int_{-\infty}^{\infty} EQE(\lambda + l)w(l)\mathrm{d}l, \tag{40.40}$$

where $EQE(\lambda)$ is the simulated QE, $w(\lambda)$ is a function approximating the spectral lineshape of the monochromator which is centered at $\lambda = 0$ and integrates to unity, and l is a variable of integration.

Besides calculating the experimentally measurable quantities R, T, EQE, and IQE, simulation also provides the opportunity for introspection – examining quantities, which would not be measurable. We can integrate various recombination rates over the volume of the device to quantify each of the loss mechanisms and explain any nonideality in the simulation. For example, integrating the SRH recombination rate over the volume of subcell i, we find a loss component

$$L_{\text{SRH},i}(\lambda) = \frac{1}{A\Phi(\lambda)} \int_{V_i} R_{\text{SRH}}(\lambda)\mathrm{d}V. \tag{40.41}$$

Similarly for interface recombination at the interfaces above and below subcell i,

$$L_{\text{surf},i}(\lambda) = \frac{1}{A\Phi(\lambda)} \left(\int_{A_{\text{em},i/\text{fsf},i}} R_{\text{surf}}\mathrm{d}A + \int_{A_{\text{base},i/\text{bsf},i}} R_{\text{surf}}\mathrm{d}A \right), \tag{40.42}$$

where the integrals are evaluated over the area of the front surface field/emitter and base/back surface field interfaces of the subcell in question. In general, it is possible to account for all EQE losses in this way, in which case the sum of all subcell EQEs and all loss components should be unity (Figure 40.8). Just as the EQE can be integrated with incident spectrum to give J_{SC} of each subcell, the loss components can be integrated to indicate the amount of photocurrent lost via a particular mechanism.

40.8 Photon Recycling and Luminescent Coupling

Some MJSCs show significant effects due to photon recycling, which can enhance the voltage of individual subcells, and luminescent coupling which can redistribute photocurrent between junctions and reduce sensitivity to spectrum [54,55].

To understand this, we can modify the circuit model of Figure 40.1b with two diodes per subcell, one of which represents the nonradiative recombination current and one the radiative current, $J_{\text{rad},i}$ (Figure 40.9). A portion of the radiatively emitted current results in generation within other junctions, according to

$$J_{\text{LC},i} = \sum_j K_{i,j} J_{\text{rad},j}, \tag{40.43}$$

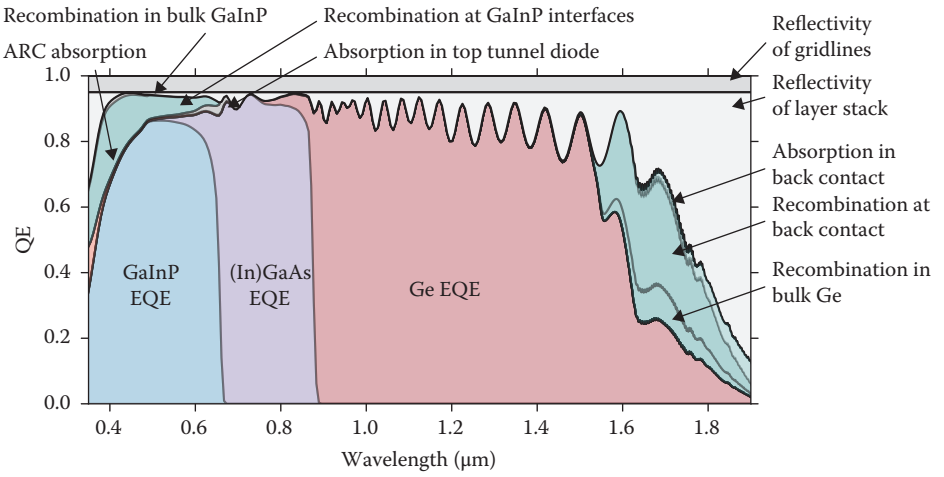

FIGURE 40.8 Quantum efficiency plot showing the breakdown of losses. (ARC, antireflection coating; EQE, external quantum efficiency.)

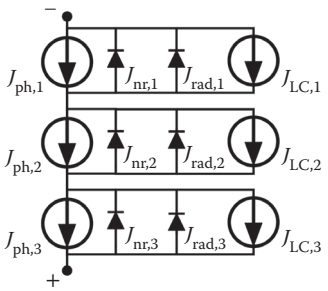

FIGURE 40.9 Equivalent circuit model of a triple-junction cell, modified to distinguish radiative and nonradiative diode currents and to include coupled generation.

where $K_{i,j}$ is a junction-to-junction coupling coefficient. If any of the photocurrents $J_{ph,i}$ are in excess of the device current J, that excess current will recombine through the nonradiative and radiative diodes in Figure 40.9. At sufficiently high bias, GaAs and GaInP junctions have predominantly (>90%) radiative recombination. These junctions are thick enough that the self-coupling coefficient $K_{i,i}$ can also be >90%, meaning that carriers can cycle repeatedly through the process of radiative recomination, photon emission, and reabsorption within the same junction. This process is called "photon recycling" and can be responsible for an increase in carrier concentrations.

A related term, "luminescent coupling," refers to the coupling between different junctions. If one of the bottom junctions is current limiting (i.e., it has the samllest J_{ph} of any subcell), then the excess current in the upper junctions can be radiatively coupled into the bottom junction, allowing a J_{SC} greater than the photocurrent in the limiting junction.

Luminescent coupling can also be observed in experimental measurements of multijunction cell QE. When bias lights are applied to the top two subcells, some of that bias light is coupled into the subcell that is being probed, interfering with the measurement [56].

These effects can be modeled in a drift-diffusion device simulator with the addition of a coupled generation term to the continuity equation [57,58], or through a process of successive approximations to the optical generation function [59].

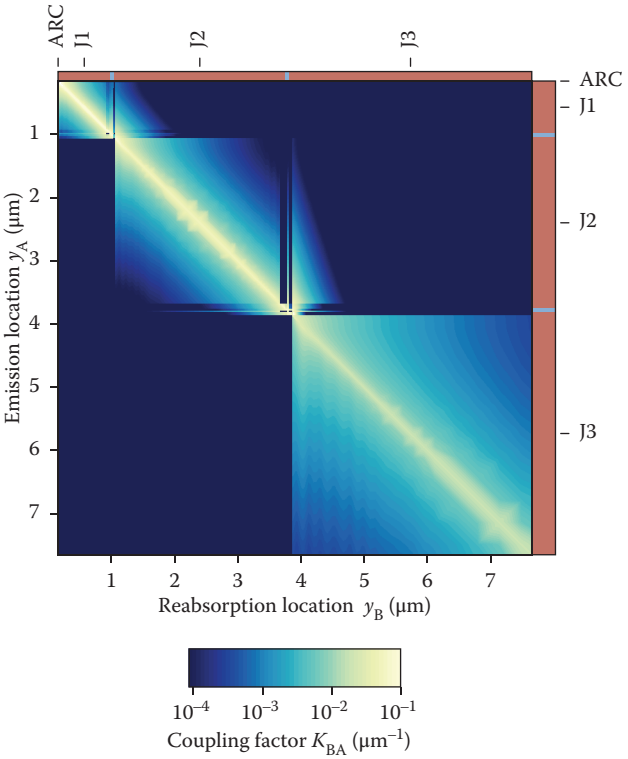

FIGURE 40.10 Luminescent coupling matrix for the triple-junction GaInP/(In)GaAs/Ge example structure. (ARC, antireflection coating.)

We discuss the first option here, and begin by modifying the process flow outlined in Figure 40.2 and insert a step after the assignment of material parameters for calculation of the coupling matrix, which gives the probability that a photon emitted from a position A will be reabsorbed at another position, B (shown in Figure 40.10). Naturally, photons have a great probability of being reabsorbed near the point where they are emitted (i.e., along the main diagonal of the matrix), or in layers of smaller bandgap such as a lower subcell. Indeed, the figure shows significant absorption in the GaAs junction (J2) due to emissions from J1, and also absorption in J3 from emissions in J2. This coupled transfer is unidirectional; there is no coupling of emitted photons from the lower subcells into the upper ones. Also, it should be noted that Figure 40.10 indicates only the efficiency with which emitted photons are reabsorbed. Some layers may not emit strongly at all depending on the local carrier concentration and the ratio of radiative and nonradiative lifetimes.

Once the coupling matrix has been calculated, the simulation can proceed as normal with a modification to the electron and hole continuity equations to add the coupled generation term which is a function of radiative recombination rates at all positions within the device. While the simulator used here, Synopsys Sentaurus, does not directly provide a method to add this term, it does provide a facility to define custom generation/recombination terms through a C++ programming interface. The custom C++ code is executed at each iteration of the nonlinear equation solver, and has access to quantities such as carrier concentrations (and by extension, recombination rates) at all mesh vertices within the simulation domain. The technique is described in detail in the appendix to Reference [58].

The result of this calculation can be seen in Figure 40.11; the "plateau" in QE between 650 and 880 nm is due to light that is radiatively coupled from the (In)GaAs junction into the germanium junction, and is an undesirable effect as it interferes with measuring the actual germanium photocurrent $J_{ph,3}$.

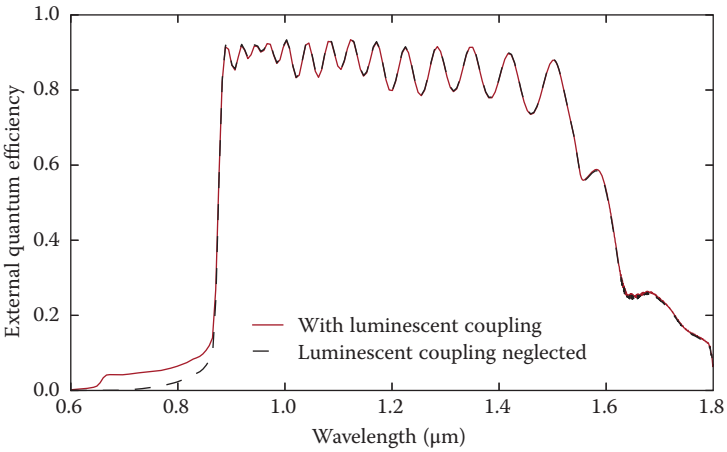

FIGURE 40.11 Calculated quantum efficiency of the germanium junction including coupled generation from light emitted by the (In)GaAs junction.

In multijunction phototransducers, which are devices closely related to MJSCs but are designed to operate under single-wavelength illumination, the luminescent coupling causes a significant broadening of the wavelength response. This broadening effect has also been modeled using this technique[58,60].

In contrast with the circuit-model representation, this treatment of luminescent coupling using a drift-diffusion simulation provides some advantages. The radiative emission rate, and the resulting coupled generation, is calculated as a function of position and reflects local changes in carrier density. The rates of emission from the emitter, the space charge region, and the base will each vary depending on the bias conditions, and the simulation will handle these changes appropriately. Areas that are shaded by gridlines and busbars will also be treated correctly. Furthermore, it is straightforward to consider coupling between all layer combinations in the device structure, not just those layers that comprise the photovoltaic junctions.

40.9 Summary

MJSCs are relatively complex devices, with typically 25 or more heterointerfaces and a number of photovoltaic junctions and tunnel diodes. With sufficient calibration, models can be developed that provide a very accurate representation of the device performance and can indicate the relative importance of various loss mechanisms. Once the model is calibrated, a detailed design of experiments can be done to pinpoint optimal design for a given requirement. In order to do accurate simulations, an extensive set of material parameters must be collected for each of the semiconductor materials. In particular, the complex refractive index, the minority carrier lifetimes, and the tunneling parameters tend to be difficult to find in the literature and need to be found through experiment.

In many cases, lateral current spreading and uneven bias across the surface of the cell have significant impacts on efficiency, which cannot be represented in a simple equivalent circuit model. In concentrating systems that have chromatic aberrations, the illumination of each subcell will have a different, nonuniform distribution over the the cell aperture. This leads to local mismatch in the subcell current densities and is an interesting problem to study through device simulation. Simulations based on drift-diffusion enable the device to be simulated including transport across all heterojunctions, without any presupposition of diode-like behavior, and hence this type of model can be used to study the effects of extreme operating conditions or manufacturing errors where the diode model is not applicable.

References

1. M. A. Green, K. Emery, Y. Hishikawa,W. Warta, and E. D. Dunlop. Solar cell efficiency tables (version 47). *Prog. Photovolt. Res. Appl.*, 24(1):3–11, 2016.
2. J. Nelson. *Physics of Solar Cells*. London: Imperial College Press, 2003.
3. C. H. Henry. Limiting efficiencies of ideal single and multiple energy gap terrestrial solar cells. *J. Appl. Phys.*, 51(8):4494, 1980.
4. R. R. King, D. Bhusari, A. Boca, D. Larrabee, X. Liu, W. Hong, C. M. Fetzer, D. C. Law, and N. H. Karam. Band gap-voltage offset and energy production in next-generation multijunction solar cells. *Prog. Photovolt. Res. Appl.*, 19:797–812, 2011.
5. R. R. King, N. H. Karam, J. H. Ermer, N. Haddad, P. Colter, T. Isshiki, H. Yoon, H. L. Cotal, D. E. Joslin, D. D. Krut, R. Sudharsanan, K. Edmondson, B. T. Cavicchi, and D. R. Lillington. Next-generation, high-efficiency III-V multijunction solar cells. In *Conference Record 28th IEEE Photovoltaic Specialists Conference.*, pages 998–1001, Anchorage, AK, 2000. IEEE.
6. D. J. Friedman and J. M. Olson. Analysis of Ge junctions for GaInP/GaAs/Ge three-junction solar cells. *Prog. Photovolt. Res. Appl.*, 9(3):179–189, 2001.
7. D. J. Friedman, J. M. Olson, S. Ward, T. Moriarty, K. Emery, S. Kurtz, A. Duda, R. R. King, H. L. Cotal, D. R. Lillington, J. H. Ermer, and N. H. Karam. Ge concentrator cells for IIIV multijunction devices. In *Conference Record 28th IEEE Photovoltaic Specialists Conference*, Anchorage, AK, pages 965–967. IEEE, 2000.
8. I. García, C. Algora, I. Rey-Stolle, and B. Galiana. Study of non-uniform light profiles on high concentration III-V solar cells using quasi-3D distributed models. In *2008 33rd IEEE Photovolatic Specialists Conference*, San Diego, CA, pages 1–6. IEEE, May 2008.
9. Synopsys Inc. Simulation of a GaAs / GaInP Dual-Junction Solar Cell, Synopsys Inc., Mountain View, CA, 2012.
10. Solarex17.in: EQE of III-V Tandem Cell. Silvaco Inc., Santa Clara, CA, 2015. http://www.silvaco.com/examples/tcad/section44/example17/index.html
11. Z. Q. Li, Y. G. Liao, and Z. M. S. Li, Modelling of multi-junction solar cells by Crosslight APSYS, High and Low Concentration for Solar Electric Applications, *Proceedings of the SPIE*, vol. 6339, pp. 633909–15, San Diego, CA, 2006.
12. W. E. McMahon, J. M. Olson, J. F. Geisz, and D. J. Friedman. An examination of 1D solar cell model limitations using 3D SPICE modeling. In *38th IEEE Photovoltaic Specialists Conference.*, pp. 002088–002091. Austin, TX, Jun 2012.
13. G. Létay, M. Hermle, and A. W. Bett. Simulating single-junction GaAs solar cells including photon recycling. *Prog. Photovolt. Res. Appl.*, 14(8):683–696, 2006.
14. T. Rahman and K. Fobelets. Efficient tool flow for 3D photovoltaic modelling. *Comput. Phys. Commun.*, 193:124–130, 2015.
15. P. Sharma, A. W. Walker, J. F. Wheeldon, K. Hinzer, and H. Schriemer. Enhanced efficiencies for high concentration, multijunction PV systems by optimizing grid spacing under nonuniform illumination. *Int. J. Photoenergy*, 2014:582083, 2014.
16. ASTM International. ASTM G173-03: Standard Tables for Reference Solar Spectral Irradiances: Direct Normal and Hemispherical on 37 degree Tilted Surface. Technical report, ASTM International, 2012.
17. D. E. Aspnes, S. M. Kelso, R. A. Logan, and R. Bhat. Optical properties of AlxGa1−xAs. *J. Appl. Phys.*, 60(2):754, 1986.
18. T. Kim, T. Ghong, Y. Kim, S. Kim, D. Aspnes, T. Mori, T. Yao, and B. Koo. Dielectric functions of InxGa1−xAs alloys. *Phys. Rev. B.*, 68(11):115323, 2003.
19. A. B. Djurisic, A. D. Rakic, P. C. K. Kwok, E. H. Li, M. L. Majewski, and J. M. Elazar. Modeling the optical constants of Al(x)Ga(1-x)As alloys. *J. Appl. Phys.*, 86(1):445–451, 1999.

20. M. Schubert, J. A. Woollam, G. Leibiger, B. Rheinlander, I. Pietzonka, T. Sab, and V. Gottschalch. Isotropic dielectric functions of highly disordered Al*x*Ga1−*x*InP (0 <= *x* <= 1) lattice matched to GaAs. *J. Appl. Phys*., 86(4):2025–2033, 1999.
21. Sentaurus Device User Guide, version K-2015. Synopsys Inc., Mountain View, CA. 1446p, 2015.
22. D. Schroeder. *Modelling of Interface Carrier Transport for Device Simulation*. Wien: Springer-Verlag, 1994.
23. K. Horio and H. Yanai. Numerical modeling of heterojunctions including the thermionic emission mechanism at the heterojunction interface. *IEEE Trans. Electron Dev.*, 37(4):1093–1098, 1990.
24. P. P. Altermatt. Models for numerical device simulations of crystalline silicon solar cells—a review. *J. Comput. Electron*., 10(3):314–330, 2011. Multi-Junction Solar Cells 40–43
25. I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan. Band parameters for III-V compound semiconductors and their alloys. *J. Appl. Phys*., 89(11):5815, 2001.
26. D. H. Levi. Effects of ordering on the optical properties of $GaInP_2$. *Proc. SPIE*., 5530:326–337, 2004.
27. G. Masetti, M. Severi, and S. Solmi. Modeling of carrier mobility against carrier concentration in arsenic-doped, phosphorus-doped, and boron-doped silicon. *IEEE Trans. Electron Dev*., 30(7):764–769, 1983.
28. N. D. Arora, J. R. Hauser, and D. J. Roulston. Electron and hole mobilities in silicon as a function of concentration and temperature. *IEEE Trans. Electron Dev*., 29(2):292–295, 1982.
29. J. Piprek. *Semiconductor Optoelectronic Devices: Introduction to Physics and Simulation*. San Diego, CA: Academic Press, 2003.
30. S. M. Sze and K. Ng Kwok. *Physics of Semiconductor Devices*, 3rd edition. Hoboken, NJ: Wiley-Interscience, 2007.
31. V. Palankovski and R. Quay. *Analysis and Simulation of Heterostructure Devices*. Vienna: Springer-Verlag, 2004.
32. S. Adachi. *Optical Constants of Crystalline and Amorphous Semiconductors*. Boston, MA: Springer, 1999.
33. E. D. Palik, editor. *Handbook of Optical Constants of Solids*. Cambridge, MA: Academic Press, 1998.
34. E. F. Schubert. *Physical Foundations of Solid-State Devices*. Rensselaer Polytechnic Institute, Troy, NY, 273 pages, 2006.
35. M. Levinshtein, S. Rumyantsev, and M. Shur, editors. *Handbook Series on Semiconductor Parameters*, vol. 2: Ternary And Quaternary III-V Compounds. Singapore: World Scientific, 1999.
36. M. Sotoodeh, A. H. Khalid, and A. A. Rezazadeh. Empirical low-field mobility model for III-V compounds applicable in device simulation codes. *J. Appl. Phys*., 87(6):2890, 2000.
37. J. F. Wheeldon, C. E. Valdivia, A. W. Walker, G. Kolhatkar, A. Jaouad, A. Turala, B. Riel, D. Masson, N. Puetz, S. Fafard, R. Arès, V. Aimez, T. J. Hall, and K. Hinzer. Performance comparison of AlGaAs, GaAs and InGaP tunnel junctions for concentrated multijunction solar cells. *Prog. Photovolt. Res. Appl*., 19(4):442–452, 2010. 40–44 Book title goes here.
38. A. W. Walker, O. Thériault, M. M. Wilkins, J. F. Wheeldon, and K. Hinzer. Tunnel-junction- limited multijunction solar cell performance over concentration. *IEEE J. Sel. Top. Quantum Electron*., 19(5):4000508, 2013.
39. J. Fontanella, C. Andeen, and D. Schuele. Low-frequency dielectric constants of α-quartz, sapphire, MgF2, and MgO. *J. Appl. Phys*., 45(7):2852–2854, 1974.
40. D. R. Lide, editor. *CRC Handbook of Chemistry and Physics,* 86th edition. Boca Raton, FL: CRC Press, 2005.
41. J. M. Siqueiros, R. Machorro, and L. E. Regalado. Determination of the optical constants of MgF(2) and ZnS from spectrophotometric measurements and the classical oscillator method. *Appl. Opt*., 27(12):2549–2553, 1988.
42. J. R. Devore. Refractive indices of rutile and sphalerite. *J. Opt. Soc. Am*., 41(6):416, 1951.

43. J. Kischkat, S. Peters, B. Gruska, M. Semtsiv, M. Chashnikova, M. Klinkmüller, O. Fedosenko, S. Machulik, A. Aleksandrova, G. Monastyrskyi, Y. Flores, and W. T. Masselink. Mid-infrared optical properties of thin films of aluminum oxide, titanium dioxide, silicon dioxide, aluminum nitride, and silicon nitride. *Appl. Opt*., 51(28):6789–6798, 2012.
44. S. R. Kurtz, J. M. Olson, D. J. Friedman, J. F. Geisz, and A. E. Kibbler, Passivation of interfaces in high-efficiency photovoltaic devices, In *MRS Online Proceedings Library Archive*, vol. 573, pp. 95–106, San Francisco, CA, 1999.
45. W. Guter and A. W. Bett. IV-Characterization of Devices Consisting of Solar Cells and Tunnel Diodes. In *4th World Conference Photovoltaic Energy Conversion*., pages 749–752. Waikoloa, HI, 2006.
46. E. O. Kane. Theory of tunneling. *J. Appl. Phys*., 32(1):83, 1961.
47. G. A. M. Hurkx, D. B. M. Klaassen, and M. P. G. Knuvers. A new recombination model for device simulation including tunneling. *IEEE Trans. Electron Dev.*, 39(2):331–338, 1992.
48. A. Ajoy. Complex bandstructure of direct bandgap III-V semiconductors: application to tunneling. In M. Katiyar, B. Mazhari, and Y. N. Mohapatra, editors, *International Workshop on Physics of Semiconductor Devices*, Kanpur, India, 2012.
49. J. F. Wheeldon, C. E. Valdivia, A. Walker, G. Kolhatkar, T. J. Hall, K. Hinzer, D. Masson, S. Fafard, A. Jaouad, A. Turala, R. Ares, and V. Aimez. AlGaAs tunnel junction for high efficiency multijunction solar cells: Simulation and measurement of temperature-dependent operation. In *34th IEEE Photovoltaic Specialists Conference*, Philadelphia, PA, pages 000106–000111, 2009.
50. M. Baudrit and C. Algora. Tunnel diode modeling, including nonlocal trap-assisted tunneling: A focus on III–V multijunction solar cell simulation. *IEEE Trans. Electron Dev*., 57(10):2564–2571, 2010.
51. M. Hermle, G. Létay, S. P. Philipps, and A. W. Bett. Numerical simulation of tunnel diodes for multijunction solar cells. *Prog. Photovolt. Res. Appl*., 16:409–418, 2008.
52. Synopsys Inc. Simulation of GaAs Tunnel Diode for Multijunction Solar Cells, Synopsys, Inc., Mountain View, CA, 9 pages, 2012.
53. K. Emery, M. Meusel, R. Beckert, A. W Bett, and W. Warta. Procedures for evaluating multijunction concentrators. In *Conference Record 28th IEEE Photovoltaic Specialists Conference*., pages 1126–1130. Anchorage, AK, 2000.
54. D. J. Friedman, J. F. Geisz, and M. A. Steiner. Effect of luminescent coupling on the optimal design of multijunction solar cells. *IEEE J. Photovolt*., 4(3):986–990, 2014.
55. D. J. Friedman, J. F. Geisz, and M. A. Steiner. Analysis of multijunction solar cell current voltage characteristics in the presence of luminescent coupling. *IEEE J. Photovolt*., 3(4):1429–1436, 2013.
56. S. H. Lim, J.-J. Li, E. H. Steenbergen, and Y.-H. Zhang. Luminescence coupling effects on multijunction solar cell external quantum efficiency measurement. *Prog. Photovolt. Res. Appl*., 21(3):344–350, 2013.
57. M. Wilkins, A. M. Gabr, A. H. Trojnar, H. Schriemer, and K. Hinzer. Effects of luminescent coupling in single- and 4-junction dilute nitride solar cells. In *Proceeding 40th Photovoltaic Specialists Conference*, pages 6–9, Denver, CO, 2014.
58. M. Wilkins, C. E. Valdivia, A. M. Gabr, D. Masson, S. Fafard, and K. Hinzer. Luminescent coupling in planar opto-electronic devices. *J. Appl. Phys*., 118(14):143102, 2015.
59. A. W. Walker, O. Hohn, D. N. Micha, L. Wagner, H. Helmers, A. W. Bett, and F. Dimroth. Impact of photon recycling and luminescence coupling on III - V photovoltaic devices. In *Proc. SPIE*, vol. 9358, San Francisco, CA, pages 11–13, 2015.
60. M. Wilkins, C. E. Valdivia, S. Chahal, M. Ishigaki, D. P. Masson, S. Fafard, and K. Hinzer. Performance impact of luminescent coupling on monolithic 12-junction phototransducers for 12 V photonic power systems. In *Proc. of SPIE*, vol. 9743, San Francisco, CA, 2016.

41

Nanostructure Solar Cells

Urs Aeberhard

41.1 Introduction: Nanostructures in Photovoltaics

The operation of solar cell (SC) devices is governed by a large variety of solid-state properties associated with the interplay of optical, electronic, and vibrational degrees of freedom of the component materials. Light-matter interaction provides the elementary optoelectronic processes of photogeneration and radiative recombination, while coupling of charge carriers to lattice vibrations leads to energy conversion losses due to thermalization, dissipative transport, and nonradiative recombination.

In the presence of spatial inhomogeneities, the optoelectronic material properties change if the characteristic length scale of the spatial variation is reduced below a certain threshold. The critical length scale depends on the physical degree of freedom (optical, electronic, vibrational) and corresponds to the wavelength of the associated wave (electromagnetic [EM] wave, de Broglie wave, sound wave). In this *subwavelength* regime, the physics is governed by the corresponding wave equations (Schrödinger, Maxwell, harmonic lattice dynamics), with the peculiar features such as interference or nonlocality. Along with the states, the interaction of different degrees of freedom is strongly modified in the presence of confinement, mainly due to the effect of localization and symmetry breaking. This tunability of physical properties via variation of material composition, size and shape of structures below a characteristic length scale—termed *nanostructures* (NSs)[†]—provides the basis for the vast field of nanotechnological applications, and the relation between configurational parameters and target NS functionality is a major topic of past and current research. Almost since the advent of semiconductor NS fabrication technologies, such as molecular beam epitaxy (MBE) or metal-organic chemical vapor deposition (MOCVD), but also lithographic processes,

[†] While the term *nano* is used for all of the above degrees of freedom (nanophotonics, -electronics, -phononics) to distinguish the confinement regime from bulk behavior, the characteristic confinement length varies strongly (0.1 μm to 1 cm for photons, 0.5 to 10 nm for electrons, 5 nm to 1 cm for phonons).

the peculiar physical properties of the NS produced in this way have also been utilized for the design of novel SC devices with the potential of increased photovoltaic (PV) conversion efficiency. In later years, new NS fabrication techniques such as nanoimprinting, catalytic vapor deposition, or solution-based processes have enlarged the zoo of NSs and NS architectures proposed for SC with either enhanced efficiency or the potential for substantial cost reduction.

A common way to characterize NSs is according to the dimensionality: The material without confinement, i.e., featuring extended states in all three spatial dimensions, is referred to as *bulk*. For electronic confinement in one, two, and three spatial dimensions, the resulting 2D, 1D, and 0D NSs are called *quantum well* (QW), *quantum wire* (QWR), and *quantum dot* (QD), respectively. At length scales where only optical confinement is present, the corresponding objects are usually called thin films, nanowires, and nanoparticles, respectively. On the other hand, the functionalities of nanostructure solar cell (NSSC) components can be categorized according to the physical degree of freedom that is to be affected: *Optical* functionalities are used for the purpose of light trapping for absorption enhancement via engineering of the EM field strength inside the absorber; the (opto-)*electronic* functionalities, which are related to the tuning of (opto-)electronic properties such as bandgaps, absorption coefficients, mobilities, and carrier relaxation rates via electronic structure engineering; and *vibrational* functionalities, which affect the dissipation of carrier energy due to coupling to lattice vibrations, for instance, via inhibition of suitable phonon modes. As explained earlier, all of the functionalities emerge as the consequence of the nanoscale dimension of the structures that leads to quantization and confinement of the photonic, electronic, and vibrational states and results in physical properties that can deviate considerably from the bulk behavior.

The main applications of NS in photovoltaics are in the so-called second and third generations of PV devices [1], where NS are used in the implementation of SC concepts aiming at either lower cost through strongly reduced material usage and cheap production technology or at increased energy conversion efficiencies through enhanced spectrum utilization and/or reduced thermalization losses. NS with optical functionality have been instrumental in the realization of thin-film SC (second generation) based on low absorption materials, such as thin-film silicon devices [2] or ultrathin absorbers [3]. Relevant representatives of such NS applications for light-trapping include plasmonic metal nanoparticles [4] and nanostructured metal films [5], dielectric scatterers [6] and diffraction gratings [7], engineering of the optical density of states (DOS) via photonic crystals [8], light focusing antenna effects in nanowires and nanorods [9], as well as spectral conversion via QDs [10]. This wide field of optical NS for PV applications is covered by a number of topical reviews [11–14].

The electronic functionality of NS concerns first of all the tuning of optical transitions and nonradiative charge carrier dynamics in high-efficiency SC architectures for concentrator and space applications (third generation). Since the early nineties, multi-QW (MQW—multiple decoupled QW) and QW superlattice (QWSL—periodic and coupled QW) structures have been implemented in different III-V semiconductor materials and have been investigated for application in single-junction SC [15] and multijunction (MJ) PV devices [16], and, more recently, also for hot-carrier SC (HCSC) [17]. The practical relevance of these structures lies mainly in the adjustability of the absorption edge to the wavelength required for efficiency optimization, in the case where suitable bulk materials are not available. Similar bandgap engineering features are provided by multi-QD (MQD) or QD superlattice (QDSL) structures implemented in III-V semiconductors [18–21] or silicon alloys [22], with the advantage of improved strain relaxation mechanisms. However, while QWSCs based on strain-balanced components have enabled efficiencies comparable to the bulk limit [23], severe recombination losses did not allow thus far a similar achievement in QDSC [24]. On the other hand, QD structures were considered to be particularly useful for the implementation of a number of specific third-generation PV concepts: for intermediate band SC (IBSC) due to the possibility to obtain an energetically separated intermediate state or band from the ground state of the QD [25]; for multiple exciton generation (MEG) via impact ionization (providing increased current) due to enhanced electron–electron interaction as a consequence of wave function localization [26]; and for the extraction of hot carriers in HCSC (providing increased voltage) due to inhibited electron–phonon scattering resulting from the low density and spectral sparsity of electronic and vibrational modes [27]. These effects have

been demonstrated in dedicated experimental setups [28–33], but they remain to be shown in working SC devices.

Besides these highly ordered epitaxial QD structures, there is a whole class of QD materials made as colloids from solution, such as, e.g., PbSe and PbS [34], which are arranged into different SC architectures, ranging from QD-sensitized TiO_2 electrodes to hybrid organic–inorganic devices and closely packed QD solids with carrier-selective contacting schemes [35–38]. While present efficiencies are still low (11.3%), these approaches promise low-cost manufacturing, and there is still the prospect of MEG exploitation as observed in the constituent particles [39].

For the sake of completeness, a word is due here on the case of QWRs. Even though they offer a perfect combination of size tunability in transverse dimensions with extended states in transport direction, there are almost no experimental structures including 1D NS with electronic functionality. The reasons are twofold. On the one hand, it is very challenging to produce wires with diameters small enough such as to exhibit electronic confinement. On the other hand, it is not possible to achieve at the same time optical and electronic confinement: At the dimensions suitable for optical confinement, the absorber is electronically bulk-like, and no guided optical modes can exist in the tiny dimensions required for electronic confinement.

As compared to the optical and electronic features emerging at the nanoscale, the vibrational functionalities of NSs have received the least attention so far. Most potential applications consider exploitation of reduced carrier cooling by engineering the phonon modes in NSs such as QWSL [40] and QDSL [22,27]. The issue of confined phonons is discussed also in the context of fast relaxation of hot photogenerated excitons in isolated QD, one of the detrimental factors suppressing MEG, which was thought to be reduced in QD. Recent experiments and simulations emphasize the influence of softened surface modes enabling efficient multiphonon relaxation [41].

All of the above functionalities might be relevant in a single SC device, for instance a HCSC based on a III-V QWSL structure (similar to [40]) with absorption enhancement induced by plasmonic nanoparticles or a cavity effect. In such a case, deviations from bulk physics need to be considered for photons, electrons, and phonons at the same time. Furthermore, while the single functionality of an NS component (e.g., enhancement of interband absorption due to wave function localization) might be beneficial for a certain aspect of device operation (e.g., photogeneration), it may at the same time have a detrimental impact on another aspect (e.g., recombination and transport). Indeed, in experimental implementations of advanced PV concepts, NSs are often found to introduce electronic and optical losses, e.g., due to high defect density at NS interfaces and parasitic absorption into localized states with insufficient carrier extraction efficiency. In the idealized models considered for the determination of efficiency bounds such as the famous Shockley–Queisser (SQ) limit [42], such losses are usually neglected. This explains to some extent the large discrepancy between theoretical ideal case predictions and experimentally achieved results. However, for a realistic assessment of the technological potential of a given NSSC concept, the detrimental aspects associated with the NS need to be considered and quantified. This underlines the importance of advanced modeling approaches reaching beyond bulk semiconductor physics for this category of PV devices, which will be in the focus of the remainder of this chapter.

Space does not permit a detailed discussion of the modeling approaches for all of the NS used in photovoltaics as outlined above. Since nanophotonic aspects of optoelectronic device modeling are covered elsewhere in this book, NS with purely optical functionality—including the important case of the nanowire SC, where the absorber is at the same time its own light-trapping structure—will not be addressed here. The other important class of NSSC that will not be covered comprises the colloidal QD SC architectures. In these devices, the PV processes are strongly influenced by the surface chemistry, such as, e.g., the properties associated with different ligand molecules, but also by strong carrier localization, large exciton binding energies, and the considerable fluctuations in size and position. As a consequence, mobility is usually limited by thermally activated hopping processes [43], resulting in poor transport properties. The disorder-induced smearing of the band edge has also unfavorable impact on the achievable open-circuit voltages, even at the radiative limit and under the assumption of perfect transport [44]. Overall, the situation is

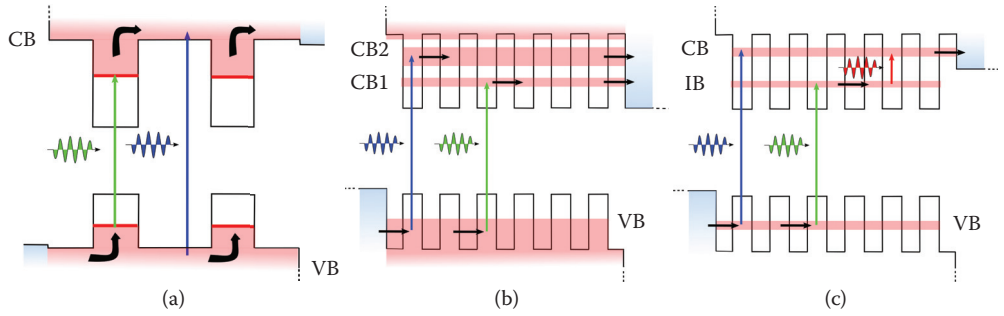

FIGURE 41.1 Common architectures of NSSC: (a) multi-QW or QD structures, where the confined states of distinct NS are not coupled, and photogenerated carriers therefore need first to escape to extended (quasi-)continuum states in order to contribute to current; (b) superlattices of QW or QD, where miniband formation or sequential tunneling enables photocarrier transport and extraction via partially localized NS states; (c) in contrast to (b), the IBSC configuration features noncontacted NS states, which, however, can still be delocalized, but, contrary to (a), should be only radiatively coupled to the contacted continuum states by which carriers are extracted. CB, conduction band; VB, valence band; IB, intermediate band.

closer to the case of organic and hybrid SC, for which a whole range of dedicated simulation approaches exist [45].

The main focus of this chapter, however, will be on the modeling of *regimented* QW and QD architectures as utilized for the implementation of high-efficiency SC concepts using *inorganic* semiconductor materials. Among these structures, three main categories can be distinguished according to the participation of the NS states in the transport of charge carriers. The first category includes devices where the NS states are localized in transport direction and need to be coupled to another set of extended (i.e., bulk-like or unconfined) states to enable charge carrier flow. Typical representatives are the MQW and MQD SC (Figure 41.1a). The second category consists of architectures where both generation-recombination and transport processes are mediated by NS states, which requires a finite degree of wave function delocalization in transport direction related to the coupling of NS, as in QDSL devices with large confinement barriers (Figure 41.1b). Finally, there are structures where coupling-induced delocalization exists between both confined and unconfined states, but only the continuum states are connected to contacts, as in some QD-IBSC implementations (Figure 41.1c).

In Section 41.2, the main physical mechanisms governing the PV device operation are discussed for the different types of NS architectures with regard to their theoretical description and modelization. In Section 41.3, existing methods are reviewed to obtain the device characteristics of NS-based SCs under consideration of the modeling requirements resulting from Section 41.2 and the associated implementation challenges, establishing a simulation hierarchy ranging from global detailed balance approaches for limiting efficiency calculations to microscopic quantum-kinetic theories for the local charge carrier dynamics in nonclassical device regions. In Section 41.4, the simulation of a prototypical NSSC in the form of a single QW photodiode is considered in the light of the approaches discussed in Section 41.3.

41.2 Physical Mechanisms of NSSC Device Operation

In the following, the different stages of the PV energy conversion process in NSSC—from the incident photon flux to the extracted charge current—are considered with focus on the modeling requirements imposed by the presence of NS components.

41.2.1 Photogeneration of Electron–Hole Pairs

The structures that are of interest in NS PV architectures are small compared to the coherence length of sunlight. The spectral *photon flux* $\Phi_\gamma(\mathbf{r}, \hbar\omega)$ at energy $\hbar\omega$, with $\hbar$ the reduced Planck constant and ω the frequency of the light, and position $\mathbf{r}$ inside the absorber of the SC under consideration is therefore well described by the propagation of coherent EM radiation in a dispersive medium, as provided by the solution of (macroscopic) Maxwell's equations. In the most general case, the linear response[†] of the medium to the transverse electrical field $\boldsymbol{\mathcal{E}}^{\mathrm{t}}$ is characterized by the complex microscopic dielectric tensor $\tilde{\boldsymbol{\varepsilon}}(\mathbf{r}, \mathbf{r}', \hbar\omega)$, which considers the electronic excitations in dependence of the energy and polarization of the incident radiation field and of the available electronic states. The corresponding local volume rate $\mathcal{G}$ for the generation of electron–hole pairs by photon absorption is obtained from the dissipated EM energy as follows [46]:

$$\mathcal{G}(\mathbf{r}, \hbar\omega) = \eta_{\mathrm{gen}}(\mathbf{r}, \hbar\omega)(2\hbar\mathcal{V})^{-1} \sum_{\mu,\nu} \int d^3r' \, \Im \left\{ \varepsilon_0 \varepsilon_{\mu\nu}(\mathbf{r}, \mathbf{r}', \hbar\omega) \mathcal{E}^{t*}_{\mu}(\mathbf{r}, \hbar\omega) \mathcal{E}^{t}_{\nu}(\mathbf{r}', \hbar\omega) \right\}, \tag{41.1}$$

where $\Im$ denotes the imaginary part, ε_0 is the vacuum permittivity, η_{gen} is the generation efficiency ($\eta_{\mathrm{gen}} = 1$ means that each photon that is absorbed generates exactly one electron-hole pair), and $\mathcal{V}$ is the absorbing volume. In the above expression, the time-harmonic monochromatic Ansatz $\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r}, t) = \Re\{\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r}, \hbar\omega) \exp(-i\omega t)\}$ was used. While the interaction with the light is inherently nonlocal from the electronic point of view, an averaged local version of the dielectric tensor can be used if the EM field varies only slowly over the length scale of electronic inhomogeneities. For an isotropic medium and unit generation efficiency, Equation 41.1 can then be rewritten as

$$\mathcal{G}(\mathbf{r}, \hbar\omega) = (2\hbar)^{-1} \varepsilon_0 \Im\varepsilon(\mathbf{r}, \hbar\omega) |\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r}, \hbar\omega)|^2 \approx \alpha(\mathbf{r}, \hbar\omega)\Phi(\mathbf{r}, \hbar\omega), \tag{41.2}$$

where

$$\alpha(\mathbf{r}, \hbar\omega) \equiv \frac{\omega}{n_{\mathrm{r}} c_0} \Im\varepsilon(\mathbf{r}, \hbar\omega) \tag{41.3}$$

denotes the (locally defined) *absorption coefficient* that reflects the decay of the optical intensity due to all optical excitations possible at a given photon energy, weighted by their respective strength (c_0: speed of light in vacuum, n_{r}: refractive index of the background).

41.2.1.1 Photon Flux

In the general time-harmonic case, the photon flux is given by the Poynting vector $\mathbf{S}$ of the transverse EM fields ($\boldsymbol{\mathcal{E}}^{\mathrm{t}}, \boldsymbol{\mathcal{B}}^{\mathrm{t}}$):

$$\mathbf{S}(\mathbf{r}, \hbar\omega) = (2\mu_0)^{-1} \left\{ \boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r}, \hbar\omega) \times \boldsymbol{\mathcal{B}}^{t}(\mathbf{r}, \hbar\omega) \right\} \equiv \Phi(\mathbf{r}, \hbar\omega)\hbar\omega \, \hat{\mathbf{n}}(\mathbf{r}, \hbar\omega), \tag{41.4}$$

where μ_0 is the vacuum permeability and $\hat{\mathbf{n}}$ is the unit vector in the direction of propagation. The evolution of the flux inside the absorber is related to the EM dissipation due to light-matter coupling via the optical conservation law:

$$\nabla \cdot \mathbf{S}(\mathbf{r}, \hbar\omega) = -\frac{1}{2}\omega\varepsilon_0 \, \Im \left\{ \boldsymbol{\varepsilon}(\mathbf{r}, \hbar\omega) [\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r}, \hbar\omega)]^* \cdot \boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r}, \hbar\omega) \right\}, \tag{41.5}$$

[†] This is appropriate for standard solar illumination conditions, i.e., the AM1.5g spectrum.

where the transverse electrical field is given by Maxwell's equations acquiring the form

$$\mu_0^{-1}\nabla\times\nabla\times\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r},\hbar\omega)=\omega^2\varepsilon_0\boldsymbol{\varepsilon}(\mathbf{r},\hbar\omega)\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r},\hbar\omega), \tag{41.6}$$

$$\nabla\cdot\varepsilon_0\boldsymbol{\varepsilon}(\mathbf{r},\hbar\omega)\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r},\hbar\omega)=0. \tag{41.7}$$

The *absorptance* of the absorbing volume is then given by the ratio $a\equiv P_{\mathrm{abs}}/P_{\mathrm{in}}$ of absorbed and incident EM power, with $P_{\mathrm{abs}}=\int_{\mathcal{V}}d\mathcal{V}(-\nabla\cdot\mathbf{S})$ and $P_{\mathrm{in}}=\int_{\mathcal{S}}d\boldsymbol{S}\cdot\mathbf{S}$, where $\mathcal{S}$ is the surface of the absorbing volume. In the absence of optical confinement—for instance, in optically thick absorbers—the decay of the flux can be written in terms of the absorption coefficient via the *Lambert–Beer* (LB) law, e.g., for propagation along the z-direction:

$$\Phi_{\mathrm{LB}}(z,\hbar\omega)=\Phi_0^{\swarrow}(\hbar\omega)\exp\left(-\int_0^z dz'\,\alpha(z',\hbar\omega)\right), \tag{41.8}$$

with $\Phi_0^{\swarrow}$ denoting the incident photon flux at $z=0$. This defines the absorptance $a_{\mathrm{LB}}(z,\hbar\omega)\equiv 1-\Phi_{\mathrm{LB}}(z,\hbar\omega)/\Phi_0^{\swarrow}(\hbar\omega)$ after an optical path of length z. For planar systems featuring optical elements such as antireflection coatings (ARC) or a back reflector, as frequently used in thin-film PV devices due to the requirement of absorption enhancement via light trapping, the required solution of Maxwell's equations can be performed using a transfer-matrix method (TMM) [47], which relates the coefficients of incident, reflected, and transmitted plane wave components for piecewise constant dielectric function. In the case of 2-D/3-D nanophotonic or plasmonic light trapping structures, Maxwell's equations (Equation 41.7) have to be solved by advanced numerical approaches such as the finite-difference time-domain (FDTD) method or the finite element method (FEM).

41.2.1.2 Absorption Coefficient

Equations 41.2 and 41.3 can be used to determine the absorption coefficient from the complex dielectric function or the linear *susceptibility* $\chi(\omega)=\varepsilon(\omega)-\varepsilon_b$ ($\varepsilon_b\equiv n_r^2$: background dielectric constant) or, equivalently, from the net absorption rate $r_{\mathrm{abs}}=\mathcal{G}/\eta_{\mathrm{gen}}$ (net means after subtraction of the rate for stimulated emission). For weak perturbation, the net volume rate of uncorrelated direct electronic transitions between states v and c by absorption of photons of energy $E_\gamma=\hbar\omega$ is given by *Fermi's Golden Rule* (FGR):

$$r_{v\to c}^{\mathrm{abs}}(E_\gamma)=\frac{2\pi}{\hbar\mathcal{V}}|\langle c|\hat{H}_{e\gamma}|v\rangle|^2\delta(\varepsilon_c-\varepsilon_v-E_\gamma)(f_v-f_c), \tag{41.9}$$

with $\hat{H}_{e\gamma}$ the electron–photon interaction Hamiltonian, $\varepsilon_{c,v}$ the energies of single electron and hole states, and $f_{c,v}$ the occupation of these states. The electron–photon Hamiltonian has the general form

$$\hat{H}_{e\gamma}(\mathbf{r},t)=\frac{e}{2m_0}\left\{\hat{\mathbf{A}}(\mathbf{r},t)\cdot\hat{\mathbf{p}}+\hat{\mathbf{p}}\cdot\hat{\mathbf{A}}(\mathbf{r},t)\right\}+\frac{e^2[A(\mathbf{r},t)]^2}{2m_0}\approx\frac{e}{m_0}\hat{\mathbf{A}}(\mathbf{r},t)\cdot\hat{\mathbf{p}}, \tag{41.10}$$

where $\hat{\mathbf{A}}$ is the EM vector potential, $\hat{\mathbf{p}}=-i\hbar\nabla$ is the momentum operator, and for the last expression (minimal coupling), the assumption of weak vector potential $\hat{\mathbf{A}}$ was made and Coulomb gauge ($\nabla\cdot\mathbf{A}=0$) was used. In the standard *semiclassical* picture of light-matter interaction in semiconductors, the light is treated as a classical wave. Using for the incident field, a monochromatic time-harmonic plane wave expansion:

$$\mathbf{A}(\mathbf{r},t)=\frac{A_0}{2}\boldsymbol{\epsilon}\left\{e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)}+c.c.\right\} \tag{41.11}$$

of the vector potential, with amplitude $A_0/2$, wave vector $\mathbf{q} = \frac{\omega}{c}\hat{\mathbf{n}}$, and polarization $\boldsymbol{\epsilon}$, together with the minimal coupling form of $H_{e\gamma}$ in the FGR rate (Equation 41.9) yields

$$r^{\text{abs}}_{v\to c}(E_\gamma) = \frac{2\pi}{\hbar V}\frac{e^2A_0^2}{4m_0^2}\mathcal{M}^2_{\text{cv}}\delta(\varepsilon_{cv} - E_\gamma)(f_v - f_c), \tag{41.12}$$

$$\mathcal{M}_{\text{cv}} \equiv |\langle c|e^{i\mathbf{q}\cdot\mathbf{r}}\boldsymbol{\epsilon}\cdot\hat{\mathbf{p}}|v\rangle|, \tag{41.13}$$

where $\varepsilon_{cv} \equiv \varepsilon_c - \varepsilon_v$. If $\mathbf{q}\cdot\mathbf{r} \ll 1$ (i.e., the wavelength of the light is much larger than the characteristic length scale of the absorbing structure), the phase factor $e^{i\mathbf{q}\cdot\mathbf{r}}$ in the matrix element can be neglected, which corresponds to the *dipole approximation*, meaning that contributions of higher order electric and magnetic multipoles are not considered. In this approximation, the coupling via the momentum operator $\hat{\mathbf{p}}$ can be replaced by a coupling via the dipole operator $\hat{\mathbf{d}} = -e\hat{\mathbf{r}}$ ($\hat{\mathbf{r}}$: position operator) due to the relation $(e/m_0)\mathbf{A}\cdot\hat{\mathbf{p}} \approx -\hat{\mathbf{d}}\cdot\boldsymbol{\mathcal{E}}^{\text{t}}$ [48]. The photon flux associated with Equation 41.11 amounts to

$$\Phi_0^{\swarrow}(\hbar\omega) = \frac{n_r c_0 \varepsilon_0 \omega A_0^2}{2\hbar}, \tag{41.14}$$

which, together with Equation 41.12, provides the general FGR absorption coefficient:

$$\alpha^{\text{FGR}}(E_\gamma) = \frac{\pi\hbar}{n_r c_0 \varepsilon_0 E_\gamma \mathcal{V}}\left(\frac{e}{m_0}\right)^2 \sum_{v,c}\mathcal{M}^2_{cv}\delta(\varepsilon_{cv} - E_\gamma)(f_v - f_c), \tag{41.15}$$

where the sum is over all initial and final states.

In Equation 41.15, the occupation of the initial and final states is assumed to be known, and in the case of interband transitions, $f_v = 1$ and $f_c = 0$ are often assumed. However, this is no longer appropriate in the case of transitions between partially filled subbands, as for instance in the case ot the IBSC concept. Thus, in general, the occupation needs to be determined by a microscopic treatment of the charge carrier dynamics under illumination in the framework of the semiconductor Bloch equations [48–50] or nonequilibrium Green's functions [51].

The main quantities determining the generation of electron-hole pairs are, thus, the local value of the transverse EM field on the one hand, and the local value of the dipole or momentum matrix elements (MMEs) and of the density of occupied initial and empty final states of the optical transitions, respectively, on the other hand. While the spatial variation of the transverse EM fields is small over the extent of semiconductor NS in the range of few nanometers, the electronic quantities reflect the impact of symmetry breaking, wave function localization, and energy quantization associated with reduced dimensionality. A prominent example is silicon QDs, where the contribution of direct transitions—absent in bulk—and an increase in oscillator strength due to larger overlap of electron and hole wave functions (in this case equivalent to larger MME) result in a strongly increased absorption [52,53].

Considering Equation 41.12, evaluation of the absorption in semiconductor NSs starts with the computation of the wave functions $\psi_{c,v}$ and energies $\varepsilon_{c,v}$ of the single-particle states $\{|v\rangle, |c\rangle\}$ participating in the optical transitions. There is a vast amount of literature on that topic, which is also covered by a number of chapters in this volume, and will thus not be treated explicitly here. With the electronic structure information, idealized absorption spectra can be obtained from Equation 41.12. However, for a realistic assessment of the PV response, the experimental spectra should be reproduced as closely as possible, which requires consideration of the actual effects that have an impact on the line shape, such as finite fields, excitonic contributions, lifetime broadening due to coupling to phonons and hybridization with continuum states, compositional disorder, and finite-size distributions. Various approaches with increasing degree of sophistication can be found in the literature.

For **QW**, excitons and field effects (i.e., electroabsorption) were included in [54] using a variational ansatz for the exciton wave functions. The consistent consideration of the bound-to-bound state and the bound-to-quasibound state transitions including excitonic and field effects was achieved using an exciton Green's function approach [55]. To describe lineshape broadening beyond phenomenological approaches, microscopic quantum-kinetic theories such as the density matrix formalism, including the different scattering mechanisms (electron–phonon, electron-electron, etc.) can be used [48–50]. Recently, excitonic absorption enhancement for QWs under consideration of the electron–phonon interaction was formulated in a general nonequilibrium Green's function framework [56]. In practice, however, some degree of fitting is used to reproduce the experimental spectra. For QWSC, a semiempirical approach was introduced by Paxman et al. [57], based on earlier work on QW optical modulators [58,59], in which the field-dependent absorption coefficient is expressed as

$$\alpha(E_\gamma, \mathcal{E}^z) = \sum_{c,v} \alpha_{cv} M_{cv}(\mathcal{E}^z) \Big[\sum_l r_{cv}^l \mathcal{L}[E_\gamma, \varepsilon_{cv}(\mathcal{E}^z) - B_{cv}^l] + \int_{\varepsilon_{cv}(\mathcal{E}^z)}^{\infty} dE' \mathcal{L}(E_\gamma, E') \Big], \tag{41.16}$$

where $\mathcal{E}^z$ is the longitudinal electric field in growth direction, $M_{cv} = \int_L dz \psi_c^*(z, \mathcal{E}^z)\psi_v(z, \mathcal{E}^z)$ (L: normalization length) is the overlap matrix element of the field-dependent single-particle electron and hole wave functions, r_{cv}^l (in units of energy) and B_{cv}^l are the relative oscillator strength and binding energy of the lth exciton, respectively, and $\mathcal{L}(E_\gamma, E') = \Gamma/(2\pi[(E_\gamma - E')^2 + \Gamma^2/4])$ is a Lorenzian lineshape function to include homogeneous broadening characterized by Γ. The parameter α_{cv} corresponds to the absorption coefficient at the onset of the continuum absorption, and, in practice, is used together with B and r to fit the experimental spectra. For the absorption at photon energies above the bulk bandgap of the barrier material, the bulk absorption coefficient of the well material is used, neglecting any field effects. A similar semiempirical approach was used in [60], based on [61].

For **QD**, the single-particle absorption spectra can be obtained using similar density matrix formalisms for the linear susceptibility [62] or the equivalent FGR, with straightforward extension from single QD to QDSL [63,64]; in the case of significant excitonic effects, the latter can be treated by means of the configuration interaction (CI) method [65]. While lineshape broadening due to dephasing via electron–phonon interactions can in principle be included in quantum-kinetic approaches, the finite-size distribution of QD ensembles requires additional convolution with some distribution function representing inhomogeneous broadening, commonly of the Gaussian type [66]. In self-assembled QD formed on a wetting layer (WL), the WL states may also contribute to the subgap absorption [67]. For practical device simulation, a common approach is the use of an effective medium absorption coefficient [68]:

$$\bar{\alpha}_{\text{eff}} = w_{\text{QD}}\bar{\alpha}_{\text{QD}} + w_{\text{WL}}\bar{\alpha}_{\text{WL}} + w_{\text{bulk(QD)}}\bar{\alpha}_{\text{bulk(QD)}} + w_{\text{barrier}}\bar{\alpha}_{\text{barrier}}, \tag{41.17}$$

which contains the contribution of the confined states in the QD, the WL, the bulk absorption of the QD material for continuum states, and the barrier absorption, weighted according to the relative volume fractions and a given distribution function reflecting the variations in size and shape. The general expression for the absorption of an ensemble of QD configurations C with distribution f using these simple approaches reads:

$$\bar{\alpha}_{\text{QD}}(E_\gamma) = \int dC f(C) \alpha_C(E_\gamma), \tag{41.18}$$

$$\alpha_C(E_\gamma) = \frac{\pi\hbar}{n_r c_0 \varepsilon_0 E_\gamma \mathcal{V}_{QD}} \left(\frac{e}{m_0}\right)^2 \sum_{c,v} \bar{p}_{cv}^2(C) \frac{\Gamma}{\left(\varepsilon_{cv}(C) - E_\gamma\right)^2 + \Gamma^2/4} (f_v - f_c), \tag{41.19}$$

where $\bar{p}_{cv}$ is the polarization-averaged MME and Γ is the homogeneous broadening. The WL absorption is given by the expression for QW (usually neglecting excitonic absorption), where consideration of thickness fluctuations results in an additional absorption tail. Since the absorbing volume of a QD is not

always well defined, especially in the case of quasibound states, it is common to write the QD contribution absorption in terms of an *absorption cross-section* σ_{QD} per QD and of the QD concentration N_{QD} via $\alpha_{QD} = N_{QD}\sigma_{QD}$.

41.2.2 Charge Separation and Extraction

In general, the extraction of photogenerated charge carriers represents one of the most critical aspects in NS-based PV devices. Indeed, while in some devices based on QW or QD, the tailoring of the absorption via band structure engineering could be successfully demonstrated [69,70], the efficiency of carrier collection has until now remained below the values found for the bulk counterparts, in some cases on a dramatically poor level, especially in devices where transport proceeds via states with increased degree of localization. There are several reasons for the observed performance issue, which are associated with different stages in the extraction process, such as exciton dissociation, carrier escape and capture between localized and extended states, mobility issues associated with scattering, and the virtual absence of true miniband formation in realistic situations, which shall be discussed below.

41.2.2.1 Exciton Dissociation

Since electrons and holes have to be extracted via separate contacts, the dissociation of the photogenerated excitons and the subsequent charge separation need to occur at some point in the device. In bipolar bulk SC, exciton binding energies are usually in the range of a few meV, resulting in thermal dissociation immediately after generation, and the generation can thus safely be described via (Equation 41.2) in terms of noninteracting electron-hole pairs. In isolated NS on the other hand, spatial confinement enhances the electron-hole interaction, and exciton binding energies can amount to multiples of the thermal energy k_BT (k_B: Boltzmann constant, T: temperature), in which case rapid thermal dissociation is no longer possible, and the dissociation process needs to be considered explicitly [53]. Inefficient exciton dissociation can result in a transport behavior dominated by exciton diffusion, similar to the situation in organic SC devices, where a specially designed bulk–heterojunction interface is required for exciton dissociation. As a SC based on exciton diffusion is not likely to reach very high efficiencies, the description of excitons in the simulation of efficient NSSC devices is primarily focused on the excitonic enhancement of optical transitions close to the effective band edge, as described above in the treatment of the absorption coefficient.

41.2.2.2 Charge Carrier Escape and Capture

Charge carriers that are generated in localized states of electronically decoupled NS need to be transferred to extended states in order to contribute to current, which normally amounts to a sequence of scattering events involving confined (2D—QW/0D—QD) subband states followed by scattering from a bound or quasibound state to a (3D) continuum state. While this *escape* process constitutes a loss mechanism in light-emitting devices, which is sought to be suppressed in optimized designs, the opposite applies in NSSC architectures. Hence, understanding the escape process via identification of dominant and limiting mechanisms and design of corresponding models has been an important focus in NSSC research from the beginning. On the other hand, any escape channel enables its inverse capture process, which removes photocarriers from contacted states and increases dark current flow.

In general, while the microscopic processes underlying the carrier dynamics are always related to the fundamental coupling between the electronic, vibrational, and optical degrees of freedom, there are several mechanisms by which photogenerated charge carriers can escape from confined NS states, which are commonly characterized by respective time constants τ^i_{esc}, with total escape rate proportional to $\tau^{-1}_{esc} = \sum_i (\tau^i_{esc})^{-1}$. For **QW** at room temperature—the relevant temperature regime for terrestrial SC operation—and with shallow confinement, the dominant escape mechanism was identified as *thermionic emission* [71], which describes the current due to charge carriers that occupy the part of the energy distribution located above the barrier edge. In the presence of a rectifying built-in field, the escape process is affected by the latter through the modified height (with respect to the center of the QW) and triangular shape of the

barrier, giving rise to escape enhancement due to *tunneling*. For the common assumption of a constant field, the effective barrier for escape in direction of and against the field is given by $V_{B\mp}(\mathcal{E}^z) = \Delta E_b \mp q\mathcal{E}^z L_w/2$, where ΔE_b is the offset of bulk band edges and L_w is the QW width. A general *phenomenological* expression for the corresponding escape time is then given by [71]

$$\tau_{\text{esc}}^{-1} = \frac{1}{L_w}\int_0^\infty \tilde{n}(E)\mathcal{T}(E)v(E)dE, \quad \tilde{n}(E) = \frac{n(E)}{\int_0^\infty n(E)dE}, \tag{41.20}$$

where $\mathcal{T}(E)$ is the probability function for transmission through the barriers, $v(E)$ is the carrier velocity, and $n(E) = g(E)f(E)$ is the spectral carrier density at an energy E measured from the center of the bottom and assuming a given distribution function $f(E)$ for charge carriers in the QW. The actual form of Equation 41.20 depends on the specific description of the spectral quantities ($\mathcal{T}, v, g, f$). The simplest assumption for $\mathcal{T}$ is a step function with unit transmission above the effective barrier edge and zero below, which neglects tunneling contributions. Tunneling can be included on the level of the Wentzel–Kramers–Brillouin (WKB) theory via [72]

$$\mathcal{T}_{\text{WKB}}(E) = \exp\left(-\frac{2}{\hbar}\int_0^{L_B} dz\sqrt{2m_B^*[V_B(z) - E]}\right), \tag{41.21}$$

where m_B^* is the effective mass of carriers in the barrier material, L_B is the width of the barrier, and $V_B(z)$ is the spatial profile of the barrier potential. More accurate approaches for determination of the transmission function are based on analytical or numerical solution of the Schrödinger equation using the exact barrier potential, e.g., by means of an electronic transfer matrix formalism [71]. For free carriers, the velocity is related to the (kinetic) carrier energy via the momentum, e.g., $v_z = \hbar k_z/m_w^* = \sqrt{2E_z/m_w^*}$ for propagation in the z-direction. In basic approximation, the thermal velocity v_{th} is used, corresponding to $E_z = k_{\text{B}}T$. The exact form of the DOS g reflects the transition of dimensionality from 2D to 3D at energies close to the barrier edge and can be captured in advanced descriptions based on scattering states in a real space basis representation, which do not discriminate between bound, quasibound, and continuum states, such as the Green's function approaches [73,74]. In most implementations, however, a combination of 2D square well DOS and 3D bulk DOS is used [71]. Finally, the occupation function f needs to encode the nonequilibrium steady-state carrier distribution under illumination and bias voltage. In general, this is achieved by using Fermi or Boltzmann statistics together with a quasi-Fermi level (QFL) for the carriers in the QW. In the case of Boltzmann statistics, the QFL drops out of the expression for the escape time due to the normalization to the total carrier density [75]. The advantage of the above approach is that it covers any combination of thermionic, direct tunneling, and thermally assisted tunneling escape. The drawback is that it requires assumptions on the carrier distribution in the QW and it does not contain any information on the microscopic mechanisms that lead to the carrier population at a certain energy. Such mechanisms include, e.g., the very fast electron-electron scattering that establishes a hot-carrier quasi-equilibrium at a fs timescale. But even the slower scattering of electrons with longitudinal optical (LO) phonons was shown to produce a hot-carrier population via subband scattering that enables fast thermally assisted carrier escape [76–78]. This latter result, while based on simple application of FGR to electron-LO-phonon scattering between electronic subbands, was confirmed by the state-of-the-art quantum-kinetic simulation of the phonon-assisted tunneling escape [79].

In **QD** devices, in addition to continuum (3D) and confined (0D) QD states, the 2D states of the WL play an important role in the dynamics of carrier transitions. Experimentally, similar thermionic and tunneling escape processes were identified [80–82]. However, thermal escape is much less efficient than in QW, and larger fields are required to enable tunneling escape. In terms of microscopic mechanisms, evidence for both Auger scattering [83] and (multi)phonon absorption [84] was found. While the carrier-carrier scattering depends strongly on carrier density and decreases quickly with QD size due to shrinking Coulomb (Cb)

matrix elements—and is, therefore, less prominent in QW—the phonon processes depend critically on the energetic spacing of the states involved in the transitions, which produces strong resonances in the size dependence of the scattering mechanism. However, the lifetime broadening of QD states due to dynamical processes and delocalization increasing with energy alleviate the impact of the discrete DOS. Finally, in addition to the nonradiative escape processes, QDs offer the possibility of photon-assisted or purely radiative escape via subband transitions as described in Section 41.2.1. In QW, the subband absorption is strongly suppressed for in-plane polarization; hence, special photon management is required to scatter light into large angles, preferentially with coupling to guided modes.

The inverse process to escape is carrier *capture*. In SCs, this process is relevant in MQW and MQD devices, where photogenerated carriers that escaped to the current-carrying continuum states are recaptured into confined states on their way to the contacts. Since on the other hand, this process is instrumental for light-emitting devices, its theoretical description and numerical simulation have received considerable attention and are extensively covered in the literature. For **QW**, traditional models consider the FGR rate for transitions from 3D bulk states to confined states via the emission of LO-phonons [85–95] or carrier-carrier scattering [96,97], in combination with population dynamics from rate equations [98,99] or Monte Carlo simulations [100–104], while more recent approaches are based on quantum-kinetic theory [105]. At the relatively low densities of injected carriers at the operating point of a SC (which is below the bias regime of LED operation), relaxation due to emission of phonons was found to dominate over the effects of carrier-carrier scattering. In the case of **QD** devices, capture from the WL and subband relaxation were evaluated in the FGR framework for Auger processes [106,107], including the effects of line broadening induced by the carrier dynamics [108], and for the emission of single phonons [109,110] and multiple phonons [111], with extension to a polaron picture with finite phonon lifetime [112] indicated due to the strong coupling of QD states to phonons. The mechanisms were also combined in unified theories on semiclassical [113] and quantum-kinetic [114] levels.

41.2.2.3 Transport to and Extraction at Carrier-Selective Contacts

The choice of the picture for charge transport is primarily dictated by the kind of NS device architecture under consideration. In the second and third types of NSSC devices displayed in Figure 41.1, the NS states are involved not only in the absorption, but also in the carrier transport. These states thus need to be to some degree extended, which amounts to the requirement of coupling between individual NS. Indeed, most of the concepts of this kind are based on the formation of minibands in superlattices of QWs or QDs. For that situation, transport may be described as band-like, with a low-field Bloch-type electron *mobility* $\mu = \sigma/(\rho e)$ defined in terms of the conductivity tensor σ, and in the diffusive limit is obtained from the linearized Boltzmann transport equation (BTE) [115]:

$$\mu_{\alpha\beta} = e\left[\sum_{n,\mathbf{K}} \tau_{n\mathbf{K}} v^{\alpha}_{n\mathbf{K}} v^{\beta}_{n\mathbf{K}} \left(-\frac{\partial f^0_{n\mathbf{K}}}{\varepsilon_{n\mathbf{K}}}\right)\right] \Big/ \sum_{n,\mathbf{K}} f^0_{n\mathbf{K}}. \tag{41.22}$$

In Equation 41.22, n is the band index, $\mathbf{K}$ the superlattice wave vector, $\tau_{n\mathbf{K}}$ is the relaxation time, $\varepsilon_{n\mathbf{K}}$ is the miniband dispersion, $f^0_{n\mathbf{K}}$ is the equilibrium Fermi-Dirac electron distribution function, and $v^{\alpha}_{n\mathbf{K}} = \hbar^{-1}\partial_{K_\alpha}\varepsilon_{n\mathbf{K}}$ is the band velocity. The relaxation time depends on the scattering processes that are present in the device and can be computed—in some approximation—from the corresponding FGR formalism [116]:

$$\tau^{-1}_{n\mathbf{K}} = \sum_{n'\mathbf{K}'} \left(\frac{1 - f^0_{n'\mathbf{K}'}}{1 - f^0_{n\mathbf{K}}}\right)\left\{1 - \cos(\theta_{\mathbf{K}\mathbf{K}'})\right\} P^{nn'}_{\mathbf{K}\mathbf{K}'} \tag{41.23}$$

where $\cos(\theta_{\mathbf{KK}'}) = (\mathbf{v}_{n\mathbf{K}} \cdot \mathbf{v}_{n'\mathbf{K}'})/(|\mathbf{v}_{n\mathbf{K}}||\mathbf{v}_{n'\mathbf{K}'}|)$ and $P^{nn'}_{\mathbf{KK}'}$ is the FGR rate for scattering from the occupied state $|n\mathbf{K}\rangle$ to the empty state $|n'\mathbf{K}'\rangle$. However, even in the ideal case of a perfect miniband, the associated Bloch mobility may be critically low, as in the case of silicon QDs in a dielectric matrix material [117]. In realistic situations, due to the presence of built-in fields and any kind of spatial, configurational and compositional disorder, the NS states usually show a high degree of localization such that tunneling between NS is restricted to nearest neighbors. In addition to inducing localization, these effects lead to a misalignment of energy levels in adjacent NS, with the formation of Wannier-Stark ladders in the extreme case of very strong fields [118], in which situation transport is only possible via an inelastic scattering process and is best described by hopping between localized states. In coupled QD systems, due to the sparsity of the DOS, the presence of charge on the NS can lead to pronounced shifts in the energy level structure [53], inhibiting transport (Coulomb blockade regime). In devices where charge transport acquires molecular character, such as in colloidal QD solids, the hopping process involves even a change in configuration and may be phenomenologically described by the Marcus theory [44]. The bulk picture of charge carrier transport in band states does thus in general not provide a proper description of the real situation, which in addition to the remaining coherence and nonlocality effects should include all the localization effects and scattering mechanisms required to overcome energetic misalignment. The modification of the electronic structure due to the presence of NS also affects the scattering processes responsible for the limitation of mobility and for carrier relaxation, such as electron–phonon interaction, which becomes especially relevant in devices where this form of energy dissipation is sought to be suppressed, as in the hot-carrier concept, and in the case where extraction distances are on the order of the mean free path and transport approaches the ballistic regime.

The final step of the PV energy conversion consists in the extraction of the charge carriers at carrier-selective contacts. Carrier selectivity—which is an essential requirement on any PV device architecture [119]—can be described on a phenomenological level via the surface recombination velocities for majority and minority carriers used in the boundary conditions for carrier density/current in macroscopic transport equations. Microscopically, transport across the boundary to the contact depends again on the electronic structure of the interface and on the prevailing scattering processes. Since those can be largely engineered using NS, the latter also play an important role in the implementation of energy-selective [22] and carrier-selective [120] contacts.

41.2.3 Recombination

On the way to or from the contacts, the interaction of optically or electronically injected charge carriers with the EM and vibrational fields can induce different types of *recombination* (the mutual annihilation of an electron and a hole), which are usually classified in *radiative*, *Auger*, and *defect-mediated* processes. Like for the generation process, a proper consideration of NS effects on the recombination processes requires a careful consideration of the wave functions and local DOS together with the occupation of these states in general nonequilibrium conditions.

41.2.3.1 Radiative

Even in ideal, defect-free absorber materials, the principle of *detailed balance* dictates the presence of radiative recombination by spontaneous and stimulated photon emission as the inverse process to carrier generation by photon absorption. Conventionally, the local volume rate of radiative recombination is related to the charge carrier densities and the local optical material constants (α, n_r) as follows [121]:

$$r^{\mathrm{em}}(\mathbf{r}) = \mathcal{B}(\mathbf{r})\left\{\rho_e(\mathbf{r})\rho_h(\mathbf{r}) - \rho_e^0(\mathbf{r})\rho_h^0(\mathbf{r})\right\}, \tag{41.24}$$

$$\mathcal{B}(\mathbf{r}) = \rho_i^{-2}\int dE_\gamma\, \alpha(\mathbf{r}, E_\gamma)\bar{\Phi}_{bb}(E_\gamma), \tag{41.25}$$

where ρ^0 and ρ_i denote equilibrium and intrinsic carrier densities, respectively, and $\bar{\Phi}_{bb} = 4\pi\Phi_{bb}$ is the angle-integrated blackbody radiation flux, with

$$\Phi_{bb}(E_\gamma) = \frac{E_\gamma^2 n_r^2}{4\pi^3\hbar^3 c_0^2}\left\{\exp\left(\frac{E_\gamma}{k_B T}\right) - 1\right\}^{-1}. \tag{41.26}$$

In Equation 41.25, the assumption of emission in an optically homogeneous medium is made, with isotropic photon DOS. In thin-film SCs with nanophotonic light trapping, this is often not appropriate, and the radiative recombination should be modeled starting from the FGR for spontaneous emission under consideration of the actual density of final photon states. For NS, the emission is enhanced as compared to the bulk, due to carrier localization [122] resulting in maximized density product in Equation 41.25 and to larger overlap of electron and hole wave functions in Equation 41.13. A further point to note in the case of NSSC is the dependence of the absorption coefficient on the operating conditions, mainly the bias voltage via the effect of the built-in field on the local electronic structure, which should be reflected in the quantities used in Equation 41.25.

41.2.3.2 Auger

In the regime of large optical injection, relevant for the operation under high optical concentration, the PV performance of a defect-free absorber is limited by the interband Auger recombination mechanism [123], which as an intrinsic effect is unavoidable and needs thus to be considered for a realistic estimate of the limiting energy conversion efficiency. In this process originating in carrier-carrier scattering, electrons recombine with holes by giving up the excess energy to excite either another electron or another hole. In the semiclassical bulk continuum formulation, this rate takes the form [124]

$$r^{\mathrm{Aug}}(\mathbf{r}) = \left[C_e(\mathbf{r})\rho_e(\mathbf{r}) + C_h(\mathbf{r})\rho_h(\mathbf{r})\right]\left\{\rho_e(\mathbf{r})\rho_h(\mathbf{r}) - \rho_e^0(\mathbf{r})\rho_h^0(\mathbf{r})\right\}, \tag{41.27}$$

where $C_{e/h}$ are the Auger coefficients for electron/hole excitation. For NS, some or all of the band indices are replaced by the labels of subbands (QW) or discrete levels (QD). The Auger coefficients can be computed microscopically by application of perturbation theory for the carrier-carrier interaction under consideration of the appropriate dimensionality of the (screened) Cb potential. For QW, FGR-type approaches were used to assess direct [125–129] and phonon-assisted [130] Auger processes, and the latter was also investigated using a Green's function technique [131]. For QD, similar FGR models were developed for colloidal QD [132] and applied to self-assembled III-V single QD [133] and QD arrays [134]. Recently, the complete dynamics of excitons in colloidal QD was also simulated on an *ab initio* level using a combination of density functional theory and molecular dynamics [135]. While in QW, the magnitude of the Auger recombination rate was found to be similar to bulk, it is strongly enhanced in QD due to the effects of wave function localization and associated relaxation of momentum selection rules.

41.2.3.3 Defect Mediated

While radiative and Auger losses are unavoidable and thus present even in structures without imperfections, recombination in real-world devices is usually dominated by defect-mediated processes where the carrier energy is dissipated as heat and is lost for the conversion process. Indeed, there is an inherent increase in surface area associated with the presence of NS, which are thus likely to act as centers of nonradiative recombination. In some cases, it is possible to reduce the NS-related defects via specially engineered strain-balancing techniques [136] or via passivation [137], but in general, the insertion of NS leads to losses in the open-circuit voltage of the SC. With regard to the modeling, the role of lower dimensional NS in the context of defect-mediated recombination has received little attention so far, and models are usually

based on the conventional bulk Shockley–Read–Hall (SRH) theory [138,139], with modification to relate the recombination to interface or surface defects rather than volume defects, e.g., for QW [60]

$$r^{\mathrm{SRH}}(\mathbf{r}) = \frac{\rho_e(\mathbf{r})\rho_h(\mathbf{r}) - \rho_e^0(\mathbf{r})\rho_h^0(\mathbf{r})}{(\rho_e(\mathbf{r}) + \rho_e^t(\mathbf{r}))\tau_e + (\rho_h(\mathbf{r}) + \rho_h^t(\mathbf{r}))\tau_h}, \tag{41.28}$$

where $\tau_s^{-1} = \sigma_s v_{th} N_S / L_w$ $(s = e, h)$ with σ_s the carrier capture cross-section, v_{th} the thermal velocity, N_S the sheet density of interface defects, L_w the QW width, and ρ^t the carrier density for a Fermi level at the trap energy. The capture cross-sections depend on the local electronic structure of the NS and defect states, as well as on the phonon modes coupling to those states in the multiphonon relaxation mechanism [124], and can be obtained based on microscopic theories on phenomenological [140] and fully *ab initio* [141] levels.

41.3 Modeling of NSSC Device Characteristics

The basic function of a SC consists in the conversion of the solar energy flux impinging on its surface into electrical power. A central requirement on any theory suitable for the modeling of SC device operation is thus the ability to predict the PV energy conversion efficiency as defined via the ratio $\eta_{\mathrm{PV}} \equiv P_{el}^{\nearrow} / P_{\gamma}^{\swarrow}$ of generated electrical power to the incident radiative power, with

$$P_{el}^{\nearrow} = \max_V \{I(V) \cdot V\}|_{T_C}, \tag{41.29}$$

$$P_{\gamma}^{\swarrow} = \int dE_\gamma \int_S \int_\Omega d\boldsymbol{\Omega} \cdot d\boldsymbol{S}\, \Phi_{\gamma}^{\swarrow}(\mathbf{r}_S, \Omega, E_\gamma) E_\gamma, \tag{41.30}$$

where V is the terminal voltage (i.e., related to the separation of chemical potentials at the contacts), $I(V)$ is the charge current extracted from the device at this voltage, T_C is the temperature of the SC, E_γ is the photon energy, and $\Phi_{\gamma}^{\swarrow}$ denotes the incident spectral photon flux impinging on a cell of (outer) surface S under a solid angle of incidence $\Omega = (\theta, \varphi)$. Hence, any PV model needs to provide I as a function of V and $\Phi_{\gamma}^{\swarrow}$ at given T_C.

As outlined in Sections 41.1 and 41.2, depending on the functionality of the NS within the SC device, some or all of the fundamental PV processes are affected by the peculiar physical properties of the NS component, with major implications for the requirements on a suitable simulation approach. The general approach reflects the multiscale and multiphysics scheme depicted in Figure 41.2: Starting from the NS states, the dynamics giving the occupation of these states is evaluated and the resulting generation and recombination rates are used to obtain the device characteristics. Three levels of sophistication that can be identified in the hierarchy of theoretical descriptions of SC device operation providing the current–voltage characteristics $I(V, \Phi_{\gamma}^{\swarrow})$, depending on the explicity of consideration of the various physical mechanisms, are: *detailed balance* theories providing efficiency limits of idealized systems, *semiclassical* transport models with detailed balance generation-recombination rates, and *quantum-kinetic* theories with consistent and fully microscopic treatment of charge carrier dynamics and transport [142]. These three approaches shall now be discussed in some more detail.

41.3.1 Global Detailed Balance Theories for Ideal Systems

On the most basic level, NS properties are considered only in generation and recombination at the radiative limit, while transport is assumed to be ideal. This means that from a device point of view, the system is treated as a black box, and there is no information on geometry or any spatial inhomogeneity in the electronic system that would depart from effective bulk properties. The most prominent examples of this

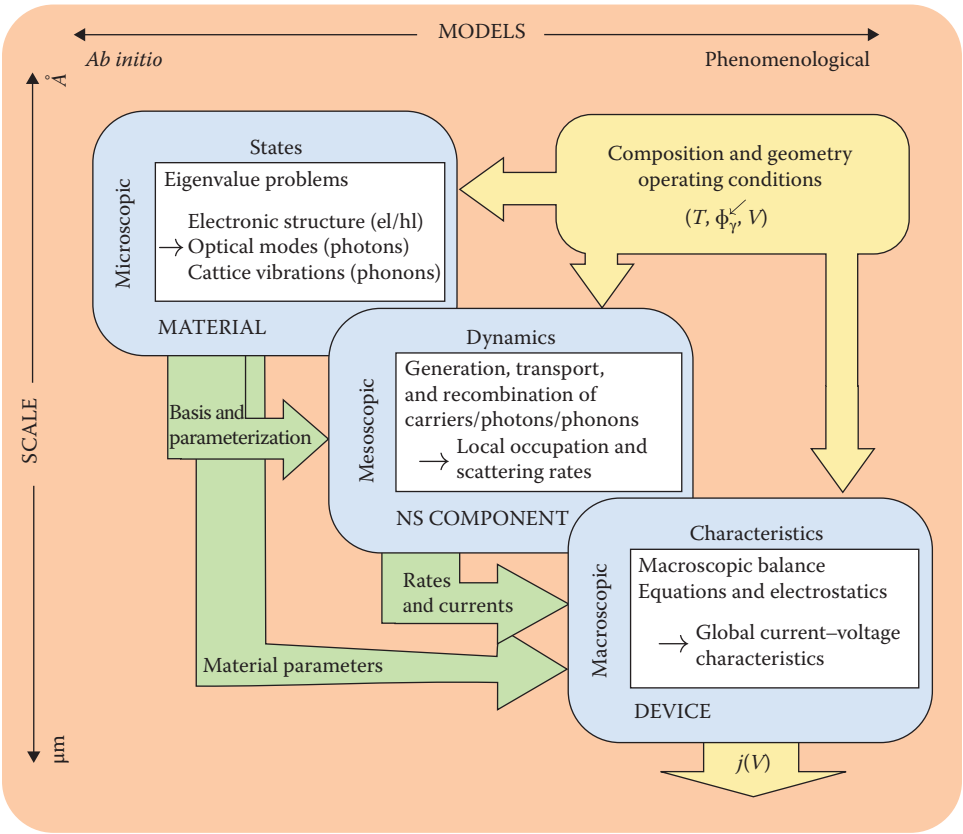

FIGURE 41.2 Schematic representation of the multiscale and multiphysics approach to the simulation of nanostructure-based solar cell devices.

category are the detailed balance models [143] used to obtain an upper bound to the energy conversion efficiency, such as the SQ limit. In this approach, the current–voltage characteristics are obtained by equating the terminal current with the difference of the number of photons per unit time entering and leaving the absorber volume at given terminal bias voltage [143,144]:

$$I(V,\Phi_\gamma^{\swarrow})|_{T_C} = -q\int \mathrm{d}E_\gamma \int_{\mathcal{S}}\int_{\Omega} \mathrm{d}\boldsymbol{\Omega}\cdot\mathrm{d}\boldsymbol{\mathcal{S}}\left\{\Phi_{\mathrm{abs}}(\mathbf{r}_{\mathcal{S}},\Omega,E_\gamma)-\Phi_{\mathrm{em}}(\mathbf{r}_{\mathcal{S}},\Omega,E_\gamma)\right\}|_{V,\Phi_\gamma^{\swarrow},T_C}. \tag{41.31}$$

At the radiative limit, where the recombination is determined by the emission, and assuming unit generation efficiency, the electronic equivalent to the above balance relation is obtained from the integration of the charge continuity equation over the absorber volume $\mathcal{V}$ (e/h: electrons/holes):

$$I(V,\Phi_\gamma^{\swarrow})|_{T_C} = \mp q\int_{\mathcal{V}} d^3\mathbf{r}\left\{\mathcal{G}_{e/h}(\mathbf{r})-\mathcal{R}_{e/h}(\mathbf{r})\right\}|_{V,\Phi_\gamma^{\swarrow},T_C}. \tag{41.32}$$

The absorbed photon flux is derived in terms of the absorptance—or, synonymously, the absorptivity—a of the device, and of the reflection coefficient r for incident light, via

$$\Phi_{\mathrm{abs}}(\mathbf{r}_{\mathcal{S}},\Omega,E_\gamma) = [1-r(\mathbf{r}_{\mathcal{S}},\Omega,E_\gamma)]a(\mathbf{r}_{\mathcal{S}},\Omega',E_\gamma)\left\{\Phi_\gamma^{\swarrow}(\mathbf{r}_{\mathcal{S}},\Omega,E_\gamma)+\Phi_{\gamma 0}(E_\gamma,\Omega,T_C)\right\}, \tag{41.33}$$

where Ω' is the solid angle after refraction, related to the former by Fresnel's law [145]. The second term in the curly brackets is the isotropic thermal equilibrium flux and is given by

$$\Phi_{\gamma 0}(E_\gamma, \Omega, T_C) = \Phi_{bb}(E_\gamma, T_C)\cos\theta, \tag{41.34}$$

where Φ_{bb} is the blackbody radiation flux (Equation 41.26) introduced in Section 41.2.3. The generation rate in terms of local photon flux and absorption coefficient is given in Equation 41.2. However, for equivalence with the above result, photon recycling needs to be considered in the evaluation of the local photon flux. At unit generation efficiency and in the absence of nonradiative recombination, the product in front of the incident flux in Equation 41.33 is identified as the *external quantum efficiency* (*EQE*). Under the assumption of perfect transport, i.e., infinite mobility, which amounts to flat QFLs split by the terminal voltage, i.e., $\Delta\mu = qV$, and of an optically isotropic medium, detailed balance then provides the emitted photon flux as a function of the absorptance [144,146]:

$$\Phi_{\mathrm{em}}(\mathbf{r}_S, \Omega, E_\gamma)|_V = [1 - r(\mathbf{r}_S, \Omega, E_\gamma)]a(\mathbf{r}_S, \Omega', E_\gamma)\tilde{\Phi}_{bb}(\Omega, E_\gamma, T_C, V) \tag{41.35}$$

$$\equiv EQE(\mathbf{r}_S, \Omega, E_\gamma)\tilde{\Phi}_{bb}(\Omega, E_\gamma, V), \tag{41.36}$$

where:

$$\tilde{\Phi}_{bb}(\Omega, E_\gamma, T_C, V) = \Phi_{bb}(E_\gamma - qV, T_C)\cos\theta \tag{41.37}$$

$$\approx \Phi_{\gamma 0}(E_\gamma, \Omega, T_C)\mathrm{e}^{qV/(k_B T_C)} \qquad (E_\gamma \gg qV). \tag{41.38}$$

The approximation in Equation 41.38 permits a compact formulation of the current–voltage characteristics:

$$I(V, \Phi_\gamma^{\swarrow})|_{T_C} \approx -q\int dE_\gamma \int_S \int_\Omega d\mathbf{\Omega}\cdot d\boldsymbol{S}\, EQE(\mathbf{r}_S, \Omega, E_\gamma) \times \left\{\Phi_\gamma^{\swarrow}(\mathbf{r}_S, \Omega, E_\gamma) - \Phi_{\gamma 0}(E_\gamma, \Omega, T_C)\left[\mathrm{e}^{qV/(k_B T_C)} - 1\right]\right\}. \tag{41.39}$$

Finally, the local recombination rate for the determination of the radiative dark current via Equation 41.32 is given by Equation 41.25.

In the SQ limit, the absorptance is a unit step function with threshold energy at the value of the bandgap E_g, which is then the only electronic parameter entering the calculation of the current–voltage characteristics. This simplifies the expressions for the absorbed and emitted photon fluxes considerably. In a first step beyond the SQ analysis, the effect of incomplete absorption with non–abrupt onset as well as the impact of light-trapping structures can be considered by using in Equations 41.33 and 41.35 the proper absorptance computed by optical models for the light propagation as outlined in Section 41.2. It is common to neglect the bias dependence of the photocurrent also in situations where a realistic absorptance is considered, which results in current–voltage characteristics given by the *superposition* of photocurrent I_γ and dark current I_d:

$$I(V, \Phi_\gamma^{\swarrow}) = I_\gamma(\Phi_\gamma^{\swarrow}) + I_\mathrm{d}(V). \tag{41.40}$$

However, this tends to be inappropriate in nanoscale absorbers due to the strong impact of built-in fields and finite state filling on the local absorption coefficient [147,148].

Due to its formal simplicity, the detailed balance analysis of the radiative limit is applied to most novel NS PV devices in a first assessment of the performance potential. On the other hand, the restrictive assumptions made are seldom met in realistic devices, and the upper efficiency bound obtained in this way may therefore be far from a realistic estimate.

Within the above setting of idealized conditions, an alternative starting point for the device characteristics is a modification of the ideal diode equation [149]:

$$I(V, \Phi_\gamma^{\swarrow}) = I_0\left(e^{qV/(k_B T_C)} - 1\right) - I_G(\Phi_\gamma^{\swarrow}) + I_R(V), \tag{41.41}$$

which is formally equivalent to Equation 41.39, but where the generation and recombination currents I_G and I_R are composed of the contributions of the bulk "baseline" cell and of the NS components. The NS contributions are related to those of the baseline using correction factors for quantities such as geometry (fraction of NS material), oscillator strength, and DOS (for radiative transitions). It needs to be noted, however, that such modifications still have to comply with the detailed balance relations between absorption and emission as given above. Furthermore, even in its original formulation devised for the radiative limit, the theory goes beyond that limit in considering in the reverse saturation current I_0 the standard nonradiative minority carrier diffusion contribution of an ideal *p*-(*i*-)*n* diode.

41.3.1.1 Application to QWSC

Soon after the proposal of the QWSC concept, the detailed balance approach was used to show that under the idealized conditions of that theory, the QWSC could not provide an efficiency in excess of that of an ideal gap bulk cell, since both photocurrent generation and radiative recombination are based on the same modified absorptivity [150,151]. An SQ-type detailed balance model was then used to demonstrate that QWSC efficiency could still exceed the bulk limit if variations of the quasi-Fermi level splitting (QFLS) between bulk and QW material were allowed for [152]. This was shown to be equivalent to the detailed balance picture of the intermediate band concept, with the difficulty of the QW featuring suppressed subband absorption and fast population equilibration due the continuous electronic DOS [153]. Indeed, the issue of QFLS variation in QWSC is still a matter of debate and ongoing research. In more recent QWSC developments, the detailed balance approach was used to model tandem configurations and the impact of directional emission in strained QW [23]. Following the original work on the ideal diode QWSC model [149], a number of variations and refinements of this semianalytical approach were devised for MQWSC [154–161] and QWSLSC [159,162–164].

41.3.1.2 Application to QDSC

For QD intermediate band SC (QDIBSC), the original detailed balance IBSC model [165] was adapted to account for the specific absorption characteristics of QD structures [166,167] and was implemented for specific material systems using a Kronig-Penney model for supracrystals of InAsN/GaAsSb QD [168], InGaN/GaN QD [169], and InGaN/InN [170], as well as a $\mathbf{k} \cdot \mathbf{p}$ model for isolated InAs/GaAs QD [171]. The ideal diode theory was applied to theoretically investigate an InAs/GaAs MQD concept with AlGaAs fences to prevent capture of photogenerated carriers [172].

41.3.2 Hybrid Models: Semiclassical Transport with Detailed Balance Rates

The main shortcoming of global detailed balance approaches and ideal diode theories is the lack of a realistic consideration of photocarrier transport, i.e., the assumption of unit collection efficiency, which is often inappropriate in NSSC devices. In order to account for the effects of finite mobility, actual transport equations have to be solved for the charge carriers. The main difficulty in that undertaking resides in the consideration of the contribution of NS states with a higher degree of localization. To obtain the self-consistent occupation of confined NS states (*c*) and bulk host states (*b*), separate but coupled rate equations

need to be formulated for the densities of carriers occupying different types of states. Under the standard assumption of continuum (i.e., no atomistic information) and complete thermalization (i.e., occupation is characterized by a local QFL), the current–voltage characteristics are obtained from the *drift-diffusion* (DD) charge current for carrier species $s \in \{e, h\}$:

$$\mathbf{J}_{s,i}(\mathbf{r}) = -q\left\{\mu_{s,i}(\mathbf{r})\rho_{s,i}(\mathbf{r})\nabla\phi(\mathbf{r}) \mp D_{s,i}(\mathbf{r})\nabla\rho_{s,i}(\mathbf{r})\right\}, \qquad i = b, c \tag{41.42}$$

where μ denotes the mobility, $\rho_{e/h}$ denotes the charge carrier densities, D is the diffusion coefficient, ϕ is the electrostatic potential, q is the elementary charge, and the upper (lower) sign is for electrons (holes). In this general formulation, no restriction is made on the charge transport in NS states, which allows for application to all three NS architectures introduced in Section 41.1. The charge carrier densities and the electrostatic potential that determine the charge currents in Equation 41.42 are obtained from the coupled solution of the steady-state continuity equations:

$$\mp\frac{1}{q}\nabla\cdot\mathbf{J}_{s,i}(\mathbf{r}) = \sum_{j\neq i}\left\{\mathcal{G}_{s,j\to i}(\mathbf{r}) - \mathcal{R}_{s,i\to j}(\mathbf{r})\right\}, \tag{41.43}$$

where $\mathcal{G}/\mathcal{R}$ is the carrier (volume) generation/recombination rate due to transitions between (sub)bands and the upper (lower) sign applies again to electrons (holes), and Poisson's equation for the total charge density:

$$\epsilon_0\nabla\cdot\left\{\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})\right\} = -q\left\{\sum_i \rho_{\mathrm{h,i}}(\mathbf{r}) - \sum_j \rho_{\mathrm{e,j}}(\mathbf{r}) + N_{\mathrm{dop}}(\mathbf{r})\right\}, \tag{41.44}$$

where ε is the static dielectric function and $q \cdot N_{\mathrm{dop}}$ is the net charge density due to ionized dopants. The boundary conditions used in the solution of the coupled equations are the same as in the bulk case, reflecting the nature of the physical contact in terms of (Schottky-)barriers and surface recombination. In Equation 41.43, the generation rate $\mathcal{G}_{s,j\to i} = \mathcal{G}_{s,j\to i}[\Phi_\gamma(\mathbf{r}), \alpha_{j\to i}(\mathbf{r}, E_\gamma)]$ $(i = b, c)$ is a functional of the local photon flux and the local absorption coefficient as given by Equation 41.2 and the recombination rate $\mathcal{R}_{s,i\to j} = \mathcal{R}_{s,i\to j}[\rho_{e,i}, \rho_{h,j}, \ldots]$ depends on the carrier densities and on recombination mechanism (radiative, Auger, defect-mediated)–specific parameters. In the case where, due to strong localization as for isolated QW or QD absorbers, direct transport between NS states can be neglected, the current term in Equation 41.43 is absent for these states, and exchange of carriers between localized states always proceeds via scattering to and from extended states. The coupling of the equations for localized and extended states is then provided on the level of a pure rate equation, where the rates are subject to the detailed balance condition and are commonly expressed via the density in the initial state and an associated lifetime depending on the scattering process, e.g.,

$$\mathcal{R}^x_{e,i\to j} = \frac{\rho_{e,i}}{\tau^x_{e,i\to j}}, \qquad \text{x: scattering mechanism.} \tag{41.45}$$

The lifetimes τ (which are local quantities) are either used as fitting parameters to reproduce experimental characteristics, or derived from an appropriate description for the microscopic mechanisms of the scattering process responsible for the coupling as reviewed in Section 41.2, e.g., via FGR, based on the solution of the Schrödinger-Poisson problem for the NS states and energies. For consistency, the macroscopic material parameters based on microscopic information, such as absorption coefficients and mobilities used in the generation term and the current expressions, should be computed on the basis of the same solutions of the microscopic equations for the electronic structure. It needs to be emphasized at this point that the lifetimes themselves still depend on the occupation of the states, e.g., for the radiative recombination, $\tau_e^{\mathrm{em}}(\mathbf{r}) \approx \left[\mathcal{B}(\mathbf{r})\rho_h(\mathbf{r})\right]^{-1}$, where the very small equilibrium term in (Equation 41.24) is neglected.

In the above formulation of the charge current, the local electronic structure does not appear as a parameter that could be determined microscopically. If the Einstein relation between mobility and diffusion coefficient, i.e., $D = \mu k_B T/q$, and the Boltzmann approximation for the carrier distribution function are valid, an equivalent formulation of the charge carrier densities and currents in terms of effective DOS $\mathcal{N}$ and QFL E_F can be used:

$$\rho_{s,i}(\mathbf{r}) = \int dE\, \mathcal{D}_{s,i}(\mathbf{r}, E) f_{s,i}(E) \approx \mathcal{N}_{s,i} \exp\left(\frac{\pm E_{F_{s,i}}(\mathbf{r}) \mp \varepsilon_{s,i}(\mathbf{r})}{k_B T}\right), \tag{41.46}$$

$$\mathcal{N}_{e,i} \equiv \int_{\varepsilon_{e,i}}^{\infty} dE\, \mathcal{D}_{e,i}(\mathbf{r}, E) \exp\left(\frac{\varepsilon_{e,i}(\mathbf{r}) - E}{k_B T}\right), \tag{41.47}$$

$$\mathcal{N}_{h,i} \equiv \int_{-\infty}^{\varepsilon_{h,i}} dE\, \mathcal{D}_{h,i}(\mathbf{r}, E) \exp\left(\frac{E - \varepsilon_{h,i}(\mathbf{r})}{k_B T}\right), \tag{41.48}$$

$$\mathbf{J}_{s,i}(\mathbf{r}) = \pm\, \mu_{s,i}(\mathbf{r}) \rho_{s,i}(\mathbf{r}) \nabla E_{F_{s,i}}(\mathbf{r}), \tag{41.49}$$

with upper (lower) sign applying to electrons (holes) and $\mathcal{D}$ denoting the DOS for carrier species s and band i, with $\varepsilon_{s,i}$ the minimum (maximum) energy associated ("band edge"). This approach, in addition to the implicit consideration of the microscopic DOS for NS states, provides the possibility to study the emergence of distinct QFL for the NS populations, which is instrumental in PV concepts such as the IBSC.

The *hybrid* (macroscopic-microscopic) approach outlined above can be used to reproduce experimental device characteristics with remarkable accuracy. However, being a macroscopic and local model, any situation requiring energy resolution, nonlocality, or coherence cannot be described properly, e.g., nonthermalized carrier distributions or resonant tunneling. To include such processes in a consistent description, transition to a truly microscopic picture of carrier transport is indicated.

41.3.2.1 Applications to QWSC

Semiclassical DD Poisson models with detailed balance rates for generation and recombination were applied to III-V MQWSC [60,173,174] and to Si-SiO_x QWSLSC [175]. Simplifications include rate equations with semianalytical solutions for current contributions [176,177] and the analytical solution of diffusion equation for excess minority carriers in neutral layers [57,154,178].

41.3.2.2 Applications to QDSC

The above mentioned DD Poisson approach with local rates was implemented in a comprehensive fashion for MQDSC [68,179,180] and QD-IBSC [181–183]. Again, there are also approximations in the form of analytical solutions of the diffusion equation [184–186], of simple rate equations for the population of bulk and NS states [187], or a combination of the two [188].

41.3.3 Fully Microscopic Quantum-Kinetic Models

In situations where the NSs are also involved in transport and the latter is not entirely band-like or purely diffusive, a consistent microscopic picture of the PV processes can be formulated by means of nonequilibrium quantum statistical mechanics using the nonequilibrium Green's function formalism (NEGF) [189]. In the NEGF picture, the device current density $\mathbf{J} = \mathbf{J}_n + \mathbf{J}_p$ is given in terms of the charge carrier Green's function components as follows:

$$\mathbf{J}_{n/p}(\mathbf{r}) = \pm \lim_{\mathbf{r}' \to \mathbf{r}} \frac{e\hbar}{m_0} \left(\nabla_{\mathbf{r}} - \nabla_{\mathbf{r}'}\right) \int \frac{dE}{2\pi} G^{<}(\mathbf{r}, \mathbf{r}'; E). \tag{41.50}$$

The Green's functions in Equation 41.50 are determined by the steady-state Keldysh equation:

$$G^{\lessgtr}(\mathbf{r},\mathbf{r}';E) = \int d\mathbf{r}_1 \int d\mathbf{r}_2 G^R(\mathbf{r},\mathbf{r}_1;E)\Sigma^{\lessgtr}(\mathbf{r}_1,\mathbf{r}_2;E)G^A(\mathbf{r}_2,\mathbf{r}';E), \tag{41.51}$$

with retarded and advanced components resulting from the Dyson equation:

$$\int d\mathbf{r}_1 \left[\left\{G_0^{R/A}\right\}^{-1}(\mathbf{r},\mathbf{r}_1;E) - \Sigma^{R/A}(\mathbf{r},\mathbf{r}_1;E)\right] G^{R/A}(\mathbf{r}_1,\mathbf{r}';E) = \delta(\mathbf{r}-\mathbf{r}'). \tag{41.52}$$

While the retarded and advanced Green's functions $G^{R/A}$ are related to the charge carrier DOS, the correlation functions $G^{\lessgtr}$ additionally contain information on the (nonequilibrium) occupation of these states. The noninteracting system is described by G_0. The self-energies Σ, on the other hand, are scattering functions describing the renormalization of the Green's functions due to coupling to the environment in the form of interactions with photons (mediating photogeneration and radiative recombination), phonons (mediating relaxation and indirect transitions), and other carriers (relevant for excitonic and Auger processes). An additional self-energy term describes injection and extraction of carriers as contacts with arbitrary chemical potential, enabling the treatment of an open nonequilibrium system. The computation of the NEGF is self-consistently coupled to the Poisson equation (Equation 41.44) for the mean-field electrostatic potential ϕ—entering the equation for G_0—via the charge carrier density:

$$\rho_{e/h}(\mathbf{r}) = \mp i \int \frac{dE}{2\pi} G^{\lessgtr}(\mathbf{r},\mathbf{r};E). \tag{41.53}$$

For an in-depth review of the general NEGF formalism and its application to NS PV, the reader is referred to [189] and references therein.

In contrast to the semiclassical models presented in Section 41.3.2, the microscopic approach provides insight into the physical mechanisms underlying the PV device operation, as those are properly captured by the formalism. On the downside, the method is extremely demanding from a computational point of view and applications are therefore limited to functional NSSC components of mesoscopic extension. An important direction of future research will thus consist in the combination of quantum-kinetic models for local carrier dynamics in NS regions with semiclassical models for the global characteristics of the extended device.

41.3.3.1 Applications to Nanowire SCs

As discussed in Section 41.1, QWRs are not in the focus of efficient solar energy harvesting systems. However, the low dimensionality allows for an efficient numerical treatment of the system. Indeed, the earliest application of NEGF for PV energy conversion considered quasi-1D carbon nanotube systems with tight-binding (TB) band structure [190–192]. More recently, a $\mathbf{k}\cdot\mathbf{p}$-NEGF approach was formulated to study photocurrents in GaAs nanowires [193].

41.3.3.2 Applications to QWSC

Single and double QW structures in III-V material systems were studied in [194,195] using a simple TB band structure and in [79,196] within an effective mass approach, with focus on the modification of absorption and emission spectra by the complex potential profile and on the escape of carriers from confined to extended, current-carrying states. Photocarrier transport regimes from ballistic extraction to phonon-mediated sequential tunneling in SL absorbers were investigated for Si-SiO_x [197] (including phonon-mediated indirect transitions) and InAlGaAs-InGaAs [198].

41.3.3.3 Applications to QDSC

The main challenge for the implementation of NEGF models for QDSC is the complexity of the electronic structure with respect to 3D spatial variation, which prevents the use of the slab models conveniently applied to planar NSs. A simple approach is to use a coarse-grained model with a basis of localized QD orbitals—e.g., the Wannier functions of a regular QDSL—in the spirit of the TB approximation in atomic systems. This method was used to formulate a NEGF model for Si QDSL absorbers embedded in the SiO_x-SiC matrix material [199], and to study the impact of inter-QD and QD-contact coupling on the PV performance of generic QD arrays [200]. Within such an approach, however, it is not possible to study on a microscopic level, similar to that achieved in MQWSC, the carrier escape and capture processes that are instrumental in the operation of MQDSC.

41.4 Case Study: Single Quantum Well p-i-n SC

In this section, the different elements and levels of sophistication in the simulation of NSSC shall be illustrated by consideration of a prototypical device architecture, for which a single quantum well (SQW) p-i-n photodiode is chosen. The device as displayed in Figure 41.3 consists of a 100-nm thin bulk GaAs baseline SC with heavily doped emitter and base components ($N_A = N_D = 10^{18}$ cm^{-3}) of 20-nm extension. In the center of the intrinsic region, a slab of 10-nm InGaAs is inserted, which forms a type-I QW. For the band offsets, $\Delta E_{C/V} = 0.15/0.10$ eV is chosen. The contacts are designed to be perfectly carrier selective, which corresponds to the assumption of ohmic boundary conditions for majorities and infinite barriers for minorities. The light is assumed to be incident from the p-side, whereas a gold reflector is applied to the *n*-side, which, however, is not considered in the electronic simulation.

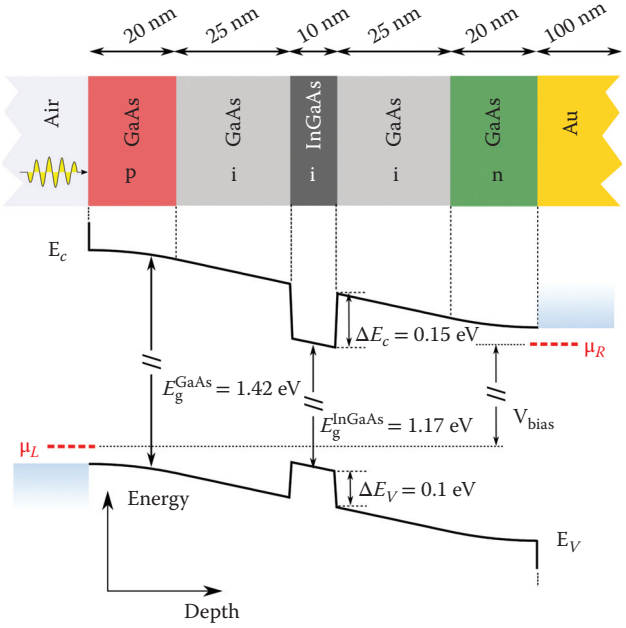

FIGURE 41.3 Single GaAs-InGaAs quantum well p-i-n solar cell architecture used as a test case for the simulation. The contacts are designed to be perfectly selective, i.e., with ohmic boundary conditions for majorities and infinite barriers for minorities. The gold reflector is used only in the optical simulation.

41.4.1 States

In the first step of the simulation, the electronic and optical device states need to be determined. For the description of the electronic structure of the component semiconductor bulk materials, a basic combination of two decoupled single-band effective mass approximation (EMA) models is used, with $m_{e/h}^{*\mathrm{GaAs}} = 0.067/0.1\ m_0$ and $m_{e/h}^{*\mathrm{InGaAs}} = 0.055/0.22\ m_0$. Under the standard thin-film assumption of a transverse continuum due to very large spatial extension perpendicular to the growth direction, the wave functions (WF) can be written as the decomposition

$$\Psi_{n\alpha\mathbf{k}_\parallel}(\mathbf{r}) = \frac{e^{i\mathbf{k}_\parallel\mathbf{r}_\parallel}}{\sqrt{\mathcal{A}}}\psi_{n\alpha\mathbf{k}_\parallel}(z)u_{n0}(\mathbf{r}), \tag{41.54}$$

where n is the bulk Bloch band index, $||$ denotes transverse quantities, $\mathcal{A}$ is the transverse cross-section area, ψ is the envelope function in the growth direction, and u_{n0} is the periodic part of the Bloch function at the Γ point. The 1D EMA Schrödinger equation for the envelope functions of the QW states is solved numerically for closed-system (Dirichlet) boundary conditions and at vanishing field. The wave functions $\psi_{c/v}(z)$ (z: depth) and eigenenergies $\varepsilon_{c/v}$ of the resulting NS states at $\mathbf{k}_\parallel = 0$ are displayed in Figure 41.4a. There are two electron and three hole states with energies below the band edge of the host material, exhibiting the expected symmetry properties and increasing barrier penetration as a function of energy. For comparison, the local density of states (LDOS) at $\mathbf{k}_\parallel = 0$ is computed using the NEGF formalism for identical EMA Hamiltonian and at flat potential (Figure 41.4b), but with open boundary conditions, which, however, does not affect either shape or energies of the confined states. While the solutions of the square well potential are readily available, the situation in the device at operating conditions is quite different, as demonstrated by Figure 41.4c, where the LDOS is shown for a realistic band profile at an operating forward bias voltage

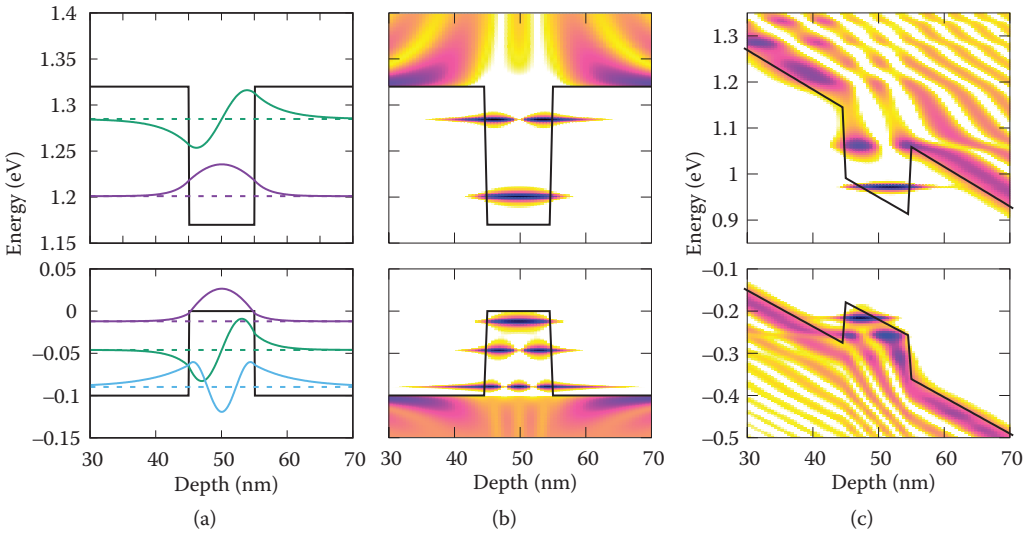

FIGURE 41.4 (a) Wave functions and energies of the eigenstates in the 10-nm InGaAs QW embedded in a GaAs host as obtained by the solution of the 1D single band effective mass Schrödinger equation along the growth direction. (b) One dimensional 1D LDOS ($\mathbf{k}_\parallel = 0$) obtained using the NEGF formalism with the same effective mass Hamiltonian and at flat band conditions. (c) 1D LDOS of the same SQW, but in the p-i-n device configuration close to operating conditions at an applied forward bias voltage $V = 0.8$ V. The most remarkable feature is the unbinding of the higher lying QW states.

of $V = 0.8$ V. Most remarkably, the higher lying QW states lose their confinement in the large built-in field and turn into quasibound resonances.

The optical modes of the device are computed under consideration of the absorbing components and of the reflector using an in-house developed wave propagation solver based on the TMM. As will be shown, this is more appropriate than simple consideration of LB law, since both slab and reflector form an optical cavity, which can lead to nontrivial spatial variation of the optical intensity inside the absorber. Since the states of the quasicontinuum in the QW material are not well described by the chosen approach with closed boundaries, the optical response will be computed only for energies below the host gap. The determination of the optical absorption strength based on the computed electronic states thus forms the next step in the simulation procedure.

41.4.2 Dynamics

The absorption coefficient is evaluated following the FGR procedure outlined in Section 41.2, using the states computed in Section 41.4.1. For energies below the host gap, assuming fully occupied valence bands and empty conduction bands, and neglecting the variation of states and energies with transverse momentum, the absorption coefficient of a square well (SW) potential of width L_w reads

$$\alpha_{\mathrm{SW}}(E_\gamma) = f \cdot E_\gamma^{-1} \sum_{i,j} |M_{c_i v_j}|^2 \Theta\left(E_\gamma - \varepsilon_{c_i v_j}\right), \tag{41.55}$$

with Θ the unit step function and

$$f = \frac{e^2 m_r^* P_{cv}^2}{m_0^2 \hbar n_r c_0 \varepsilon_0 L_w}, \tag{41.56}$$

where m_r^* is the reduced effective mass, $P_{cv} = \sqrt{E_P m_0/6}$ is the bulk MME for Kane energy $E_P = 26.9$ eV, $n_r = 3.7$ is the refractive index of the bulk material, and c_0 and ε_0 are speed of light and permittivity in vacuum, respectively. The overlap matrix elements are

$$M_{c_i v_j} = \int_L dz\, \psi_{c_i}^*(z) \psi_{v_j}(z), \tag{41.57}$$

where L ($> L_w$) is the normalization length of the wave functions, introduce the optical selection rules according to the WF symmetry. As can be verified in Figure 41.5, displaying the above FGR absorption coefficient (full line), only two of the six transitions between the three hole and two electron levels are not suppressed by symmetry. The validity of the selection rules in the absorption coefficient computed via NEGF (dotted line), while evident, is a nontrivial result, which requires careful consideration of the nonlocality in the Green's function picture of electron-photon interaction. In the absence of a symmetric potential, which is not found in the real operating device, the selection rules are strongly relaxed, and the absorption coefficient exhibits a shape that interpolates between 2D and 3D DOS (dashed line).

Based on this absorption coefficient, the emission coefficient $\mathcal{B}$ of the radiative recombination rate introduced in Section 41.2 is determined using the Van Roosbroeck–Shockley (VRS) relation (Equation 41.25). These two quantities are the main ingredients for the assessment of the device characteristics at the radiative limit.

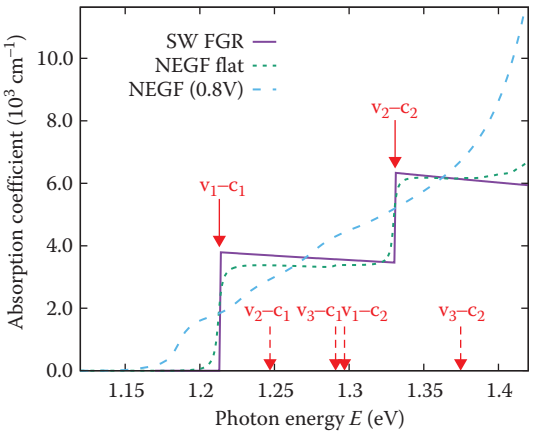

FIGURE 41.5 Absorption coefficient as computed within the Fermi's golden rule approach based on the solution of the single-band effective mass Schrödinger equation for the square well potential (SW FGR, full line), together with the NEGF absorption coefficients (spatially averaged over the QW region) for flat band (dotted) and the real potential at $V = 0.8$ V (dashed). While the selection rules are properly reflected in the NEGF approach, the real spectra feature subband tails and deviate from purely 2-D character due to transitions between quasibound states as a consequence of the large built-in fields present in the intrinsic region of the p-i-n diode.

41.4.3 Characteristics

First, the current–voltage characteristics are evaluated on the level of a detailed balance analysis, i.e., photocurrent and radiative dark current are computed from the absorptance as explained in Section 41.3. To this end, the absorptance is determined both from the LB law (a_{LB}) and from TMM (a_{TMM}) using the previously computed absorption coefficient α_{SW}. The photocurrent density can be evaluated at normal incidence. Dark current, on the other hand, should in principle be evaluated under consideration of the finite opening angle of the loss cone, as determined by the condition of total internal reflection. In the case of emission from high to low refractive index medium, which applies in the situation under consideration here ($n_r^{\mathrm{GaAs}} = 3.6$ to $n_r^{\mathrm{air}} = 1$), the angular dependence may be dropped, i.e., $a(\Omega, E_\gamma) \equiv a(\theta, E_\gamma) \approx a(E_\gamma)$, where the first equality is due to the planar geometry. The angular integration in Equation 41.39 thus provides just a factor of π, and the current–voltage characteristics can be written as

$$J(V) = J_{SC} - J_{0r}\left(e^{qV/(k_B T)} - 1\right), \tag{41.58}$$

with

$$J_{SC} = -q \int \mathrm{d}E_\gamma a(E_\gamma)\Phi_\gamma^{\swarrow}(E_\gamma), \tag{41.59}$$

$$J_{0r} = -q \int \mathrm{d}E_\gamma a(E_\gamma)\pi\Phi_{bb}(E_\gamma). \tag{41.60}$$

For simplicity, a monochromatic incident photon flux at $E_\gamma = 1.3$ eV and intensity of $I_\gamma \equiv \Phi_\gamma^{\swarrow}(E_\gamma)E_\gamma = 0.1$ kW/cm^2 is used for the evaluation of the characteristics. Figure 41.6 displays the JV curves obtained in this way using the absorptances a_{LB} and a_{TMM} based on α_{SW}, which shows that in this unoptimized situation, the reflector does not increase the absorptance of the QW significantly. Comparison with a 1D numerical DD-Poisson-TMM solver (*ASA*-TU Delft) using the same absorption coefficient α_{SW} and corresponding VRS emission coefficient, but otherwise the transport properties of bulk GaAs ($\mu_{n/p} = 400/40$ cm^2/Vs, $\mathcal{N}_{C/V} = 4.35/7.93 \times 10^{17}$ cm^{-3}, $\varepsilon_r = 13.18$, $S_{\mathrm{min/maj}} = 0/10^8$ m/s), reveals—as to be expected—coincidence of the photocurrent with that from a_{TMM}, but also a significant underestimation

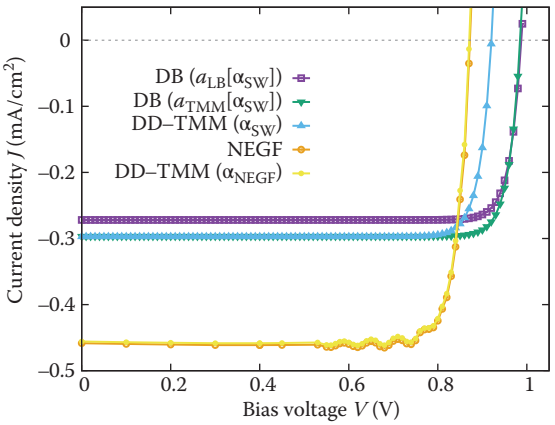

FIGURE 41.6 Current–voltage characteristics as provided by different simulation approaches: detailed balance (DB) currents based on the absorptance from Lambert–Beer's (LB) law or the TMM method, both using the square well potential (SW) absorption coefficient; the semiclassical drift-diffusion-Poisson-TMM model (DD-TMM) for the SW absorption coefficient; the fully microscopic NEGF formalism, revealing a bias-dependent structure in the level of photogeneration. The NEGF characteristics can be reproduced by using in the DD-TMM model the optical data as computed by NEGF, which hints at close to ideal transport in terms of carrier escape from the QW.

of the recombination due to the neglectance of the proper loss cone. The insensitivity to transport is a consequence of the high mobility of GaAs in combination with the short extraction distance and the absence of fast recombination, which results in the flat Fermi levels assumed in the global detailed balance approach. In the full NEGF simulation, on the other hand, both photocurrent and dark current are increased as compared to the characteristics based on the SW absorptance. Remarkably, the NEGF characteristics can be recovered almost completely in the DD-TMM model if local absorption and emission coefficients are extracted from NEGF. This hints at almost perfect transport in terms of unit carrier escape probability (implicitly assumed in the semiclassical model), in consistency with experimental observation for QW structures of similar depth and operated at room temperature. Deviations between the results of the two models are expected in the case of deep wells with presence of fast recombination processes competing with the slow carrier escape: While the NEGF picture considers that a fraction of the photogenerated carriers recombines in the QW prior to escape to the continuum, the semiclassical bulk model assumes the whole generation in bulk states, in which the carriers diffuse quickly from the highly absorbing and—due to detailed balance—emitting QW locations to regions with lower optical coupling strength. That regime, however, is currently not accessible due to the lack of a suitable NEGF picture of nonradiative recombination.

41.5 Summary

The use of NSs in photovoltaics is manifold, and many applications require dedicated models to assess their impact on the PV device performance. There is, however, a common rationale in the consideration of NS in PV device models, according to the device architecture and the NS functionality, which in general amounts to the application of a multiscale and multiphysics scheme. Hence, NS device simulation generally involves many different levels of modeling and simulation, from the determination of NS states to their consideration in the dynamics of carrier generation, transport, and recombination and the final synthesis in the propagation of NS properties to features in the global device characteristics. Reliable consideration of NS effects hence demands a derivation of the key PV material parameters (α, μ, τ) based on a consistent microscopic picture. In the case of ideal transport, an assessment in terms of the optical properties and detailed balance relations between absorption and emission gives already a good estimate of the device

performance. For less ideal systems, the individual loss mechanisms need to be considered in more detail, as provided by semiclassical models on device level, down to the microscopic picture of the fundamental physical mechanisms as delivered by quantum-kinetic approaches in the case where nonclassical effects are relevant, such as transport away from diffusive and band-like regimes.

References

1. M. A. Green. Third generation photovoltaics: Ultra-high conversion efficiency at low cost. *Prog. Photovolt. Res. Appl.*, 9:123, 2001.
2. J. Müller, B. Rech, J. Springer, and M. Vanecek. TCO and light trapping in silicon thin film solar cells. *Sol. Energy*, 77(6):917–930, 2004.
3. S. Liu, D. Ding, S. R. Johnson, and Y.-H. Zhang. Approaching single-junction theoretical limit using ultra-thin GaAs solar cells with optimal optical designs. In *38th IEEE Photovoltaic Specialists Conference (PVSC)*, pages 002082–002087, 2012.
4. S. Pillai, K. R. Catchpole, T. Trupke, and M. A. Green. Surface plasmon enhanced silicon solar cells. *J. Appl. Phys.*, 101(9):093105, 2007.
5. I. Massiot, C. Colin, N. Péré-Laperne, P. Roca i Cabarrocas, C. Sauvan, P. Lalanne, J.-L. Pelouard, and S. Collin. Nanopatterned front contact for broadband absorption in ultra-thin amorphous silicon solar cells. *Appl. Phys. Lett.*, 101(16):163901, 2012.
6. J. R. Nagel and M. A. Scarpulla. Enhanced absorption in optically thin solar cells by scattering from embedded dielectric nanoparticles. *Opt. Express*, 18(S2):A139–A146, 2010.
7. C. Heine and R. H. Morf. Submicrometer gratings for solar energy applications. *Appl. Opt.*, 34(14):2476–2482, 1995.
8. P. Bermel, C. Luo, L. Zeng, L. C. Kimerling, and J. D. Joannopoulos. Improving thin-film crystalline silicon solar cell efficiencies with photonic crystals. *Opt. Express*, 15(25):16986–17000, 2007.
9. L. Cao, P. Fan, A. P. Vasudev, J. S. White, Z. Yu, W. Cai, J. A. Schuller, S. Fan, and M. L. Brongersma. Semiconductor nanowire optical antenna solar absorbers. *Nano Lett.*, 10(2):439–445, 2010.
10. K. W. J. Barnham, I. I. Ballard, J. G. Connolly, B. G. Kluftinger N. Ekins-Daukes, J. Nelson, C. Rohr, and M. Mazzer. Recent results on quantum well solar cells. *J. Mater. Sci. Mater. Electron.*, 11:531, 2000.
11. K. R. Catchpole and A. Polman. Plasmonic solar cells. *Opt. Express*, 16(26):21793–21800, 2008.
12. H. A. Atwater and A. Polman. Plasmonics for improved photovoltaic devices. *Nat. Mater.*, 9(3):205–213, 2010.
13. K. R. Catchpole, S. Mokkapati, F. Beck, E.-C. Wang, A. McKinley, A. Basch, and J. Lee. Plasmonics and nanophotonics for photovoltaics. *MRS Bull.*, 36:461–467, 2011.
14. S. Mokkapati and K. R. Catchpole. Nanophotonic light trapping in solar cells. *J. Appl. Phys.*, 112(10):101101, 2012.
15. K. W. J. Barnham and G. Duggan. A new approach to high-efficiency multi-band-gap solar cells. *J. Appl. Phys.*, 67:3490–3493, 1990.
16. D. B. Bushnell, K. W. J. Barnham, J. P. Connolly, M. Mazzer, N. J. Ekins-Daukes, J. S. Roberts, G. Hill, R. Airey, and L. Nasi. Effect of barrier composition and well number on the dark current of quantum well solar cells. In *Proceeding of 3rd World Conference on Photovoltaic Energy Conversion*, Osaka, Japan, page 2709, 2003.
17. L. C. Hirst, M. Fuhrer, D. J. Farrell, A. Le Bris, J.-F. Guillemoles, M. J. Y. Tayebjee, R. Clady, T. W. Schmidt, M. Sugiyama, Y. Wang, H. Fujii, and N. J. Ekins-Daukes. InGaAs/GaAsP quantum wells for hot carrier solar cells. *Proc. SPIE.*, 8256:82560X, 2012.
18. R. Oshima, A. Takata, and Y. Okada. Strain-compensated InAs/GaNAs quantum dots for use in high-efficiency solar cells. *Appl. Phys. Lett.*, 93(8):083111, 2008.
19. S. M. Hubbard, C. D. Cress, C. G. Bailey, R. P. Raffaelle, S. G. Bailey, and D. M. Wilt. Effect of strain compensation on quantum dot enhanced GaAs solar cells. *Appl. Phys. Lett.*, 92(12):123512, 2008.

20. Y. Okada, R. Oshima, and A. Takata. Characteristics of InAs/GaNAs strain-compensated quantum dot solar cell. *J. Appl. Phys.*, 106(2):024306, 2009.
21. C. G. Bailey, D. V. Forbes, R. P. Raffaelle, and S. M. Hubbard. Near 1 V open circuit voltage InAs/GaAs quantum dot solar cells. *Appl. Phys. Lett.*, 98(16):163105, 2011.
22. G. Conibeer, M. Green, E.-C. Cho, D. König, Y.-H Cho, T. Fangsuwannarak, G. Scardera, E. Pink, Y. Huang, T. Puzzer, S. Huang, D. Song, C. Flynn, S. Park, X. Hao, and D. Mansfield. Silicon quantum dot nanostructures for tandem photovoltaic cells. *Thin Solid Films*, 516(20):6748–6756, 2008.
23. J. G. J. Adams, B. C. Browne, I. M. Ballard, J. P. Connolly, N. L. A. Chan, A. Ioannides, W. Elder, P. N. Stavrinou, K. W. J. Barnham, and N. J. Ekins-Daukes. Recent results for single-junction and tandem quantum well solar cells. *Prog. Photovolt. Res. Appl.*, 19(7):865–877, 2011.
24. D. Alonso-Álvarez, A. G. Taboada, J. M. Ripalda, B. Alen, Y. González, L. González, J. M. García, F. Briones, A. Martí, A. Luque, A. M. Sánchez, and S. I. Molina. Carrier recombination effects in strain compensated quantum dot stacks embedded in solar cells. *Appl. Phys. Lett.*, 93(12):123114, 2008.
25. A. Martí, L. Cuadra, and A. Luque. Partial filling of a quantum dot intermediate band for solar cells. *IEEE Trans. Electron Dev.*, 48(10):2394–2399, 2001.
26. A. J. Nozik. Quantum dot solar cells. *Physica E*, 14(1–2):PII S1386–9477(02)00374–0, 2002.
27. R. Patterson, M. Kirkengen, B. Puthen Veettil, D. König, M. A. Green, and G. Conibeer. Phonon lifetimes in model quantum dot superlattice systems with applications to the hot carrier solar cell. *Sol. Energy Mater. Sol. Cells*, 94(11):1931–1935, 2010.
28. V. I. Klimov. Detailed-balance power conversion limits of nanocrystal-quantum-dot solar cells in the presence of carrier multiplication. *Appl. Phys. Lett.*, 89(12):123118, 2006.
29. O. E. Semonin, J. M. Luther, S. Choi, H.-Y. Chen, J. Gao, A. J. Nozik, and M. C. Beard. Peak external photocurrent quantum efficiency exceeding 100% via MEG in a quantum dot solar cell. *Science*, 334(6062):1530–1533, 2011.
30. A. Martí, E. Antolín, C. R. Stanley, C. D. Farmer, N. López, P. Diaz, E. Cánovas, P. G. Linares, and A. Luque. Production of photocurrent due to intermediate-to-conduction-band transitions: A demonstration of a key operating principle of the intermediate-band solar cell. *Phys. Rev. Lett.*, 97(24):247701, 2006.
31. P. G. Linares, A. Martí, E. Antolín, and A. Luque. III-V compound semiconductor screening for implementing quantum dot intermediate band solar cells. *J. Appl. Phys.*, 109(1):014313, 2011.
32. L. C. Hirst, R. J. Walters, M. F. Fuehrer, and N. J. Ekins-Daukes. Experimental demonstration of hot-carrier photo-current in an InGaAs quantum well solar cell. *Appl. Phys. Lett.*, 104(23):231115, 2014.
33. J. A. R. Dimmock, S. Day, M. Kauer, K. Smith, and J. Heffernan. Demonstration of a hot-carrier photovoltaic cell. *Prog. Photovolt. Res. Appl.*, 22(2):151–160, 2014.
34. F. W. Wise. Lead salt quantum dots: The limit of strong quantum confinement. *Acc. Chem. Res.*, 33(11):773–780, 2000.
35. A. J. Nozik, M. C. Beard, J. M. Luther, M. Law, R. J. Ellingson, and J. C. Johnson. Semiconductor quantum dots and quantum dot arrays and applications of multiple exciton generation to third-generation photovoltaic solar cells. *Chem. Rev.*, 110(11):6873–6890, 2010.
36. Y. Kanai, Z. Wu, and J. C. Grossman. Charge separation in nanoscale photovoltaic materials: Recent insights from first-principles electronic structure theory. *J. Mater. Chem.*, 20(6):1053, 2010.
37. A. Ruland, C. Schulz-Drost, V. Sgobba, and D. M. Guldi. Enhancing photocurrent efficiencies by resonance energy transfer in CdTe quantum dot multilayers: Towards rainbow solar cells. *Adv. Mater.*, 23(39):4573–7, 2011.
38. I. J. Kramer and E. H. Sargent. The architecture of colloidal quantum dot solar cells: Materials to devices. *Chem. Rev.*, 114(1):863–882, 2014.
39. R. J. Ellingson, M. C. Beard, J. C. Johnson, P. Yu, O. I. Micic, A. J. Nozik, A. Shabaev, and A. L. Efros. Highly efficient multiple exciton generation in colloidal PbSe and PbS quantum dots. *Nano Lett.*, 5(5):865–871, 2005.

40. D. König, Y. Takeda, and B. Puthen-Veettil. Technology-compatible hot carrier solar cell with energy selective hot carrier absorber and carrier-selective contacts. *Appl. Phys. Lett.*, 101(15), 2012.
41. D. Bozyigit, N. Yazdani, M. Yarema, O. Yarema, W. M. M. Lin, S. Volk, K. Vuttivorakulchai, M. Luisier, F. Juranyi, and V. Wood. Soft surfaces of nanomaterials enable strong phonon interactions. *Nature*, 531(7596):618–622, 2016.
42. W. Shockley and H. J. Queisser. Detailed balance limit of efficiency of p-n junction solar cells. *J. Appl. Phys.*, 32(3):510–519, 1961.
43. R. E. Chandler, A. J. Houtepen, J. Nelson, and D. Vanmaekelbergh. Electron transport in quantum dot solids: Monte Carlo simulations of the effects of shell filling, Coulomb repulsions, and site disorder. *Phys. Rev. B*, 75(8):085325, 2007.
44. P. Guyot-Sionnest. Electrical transport in colloidal quantum dot films. *J. Phys. Chem. Lett.*, 3(9):1169–1175, 2012.
45. D. Beljonne and J. Cornil, editors. *Multiscale Modelling of Organic and Hybrid Photovoltaics*. Berlin: Springer, Berlin, Heidelberg, 2014.
46. J. D Jackson. *Classical Electrodynamics*. Hoboken, NJ: John Wiley & Sons, 1998.
47. P. Yeh, A. Yariv, and C. S. Hong. Electromagnetic propagation in periodic stratified media.1. General theory. *J. Opt. Soc. Am.*, 67(4):423–438, 1977.
48. W. Schäfer and M. Wegener. *Semiconductor Optics and Transport Phenomena*. Berlin: Springer, 2002.
49. H. Haug and A. P. Jauho. *Quantum Kinetics in Transport and Optics of Semiconductors*. Berlin: Springer, 1996.
50. H. Haug and S. W. Koch. *Quantum Theory of the Optical and Electronic Properties of Semiconductors*. Singapore: World Scientific, 2004.
51. U. Aeberhard. Quantum-kinetic theory of photocurrent generation via direct and phonon-mediated optical transitions. *Phys. Rev. B*, 84(3):035454, 2011.
52. G. Allan and C. Delerue. Efficient intraband optical transitions in Si nanocrystals. *Phys. Rev. B*, 66(23):233303, 2002.
53. J.-W. Luo, P. Stradins, and A. Zunger. Matrix-embedded silicon quantum dots for photovoltaic applications: A theoretical study of critical factors. *Energy Environ. Sci.*, 4(7):2546, 2011.
54. D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus. Electric field dependence of optical absorption near the band gap of quantum-well structures. *Phys. Rev. B*, 32:1043–1060, 1985.
55. S.-L. Chuang, S. Schmitt-Rink, D. A. B. Miller, and D. S. Chemla. Exciton Green's-function approach to optical absorption in a quantum well with an applied electric field. *Phys. Rev. B*, 43:1500–1509, 1991.
56. U. Aeberhard. Nonequilibrium Green's function theory of coherent excitonic effects in the photocurrent response of semiconductor nanostructures. *Phys. Rev. B*, 86:115317, 2012.
57. M. Paxman, J. Nelson, B. Braun, J. Connolly, K. W. J. Barnham, C. T. Foxon, and J. S. Roberts. Modeling the spectral response of the quantum well solar cell. *J. Appl. Phys.*, 74:614, 1993.
58. P. J. Stevens, M. Whitehead, G. Parry, and K. Woodbridge. Computer modeling of the electric field dependent absorption spectrum of multiple quantum well material. *IEEE J. Quant. Electron.*, 24:2007, 1988.
59. D. Atkinson, G. Parry, and E. J. Austin. Modelling of electroabsorption in coupled quantum wells with applications to voltage optical modulation. *Semicond. Sci. Technol.*, 5:516, 1990.
60. S. M. Ramey and R. Khoie. Modeling of multiple-quantum-well solar cells including capture, escape, and recombination of photoexcited carriers in quantum wells. *IEEE Trans. Electron Dev.*, 50:1179, 2003.
61. G. Lengyel, K. W. Jelley, and R. W. H. Engelmann. A semi-empirical model for electroabsorption in GaAs/AlGaAs multiple quantum well modulator structures. *IEEE J. Quant. Electron.*, 26(2):296–304, 1990.

62. S. Schmitt-Rink, D. A. B. Miller, and D. S. Chemla. Theory of the linear and nonlinear optical properties of semiconductor microcrystallites. *Phys. Rev. B*, 35:8113–8125, 1987.
63. D. L. Nika, E. P. Pokatilov, Q. Shao, and A. A. Balandin. Charge-carrier states and light absorption in ordered quantum dot superlattices. *Phys. Rev. B*, 76(12):125417, 2007.
64. S. Tomić, T. S. Jones, and N. M. Harrison. Absorption characteristics of a quantum dot array induced intermediate band: Implications for solar cell design. *Appl. Phys. Lett.*, 93(26):263105, 2008.
65. Y. Z. Hu, M. Lindberg, and S. W. Koch. Theory of optically excited intrinsic semiconductor quantum dots. *Phys. Rev. B*, 42(3):1713–1723, 1990.
66. W.-Y. Wu, J. N. Schulman, T. Y. Hsu, and U. Efron. Effect of size nonuniformity on the absorption spectrum of a semiconductor quantum dot system. *Appl. Phys. Lett.*, 51(10):710–712, 1987.
67. A. Vasanelli, R. Ferreira, and G. Bastard. Continuous absorption background and decoherence in quantum dots. *Phys. Rev. Lett.*, 89:216804, 2002.
68. A. W. Walker, O. Thériault, J. F. Wheeldon, and K. Hinzer. The effects of absorption and recombination on quantum dot multijunction solar cell efficiency. *IEEE J. Photovolt.*, 3(3):1118–1124, 2013.
69. R. Riölver, B. Berghoff, D. Bätzner, B. Spangenberg, H. Kurz, M. Schmidt, and B. Stegemann. Si/SiO2 multiple quantum wells for all silicon tandem cells: Conductivity and photocurrent measurements. *Thin Solid Films*, 516(20):6763–6766, 2008.
70. M. Zacharias, J. Heitmann, R. Scholz, U. Kahler, M. Schmidt, and J. Bläsing. Size-controlled highly luminescent silicon nanocrystals: A SiO/SiO2 superlattice approach. *Appl. Phys. Lett.*, 80(4):661–663, 2002.
71. J. Nelson, M. Paxman, K. W. J. Barnham, J. S. Roberts, and C. Button. Steady state carrier escape rates from single quantum wells. *IEEE J. Quant. Electron.*, 29:1460–1468, 1993.
72. O. Y. Raisky, W. B. Wang, R. R. Alfano, C. L. Reynolds Jr., and V. Swaminathan. Investigation of photoluminescence and photocurrent in InGaAsP/InP strained multiple quantum well heterostructures. *J. Appl. Phys*., 81:339, 1997.
73. K. R. Lefebvre and A. F. M. Anwar. Electron and hole escape times in single quantum wells. *J. Appl. Phys.*, 80(6):3595–3597, 1996.
74. K. R. Lefebvre and A. F. M. Anwar. Electron escape time from single quantum wells. *IEEE J. Quant. Electron*., 33(2):187–191, 1997.
75. H. Schneider and K. V. Klitzing. Thermionic emission and Gaussian transport of holes in a GaAs/Al*x*Ga1−*x*As multiple-quantum-well structure. *Phys. Rev. B*, 38(9):6160, 1988.
76. A. Larsson, P. A. Andrekson, S. T. Eng, and A. Yariv. Tunable superlattice p-i-n photodetectors: characteristics, theory, and application. *IEEE J. Quant. Electron.*, 24(5):787, 1988.
77. D. J. Moss, T. Ido, and H. Sano. Calculation of photogenerated carrier escape rates from GaAs/Al*x*Ga1−*x*As quantum wells. *IEEE J. Quant. Electron.*, 30(4):1015, 1994.
78. C.-Y. Tsai, L. F. Eastman, Y.-H. Lo, and C.-Y. Tsai. Breakdown of thermionic emission theory for quantum wells. *Appl. Phys. Lett.*, 65(4):469–471, 1994.
79. U. Aeberhard. Microscopic theory and numerical simulation of quantum well solar cells. *Proc. SPIE*, 7597:759702, 2010.
80. P. N. Brounkov, A. A. Suvorova, M. V. Maximov, A. F. Tsatsul'nikov, A. E. Zhukov, A. Yu Egorov, A. R. Kovsh, S. G. Konnikov, T. Ihn, S. T. Stoddart, L. Eaves, and P. C. Main. Electron escape from self-assembled InAs/GaAs quantum dot stacks. *Physica B*, 249–251:267–270, 1998.
81. C. M. A. Kapteyn, F. Heinrichsdorff, O. Stier, R. Heitz, M. Grundmann, N. D. Zakharov, D. Bimberg, and P. Werner. Electron escape from InAs quantum dots. *Phys. Rev. B*, 60:14265–14268, 1999.
82. P. W. Fry, J. J. Finley, L. R. Wilson, A. Lemaitre, D. J. Mowbray, M. S. Skolnick, M. Hopkinson, G. Hill, and J. C. Clark. Electric-field-dependent carrier capture and escape in self-assembled InAs/GaAs quantum dots. *Appl. Phys. Lett.*, 77(26):4344–4346, 2000.
83. S. Marcinkevius and R. Leon. Carrier capture and escape in In*x*Ga1−*x*As/GaAs quantum dots: Effects of intermixing. *Phys. Rev. B*, 59:4630–4633, 1999.

84. M. De Giorgi, C. Lingk, G. von Plessen, J. Feldmann, S. De Rinaldis, A. Passaseo, M. De Vittorio, R. Cingolani, and M. Lomascolo. Capture and thermal re-emission of carriers in long-wavelength InGaAs/GaAs quantum dots. *Appl. Phys. Lett.*, 79(24):3968–3970, 2001.
85. J. A. Brum and G. Bastard. Resonant carrier capture by semiconductor quantum wells. *Phys. Rev. B*, 33:1420–1423, 1986.
86. P. W. M. Blom, J. E. M. Haverkort, and J. H. Wolter. Optimization of barrier thickness for efficient carrier capture in graded-index and separate-confinement multiple quantum well lasers. *Appl. Phys. Lett.*, 58(24):2767–2769, 1991.
87. S. C. Kan, D. Vassilovski, T. C. Wu, and K. Y. Lau. Quantum capture and escape in quantum-well lasers—Implications on direct modulation bandwidth limitations. *IEEE Photonic. Technol. Lett.*, 4(5):428–431, 1992.
88. P. Sotirelis, P. von Allmen, and K. Hess. Electron intersubband relaxation in doped quantum wells. *Phys. Rev. B*, 47:12744–12753, 1993.
89. P. Sotirelis and K. Hess. Electron capture in GaAs quantum wells. *Phys. Rev. B*, 49:7543–7547, 1994.
90. B. K. Ridley. Space-charge-mediated capture of electrons and holes in a quantum well. *Phys. Rev. B*, 50:1717–1724, 1994.
91. C.-Y. Tsai, L. F. Eastman, Y.-H. Lo, and C.-Y. Tsai. Carrier capture and escape in multisubband quantum well lasers. *IEEE Photonic. Technol. Lett.*, 6(9):1088–1090, 1994.
92. L. Thibaudeau and B. Vinter. Phonon-assisted carrier capture into a quantum well in an electric field. *Appl. Phys. Lett.*, 65(16):2039–2041, 1994.
93. D. Bradt, Y. M. Sirenko, and V. Mitin. Inelastic and elastic mechanisms of electron capture to a quantum well. *Semicond. Sci. Technol.*, 10(3):260, 1995.
94. K. Kalna and M. Mosko. Electron capture in quantum wells via scattering by electrons, holes, and optical phonons. *Phys. Rev. B*, 54:17730–17737, 1996.
95. M. Mosko and K. Kalna. Carrier capture into a GaAs quantum well with a separate confinement region: Comment on quantum and classical aspects. *Semicond. Sci. Technol.*, 14(9):790, 1999.
96. J. A. Brum and G. Bastard. Direct and indirect carrier capture by semiconductor quantum wells. *Superlattice. Microst.*, 3(1):51–55, 1987.
97. P. W. M. Blom, C. Smit, J. E. M. Haverkort, and J. H. Wolter. Carrier capture into a semiconductor quantum well. *Phys. Rev. B*, 47(4):2072–2081, 1993.
98. C.-Y. Tsai, C.-Y. Tsai, Y.-H. Lo, and L. F. Eastman. Carrier dc and ac capture and escape times in quantum-well lasers. *IEEE Photonic. Technol. Lett.*, 7(6):599–601, 1995.
99. C.-Y. Tsai, C.-Y. Tsai, Y.-H. Lo, R. M. Spencer, and L. F. Eastman. Nonlinear gain coefficients in semiconductor quantum-well lasers: Effects of carrier diffusion, capture, and escape. *IEEE J. Sel. Top. Quant. Electron.*, 1(2):316–330, 1995.
100. Y. Lam and J. Singh. Monte Carlo studies on the well-width dependence of carrier capture time in graded-index separate confinement heterostructure quantum well laser structures. *Appl. Phys. Lett.*, 63(14):1874–1876, 1993.
101. L. Davis, Y. L. Lam, Y. C. Chen, J. Singh, and P. K. Bhattacharya. Carrier capture and relaxation in narrow quantum wells. *IEEE J. Quant. Electron.*, 30(11):2560–2564, 1994.
102. L. F. Register and K. Hess. Simulation of carrier capture in quantum well lasers due to strong inelastic scattering. *Superlattice. Microst.*, 18(3):223–228, 1995.
103. G. C. Crow and R. A. Abram. Monte Carlo simulations of capture into quantum well structures. *Semicond. Sci. Technol.*, 14(1):1, 1999.
104. M. Ryzhii and V. Ryzhii. Monte Carlo modeling of electron transport and capture processes in AlGaAs/GaAs multiple quantum well infrared photodetectors. *Jpn. J. Appl. Phys.*, 38(10R):5922, 1999.
105. J. Hader, J. V. Moloney, and S. W. Koch. Structural dependence of carrier capture time in semiconductor quantum-well lasers. *Appl. Phys. Lett.*, 85(3):369–371, 2004.

106. A. V. Uskov, J. McInerney, F. Adler, H. Schweizer, and M. H. Pilkuhn. Auger carrier capture kinetics in self-assembled quantum dot structures. *Appl. Phys. Lett.*, 72(1):58–60, 1998.
107. I. Magnusdottir, S. Bischoff, A. V. Uskov, and J. Mørk. Geometry dependence of Auger carrier capture rates into cone-shaped self-assembled quantum dots. *Phys. Rev. B*, 67:205326, 2003.
108. A. V. Uskov, I. Magnusdottir, B. Tromborg, J. Mørk, and R. Lang. Line broadening caused by Coulomb carrier-carrier correlations and dynamics of carrier capture and emission in quantum dots. *Appl. Phys. Lett.*, 79(11):1679–1681, 2001.
109. R. Ferreira and G. Bastard. Phonon-assisted capture and intradot Auger relaxation in quantum dots. *Appl. Phys. Lett.*, 74(19):2818–2820, 1999.
110. J. M. Miloszewski, M. S. Wartak, S. G. Wallace, and S. Fafard. Theoretical investigation of carrier capture and escape processes in cylindrical quantum dots. *J. Appl. Phys.*, 114(15):154311, 2013.
111. J.-Z. Zhang and I. Galbraith. Rapid hot-electron capture in self-assembled quantum dots via phonon processes. *Appl. Phys. Lett.*, 89(15):153119, 2006.
112. I. Magnusdottir, A. V. Uskov, R. Ferreira, G. Bastard, J. Mørk, and B. Tromborg. Influence of quasibound states on the carrier capture in quantum dots. *Appl. Phys. Lett.*, 81(23):4318–4320, 2002.
113. T. R. Nielsen, P. Gartner, and F. Jahnke. Many-body theory of carrier capture and relaxation in semiconductor quantum-dot lasers. *Phys. Rev. B*, 69:235314, 2004.
114. J. Seebeck, T. R. Nielsen, P. Gartner, and F. Jahnke. Polarons in semiconductor quantum dots and their role in the quantum kinetics of carrier relaxation. *Phys. Rev. B*, 71:125327, 2005.
115. G. D. Mahan. *Many-Particle Physics*, 2nd ed. New York, NY: Plenum, 1990.
116. T. Gunst, T. Markussen, K. Stokbro, and M. Brandbyge. First-principles method for electron-phonon coupling and electron mobility: Applications to two-dimensional materials. *Phys. Rev. B*, 93:035414, 2016.
117. C.-W. Jiang and M. A. Green. Silicon quantum dot superlattices: Modeling of energy bands, densities of states, and mobilities for silicon tandem solar cell applications. *J. Appl. Phys.*, 99(11):114902, 2006.
118. A. Wacker. Semiconductor superlattices: A model system for nonlinear transport. *Phys. Rep.*, 357:1–111, 2002.
119. P. Würfel. *Physics of Solar Cells: From Basic Principles to Advanced Concepts*. Weinheim: Wiley-VCH, 2009.
120. I. Mora-Sero, L. Bertoluzzi, V. Gonzalez-Pedro, S. Gimenez, F. Fabregat-Santiago, K. W. Kemp, E. H. Sargent, and J. Bisquert. Selective contacts drive charge extraction in quantum dot solids via asymmetry in carrier transfer kinetics. *Nat. Commun.*, 4:2272, 2013.
121. W. Van Roosbroeck and W. Shockley. Photon-radiative recombination of electrons and holes in germanium. *Phys. Rev.*, 94:1558, 1954.
122. E. O. Göbel, H. Jung, J. Kuhl, and K. Ploog. Recombination enhancement due to carrier localization in quantum well structures. *Phys. Rev. Lett.*, 51(17):1588–1591, 1983.
123. A. R. Beattie and P. T. Landsberg. Auger effect in semiconductors. *Proc. R. Soc. London Ser. A*, 249:16, 1959.
124. P. T. Landsberg. *Recombination in Semiconductors.* Cambridge: Cambridge University Press, 1991.
125. L. Chiu and A. Yariv. Auger recombination in quantum-well InGaAsP heterostructure lasers. *IEEE J. Quant. Electron.*, 18(10):1406–1409, 1982.
126. C. Smith, R. A. Abram, and M. G. Burt. Auger recombination in a quantum well heterostructure. *J. Phys. C Solid State Phys.*, 16(5):L171, 1983.
127. N. K. Dutta. Calculation of Auger rates in a quantum well structure and its application to InGaAsP quantum well lasers. *J. Appl. Phys.*, 54(3):1236–1245, 1983.
128. A. Sugimura. Auger recombination effect on threshold current of InGaAsP quantum well lasers. *IEEE J. Quant. Electron.*, 19(6):932–941, 1983.
129. P. K. Basu. Auger recombination rate in quantum well lasers: Modification by electronelectron interaction in quasi two dimensions. *J. Appl. Phys.*, 56(11):3344–3346, 1984.

130. A. Haug. Phonon-assisted auger recombination in quantum well semiconductors. *Appl. Phys. A*, 51:354, 1990.
131. M. Takeshima. Phonon-assisted Auger recombination in a quasi-two-dimensional structure semiconductor. *Phys. Rev. B*, 30(6):3302–3308, 1984.
132. L.-W. Wang, M. Califano, A. Zunger, and A. Franceschetti. Pseudopotential theory of auger processes in CdSe quantum dots. *Phys. Rev. Lett.*, 91:056404, 2003.
133. G. A. Narvaez, G. Bester, and A. Zunger. Carrier relaxation mechanisms in self-assembled (In,Ga)As/GaAs quantum dots: Efficient $P \rightarrow S$ Auger relaxation of electrons. *Phys. Rev. B*, 74:075403, 2006.
134. S. Tomić. Intermediate-band solar cells: Influence of band formation on dynamical processes in InAs/GaAs quantum dot arrays. *Phys. Rev. B*, 82(19):195321, 2010.
135. K. Hyeon-Deuk and O. V. Prezhdo. Time-domain ab initio study of Auger and phonon-assisted auger processes in a semiconductor quantum dot. *Nano Lett.*, 11(4):1845–1850, 2011.
136. N. J. Ekins-Daukes, K. W. J. Barnham, J. P. Connolly, J. S. Roberts, J. C. Clark, G. Hill, and M. Mazzer. Strain-balanced GaAsP/InGaAs quantum well solar cells. *Appl. Phys. Lett.*, 75(26):4195–4197, 1999.
137. K. Ding, U. Aeberhard, O. Astakhov, U. Breuer, M. Beigmohamadi, S. Suckow, B. Berghoff, W. Beyer, F. Finger, R. Carius, and U. Rau. Defect passivation by hydrogen reincorporation for silicon quantum dots in SiC/SiO*x*hetero-superlattice. *J. Non-Cryst. Solids*, 358(17):2145–2149, 2012.
138. W. Shockley and W. T. Read. Statistics of the recombinations of holes and electrons. *Phys. Rev.*, 87:835–842, 1952.
139. R. N. Hall. Electron-hole recombination in germanium. *Phys. Rev.*, 87:387, 1952.
140. A. Schenk. An improved approach to the Shockley-Read-Hall recombination in inhomogeneous fields of space-charge regions. *J. Appl. Phys.*, 71(7):3339–3349, 1992.
141. A. Alkauskas, Q. Yan, and C. G. Van de Walle. First-principles theory of nonradiative carrier capture via multiphonon emission. *Phys. Rev. B*, 90:075202, 2014.
142. U. Aeberhard. Simulation of nanostructure-based high-efficiency solar cells: Challenges, existing approaches, and future directions. *IEEE J. Sel. Top. Quant. Electron*, 19(5):4000411, 2013.
143. G. L. Araújo and A. Martí. Absolute limiting efficiencies for photovoltaic energy conversion. *Sol. Energy Mater. Sol. Cells*, 33:213, 1994.
144. U. Rau. Reciprocity relation between photovoltaic quantum efficiency and electroluminescent emission of solar cells. *Phys. Rev. B*, 76(8):085303, 2007.
145. M. Born and E. Wolf. *Principles of Optics*, 7th ed. Cambridge: Cambridge University Press, 1999.
146. P. Würfel. The chemical potential of radiation. *J. Phys. C Solid State Phys.*, 15:3967, 1982.
147. U. Aeberhard. Simulation of ultra-thin solar cells beyond the limits of the semi-classical bulk picture. *IEEE J. Photovolt.*, 6(3):654–660, 2016.
148. U. Aeberhard. Impact of built-in fields and contact configuration on the characteristics of ultra-thin GaAs solar cells. *Appl. Phys. Lett.*, 109(3), 2016.
149. N. G. Anderson. Ideal theory of quantum well solar cells. *J. Appl. Phys.*, 78:1861, 1995.
150. G. L. Araújo, A. Martí, F. W. Ragay, and J. H. Wolter. Efficiency of multiple quantum well solar cells. In *Proceeding of 12th European Photovoltaic Solar Energy Conference*, Amsterdam, The Netherlands, page 1481, 1994.
151. G. L. Araújo and A. Martí. Electroluminescence coupling in multiple quantum well diodes and solar cells. *Appl. Phys. Lett.*, 66(7):894–895, 1995.
152. S. P. Bremner, R. Corkish, and C. B. Honsberg. Detailed balance efficiency limits with quasi-Fermi level variations. *IEEE Trans. Electron Dev.*, 46(10):1932–1939, 1999.
153. A. Luque, A. Martí, and L Cuadra. Thermodynamic consistency of sub-bandgap absorbing solar cell proposals. *IEEE Trans. Electron Dev.*, 48:2118, 2001.
154. J. P. Connolly, J. Nelson, K. W. J. Barnham, I. Ballard, C. Roberts, J. S. Roberts, and C. T. Foxon. Simulating multiple quantum well solar cells. In *Proceeding of 28th IEEE Photovoltaic Specialists Conference*, Anchorage, Alaska, USA, page 1304, 2000.

155. J. C. Rimada and L. Hernández. Modelling of ideal AlGaAs quantum well solar cells. *Microelectron. J.*, 32(9):719–723, 2001.
156. S. J. Lade and A. Zahedi. A revised ideal model for AlGaAs/GaAs quantum well solar cells. *Microelectron. J.*, 35(5):401–410, 2004.
157. J. C. Rimada, L. Hernandez, J. P. Connolly, and K. W. J. Barnham. Quantum and conversion efficiency calculation of AlGaAs/GaAs multiple quantum well solar cells. *Phys. Status Solidi B*, 242(9):1842–1845, 2005.
158. J. C. Rimada, L. Hernández, J. P. Connolly, and K. W J Barnham. Conversion efficiency enhancement of AlGaAs quantum well solar cells. *Microelectron. J.*, 38(4–5):513–518, 2007.
159. M. Courel, J. C. Rimada, and L. Hernández. GaAs/GaInNAs quantum well and superlattice solar cell. *Appl. Phys. Lett.*, 100(7):073508, 2012.
160. C. I. Cabrera, J. C. Rimada, M. Courel, L. Hernandez, J. P. Connolly, A. Enciso, and D. A. Contreras-Solorio. Modeling multiple quantum well and superlattice solar cells. *Nat. Resour.*, 4(03):235–245, 2013.
161. A. Chatterjee, A. Kumar Biswas, and A. Sinha. An analytical study of the various current components of an AlGaAs/GaAs multiple quantum well solar cell. *Physica E*, 72:128–133, 2015.
162. M. Courel, J. C. Rimada, and L. Hernandez. An approach to high efficiencies using GaAs/GaInNAs multiple quantum well and superlattice solar cell. *J. Appl. Phys.*, 112(5):054511, 2012.
163. A. Alemu and A. Freundlich. Resonant thermotunneling design for high-performance single-junction quantum-well solar cells. *IEEE J. Photovolt.*, 2(3):256–260, 2012.
164. M. Courel, J. C. Rimada, and L. Hernández. AlGaAs/GaAs superlattice solar cells. *Prog. Photovolt. Res. Appl.*, 21(3):276–282, 2013.
165. A. Luque and A. Martí. Increasing the efficiency of ideal solar cells by photon induced transitions at intermediate levels. *Phys. Rev. Lett.*, 78(26):5014–5017, 1997.
166. Z. Wang, T. P. White, and K. R. Catchpole. Plasmonic near-field enhancement for planar ultra-thin photovoltaics. *IEEE Photon. J.*, 5(5):8400608, 2013.
167. W. Hu, M. Maksudur Rahman, M.-Y. Lee, Y. Li, and S. Samukawa. Simulation study of type-II Ge/Si quantum dot for solar cell applications. *J. Appl. Phys.*, 114(12):124509, 2013.
168. Q. Shao, A. A. Balandin, A. I. Fedoseyev, and M. Turowski. Intermediate-band solar cells based on quantum dot supracrystals. *Appl. Phys. Lett.*, 91(16):163503, 2007.
169. Q. Deng, X. Wang, H. Xiao, C. Wang, H. Yin, H. Chen, Q. Hou, D. Lin, J. Li, Z. Wang, and X. Hou. An investigation on $In_xGa_{1-x}N$/GaN multiple quantum well solar cells. *J. Phys. D Appl. Phys.*, 44(26):265103, 2011.
170. W. Wei, Q. Zhang, S. Zhao, and Y. Zhang. Two intermediate bands solar cells of InGaN/InN quantum dot supracrystals. *Appl. Phys. A*, 116(3):1009–1016, 2014.
171. A. Mellor, A. Luque, I. Tobias, and A. Marti. Realistic detailed balance study of the quantum efficiency of quantum dot solar cells. *Adv. Funct. Mater.*, 24(3):339–345, 2014.
172. G. Wei, K.-T. Shiu, N. C. Giebink, and S. R. Forrest. Thermodynamic limits of quantum photovoltaic cell efficiency. *Appl. Phys. Lett.*, 91(22), 2007.
173. J. Nelson, I. Ballard, K. Barnham, J. P. Connolly, J. S. Roberts, and M. Pate. Effect of quantum well location on single quantum well p-i-n photodiode dark currents. *J. Appl. Phys.*, 86:5898, 1999.
174. P. Kailuweit, R. Kellenbenz, S. P. Philipps, W. Guter, A. W. Bett, and F. Dimroth. Numerical simulation and modeling of GaAs quantum-well solar cells. *J. Appl. Phys.*, 107(6):064317, 2010.
175. T. Kirchartz, K. Seino, J.-M. Wagner, U. Rau, and F. Bechstedt. Efficiency limits of Si/SiO2 quantum well solar cells from first-principles calculations. *J. Appl. Phys.*, 105(10):104511, 2009.
176. C.-Y. Tsai and C.-Y. Tsai. Effects of carrier escape and capture processes on quantum well solar cells: A theoretical investigation. *IET Optoelectron.*, 3(6):300–304, 2009.
177. O. Kengradomying, S. Jiang, Q. Wang, N. Vogiatzis, and J. M. Rorison. Modelling escape and capture processes in GaInNAs quantum well solar cells. *Phys. Status Solidi C Curr. Top. Solid State Phys.*, 10(4):585–588, 2013.

178. J. Nelson, K. W. J Barnham, J. P Connolly, G. Haarpaintner, C. Button, and J. Roberts. Quantum well solar cell dark currents. In *Proceeding of 12th EU PVSEC*, Amsterdam, The Netherlands, 1994.
179. M. Gioannini, A. P. Cedola, N. Di Santo, F. Bertazzi, and F. Cappelluti. Simulation of quantum dot solar cells including carrier intersubband dynamics and transport. *IEEE J. Photovolt.*, 3(4):1271–1278, 2013.
180. K. Driscoll, M. F. Bennett, S. J. Polly, D. V. Forbes, and S. M. Hubbard. Effect of quantum dot position and background doping on the performance of quantum dot enhanced GaAs solar cells. *Appl. Phys. Lett.*, 104(2):023119, 2014.
181. K. Yoshida, Y. Okada, and N. Sano. Self-consistent simulation of intermediate band solar cells: Effect of occupation rates on device characteristics. *Appl. Phys. Lett.*, 97(13):133503, 2010.
182. Y.-X. Gu, X.-G. Yang, H.-M. Ji, P.-F. Xu, and T. Yang. Theoretical study of the effects of InAs/GaAs quantum dot layer's position in i-region on current-voltage characteristic in intermediate band solar cells. *Appl. Phys. Lett.*, 101(8):081118, 2012.
183. T. Sogabe, T. Kaizu, Y. Okada, and S. Tomić. Theoretical analysis of GaAs/AlGaAs quantum dots in quantum wire array for intermediate band solar cell. *J. Renew. Sustain. Ener.*, 6(1):011206, 2014.
184. V. Aroutiounian, S. Petrosyan, A. Khachatryan, and K. Touryan. Quantum dot solar cells. *J. Appl. Phys.*, 89(4):2268, 2001.
185. A. Martí, L. Cuadra, and A. Luque. Quasi-drift diffusion model for the quantum dot intermediate band solar cell. *IEEE Trans. Electron Dev.*, 49(9):1632–1639, 2002.
186. C.-C. Lin, M.-H. Tan, C.-P. Tsai, K.-Y. Chuang, and T. S. Lay. Numerical study of quantum-dot-embedded solar cells. *IEEE J. Sel. Top. Quant. Electron.*, 19(5), 2013.
187. V. Aroutiounian, S. Petrosyan, and A. Khachatryan. Studies of the photocurrent in quantum dot solar cells by the application of a new theoretical model. *Sol. Energy Mater. Sol. Cells*, 89(2–3):165–173, 2005.
188. S. M. Willis, J. A. R. Dimmock, F. Tutu, H. Y. Liu, M. G. Peinado, H. E. Assender, A. A. R. Watt, and I. R. Sellers. Defect mediated extraction in InAs/GaAs quantum dot solar cells. *Sol. Energy Mater. Sol. Cells*, 102:142–147, 2012.
189. U. Aeberhard. Theory and simulation of quantum photovoltaic devices based on the non-equilibrium Green's function formalism. *J. Comput. Electron.*, 10:394–413, 2011.
190. D. A. Stewart and F. Leonard. Photocurrents in nanotube junctions. *Phys. Rev. Lett.*, 93:107401, 2004.
191. D. A. Stewart and F. Leonard. Energy conversion efficiency in nanotube optoelectronics. *Nano Lett.*, 5:219, 2005.
192. J. Guo, M. A. Alam, and Y. Yoon. Theoretical investigation on photoconductivity of single intrinsic carbon nanotubes. *Appl. Phys. Lett.*, 88(13):133111, 2006.
193. A. Buin, A. Verma, and S. Saini. Optoelectronic response calculations in the framework of k.p coupled to non-equilibrium Green's functions for 1D systems in the ballistic limit. *J. Appl. Phys.*, 114(3):033111, 2013.
194. U. Aeberhard and R. H. Morf. Microscopic nonequilibrium theory of quantum well solar cells. *Phys. Rev. B*, 77:125343, 2008.
195. U. Aeberhard. Spectral properties of photogenerated carriers in quantum well solar cells. *Sol. Energy Mater. Sol. Cells*, 94(11):1897–1902, 2010.
196. N. Cavassilas, F. Michelini, and M. Bescond. Theoretical comparison of multiple quantum wells and thick-layer designs in InGaN/GaN solar cells. *Appl. Phys. Lett.*, 105(6), 2014.
197. U. Aeberhard. Theory and simulation of photogeneration and transport in Si-SiOx superlattice absorbers. *Nanoscale Res. Lett.*, 6:242, 2011.
198. U. Aeberhard. Simulation of nanostructure-based and ultra-thin film solar cell devices beyond the classical picture. *J. Photon. Energy*, 4(1):042099, 2014.
199. U. Aeberhard. Effective microscopic theory of quantum dot superlattice solar cells. *Opt. Quant. Electron.*, 44:133–140, 2012.
200. A. Berbezier and U. Aeberhard. Impact of nanostructure configuration on the photovoltaic performance of quantum-dot arrays. *Phys. Rev. Applied*, 4:044008, 2015

42

Nanowire Solar Cells: Electro-Optical Performance

Bernd Witzigmann

42.1 Introduction

Nanowire arrays show unique optical and electronic features for photovoltaic applications. It has been demonstrated that they exhibit intrinsic optical antireflection properties [1,2]. Moreover, they also can act as optical microconcentrators [3] that focus light to an active region. Nanowires have a large aspect ratio, improving the electronic properties by combining a long optical absorption path along the wire axis and a short lateral carrier extraction [4]. Short carrier extraction paths are beneficial for the carrier collection efficiency. These mechanisms are illustrated in Figure 42.1. On the left, a conventional bulk solar cell collects light in the vertical direction, and the photogenerated carriers move to the respective contacts causing the photo current. A nanostructured solar cell (Figure 42.1a) can act as an optical concentrator in order to focus the incoming light to a small area. Moreover, the carrier extraction path length to the contacts is reduced, which minimizes the probability of carrier loss by unwanted recombination.

Finally, multijunction cells can be realized with strain-relaxed axial heterointerfaces. This allows optimum bandgap combinations [5] and low-defect densities at the same time. As a potential drawback, the large surface to volume ratio can lead to enhanced surface recombination. Epitaxial passivation techniques have been developed in order to suppress surface recombination [6].

In this chapter, a detailed simulation study of the electro-optical characteristics of nanowire array solar cells is presented. An efficiency analysis is derived that treats solar cells with nanostructures on a broader basis using the Shockley–Queisser framework [7]. As addition, the modification of optical density of states (DOS) and wave optics effects are included. Using this framework, the numerical simulation of nanostructured solar cells is discussed. The simulation model consists of the three-dimensional solution of the vectorial Helmholtz equation for solar illumination and a drift-diffusion/Poisson equation system for

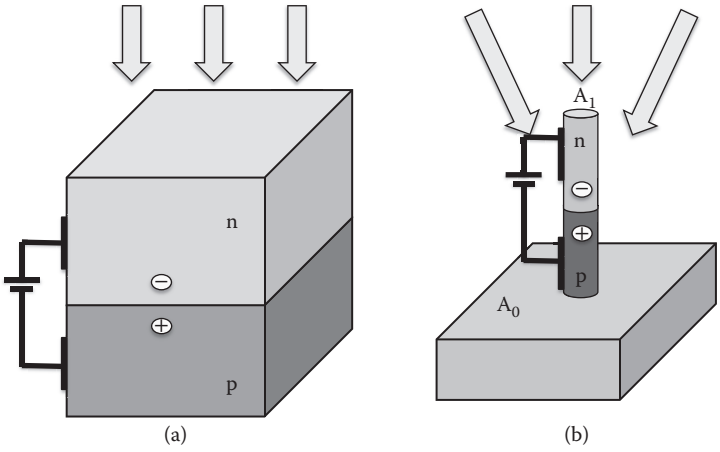

FIGURE 42.1 (a) Bulk solar cell with p-n junction. (b) Solar cell structured as a nanowire.

carrier extraction. In the results section, it is first shown how a nanowire array can act as an electromagnetic broadband absorber. Second, the choice of material composition and dopant placement is discussed with respect to efficiency in single-junction designs. Surface recombination is discussed for both radial and axial p-n junction arrangement.

42.2 Efficiency of Nanostructured Solar Cells

In this section, the efficiency of solar cells that contain nanostructures is discussed. Nanostructures can modify the optical or the electronic properties. As a basic theory, the established derivation by Shockley and Queisser [7] is used, which assumes detailed balance in the radiative limit, i.e., nonradiative recombination processes are not included. The main extension of Shockley and Queisser's classical model is to include the fact that light is viewed as electromagnetic waves and their modes in resonant structures. This leads to a modal dispersion and a strong local variation of the electromagnetic power. Moreover, resonant structures modify the optical DOS, and those enter the absorption as well as the emission properties of a cell, which is based on the Purcell effect [8].

The focus will be put on semiconductor materials, although the derivation can be transferred to other materials in a straightforward manner.

42.2.1 Solar Cell Current Density

The electronic current density of the solar cell can be easily obtained from current continuity in the steady state. It is derived from Maxwell's equations and builds on the mass conservation principle of charged particles:

$$\nabla \vec{J} = q(G - R), \tag{42.1}$$

with $\vec{J}$ being the local current density, G and R representing generation and recombination rates, respectively, and q being the elementary charge. In the integral form, the continuity equation reads

$$\int \vec{J} \mathrm{d}\vec{A} = q \int (G - R)\, \mathrm{d}V. \tag{42.2}$$

The surface area integration on the left hand side is carried out over the contacts (where carriers can leave or enter the solar cell), and the right hand side integration includes the solar cell volume.

42.2.2 Thermal Equilibrium

In thermal equilibrium, all of the emission and absorption processes of the solar cell can be equated, as there is no net current at the contacts. The left hand side of Equation 42.2 is zero, therefore

$$\int G\mathrm{d}V = \int R\mathrm{d}V = \int G^0\mathrm{d}V = \int R^0\mathrm{d}V. \tag{42.3}$$

The local generation rate per volume is

$$G^0 = \int \alpha \mathrm{d}j_{\mathrm{phot}}, \tag{42.4}$$

with α being the absorption coefficient (a material property) and j_{phot} is the photon current density (in units $1/\mathrm{m}^2\mathrm{s}$), which in thermal equilibrium is basically a blackbody radiation at the given reservoir temperature T_0. In order to relate the photon current density to the blackbody radiation (which is an energy density per spectral width), the following relation is used:

$$\mathrm{d}j_{\mathrm{phot}} = \mathrm{d}E_{\mathrm{phot}} \frac{c}{n} \frac{1}{\hbar\omega} \tag{42.5}$$

and

$$dE_{\mathrm{phot}} = D_{\mathrm{opt}} f_{\mathrm{phot}} \hbar\omega \mathrm{d}\hbar\omega. \tag{42.6}$$

The optical DOS D_{opt} depends on the dielectric surrounding of the solar cell and f_{phot} is the photon distribution function. Within an electromagnetic theory, D_{opt} can be calculated as the electromagnetic response of a local dipole excitation [9]:

$$D_{\mathrm{opt}}(\omega, \vec{x}) = \frac{1}{2\pi} \sum_m \int_{\vec{k}} \mathrm{d}\vec{k}\delta(\omega - \omega_m) |\vec{e}_{\mathrm{d}} \vec{E}_{m,\vec{k}}(\vec{r}|^2, \tag{42.7}$$

where $\vec{E}$ is the electric field as obtained from the solution of Maxwell's equations, $\vec{e}_{\mathrm{d}}$ is the dipole unit source, and ω_m the eigenfrequencies of the system.

The photon distribution function in equilibrium is

$$f_{\mathrm{phot}} = \frac{1}{\exp\left(\frac{\hbar\omega}{k_B T_0}\right) - 1}. \tag{42.8}$$

The generation rate, therefore, can be expressed as

$$G^0(\vec{x}) = R^0(\vec{x}) = \int \alpha(\vec{x}, \hbar\omega) \frac{c}{n_r} D^{\mathrm{sc}}_{\mathrm{opt}}(\vec{x}, \hbar\omega) f_{\mathrm{phot}}(T_0) d\hbar\omega, \tag{42.9}$$

which is equal to the recombination in equilibrium.

42.2.3 Nonequilibrium

In the nonequilibrium case, the solar cell is illuminated by solar radiation power, which in our case is represented by an external blackbody source at the temperature T_{s}. This leads to an energy density outside of the solar cell of

$$\mathrm{d}E^{\mathrm{s}}_{\mathrm{phot}} = D^{\mathrm{vac}}_{\mathrm{opt}} f_{\mathrm{phot}}(T_{\mathrm{s}}) \hbar\omega d\hbar\omega. \tag{42.10}$$

The optical DOS here is the vacuum DOS, and the solar temperature T_s enters the Bose–Einstein photon distribution function. In principle, other illumination sources can be used, which then enter the calculation, e.g., as tabulated energy density. The external photon source generates additional electron–hole pairs ΔG inside the solar cell, and therefore

$$G^s = \int \alpha a \frac{c}{n_r^{\text{vac}}} D_{\text{opt}}^{\text{vac}} f_{\text{phot}}(T_s) \mathrm{d}\hbar\omega. \tag{42.11}$$

Here, a is the relative absorptivity, the amount of electromagnetic solar power entering the solar cell. It can be calculated using electromagnetic theory by

$$a(\vec{x}) = \frac{\langle \int Re\left\{\nabla \vec{S}\right\} \mathrm{d}V\rangle}{\int Re\left\{\vec{S}\right\} \mathrm{d}\vec{A}}, \tag{42.12}$$

where $\vec{S}$ is the Poynting vector, and the integration in the denominator is performed over a closed surface around the solar cell. This is especially important for nonplanar devices where the solar energy is captured not only from the surface, but from an angle of incidence larger than the half space π. The $\langle\rangle$ symbol in the numerator denotes an averaging over a volume that contains several optical wavelengths so that an inclusion into a drift-diffusion carrier transport framework is possible. The absorptivity $a(\vec{x})$ is the local relative electromagnetic energy entering in the solar cell, and in the ray optics approximation often is expressed as $(1 - r)\exp(-\alpha x)$, where the first factor is the back-reflected power from the air–top surface interface, and the second factor is the diminishing optical intensity inside the device due to internal absorption (Beer–Lambert law).

In the nonequilibrium case, the generation in the cell is increased, which in turn leads to an increased radiative recombination rate. Assuming that the radiative recombination is proportional to the product of electron and hole density, one gets

$$R^s = R^0 \frac{np}{n_i^2} = \frac{N_c N_v}{n_i^2} \exp\left(\frac{\eta_c \eta_v}{k_B T_0}\right) G^0 = G^0 \exp\left(\frac{\Delta E_F}{k_B T_0}\right), \tag{42.13}$$

with the following nomenclature: The intrinsic density is n_i, the effective state densities are N_c or N_v, the reduced Fermi energies are $\eta_c = E_F^n - E^c$ and $\eta_v = -E_F^p + E^v$, and ΔE_F is the quasi-Fermi level separation.

With these quantities, the current densities of the solar cell can be calculated. Using the current continuity equation, one gets

$$\int \vec{J} \mathrm{d}\vec{A} = q \int (G^s - R^s) \mathrm{d}V = q \int \left(G^s - G^0 \exp(\frac{\Delta E_F}{k_B T_0})\right) \mathrm{d}V = I_{\text{gen}} - I_{\text{rec}}. \tag{42.14}$$

The quasi-Fermi level separation can be approximated by the voltage at the contacts, and after integration of the left hand side, the current of the solar cell becomes

$$I = I_{\text{gen}} - I_{\text{sat}} \exp\left(\frac{qV}{k_B T_0}\right). \tag{42.15}$$

The currents introduced here are the total generation and recombination currents I_{gen} and I_{rec}, and the reverse saturation current I_{sat}. They can be evaluated from

$$I_{\text{sat}} = q \int \int \alpha \frac{c}{n_r^{\text{sc}}} D_{\text{opt}}^{\text{sc}} f_{\text{phot}}(T_0) \, \mathrm{d}\hbar\omega \mathrm{d}V \tag{42.16}$$

and

$$I_{\text{gen}} = q \int \int \alpha a \frac{c}{n_r^{\text{vac}}} D_{\text{opt}}^{\text{vac}} f_{\text{phot}}(T_s) \mathrm{d}\hbar\omega \mathrm{d}V. \tag{42.17}$$

The short-circuit current of a solar cell is the current under illumination with zero voltage at the contacts, and therefore accounts to

$$I_{\text{sc}} = I_{\text{gen}} - I_{\text{sat}}. \tag{42.18}$$

With this definition, the current–voltage relationship can be written in the more conventional Shockley form by replacing I_{gen} in Equation 42.15:

$$I = I_{\text{sat}} \left(1 - \exp(\frac{qV}{k_B T_0}) \right) + I_{\text{sc}}. \tag{42.19}$$

The open-circuit voltage is the voltage where the net current flow at the contacts is zero, and it can be evaluated from the relations obtained here as

$$V_{\text{oc}} = \frac{k_B T_0}{q} \log \left(\frac{I_{\text{gen}}}{I_{\text{sat}}} \right) = \frac{k_B T_0}{q} \log \left(1 + \frac{I_{\text{sc}}}{I_{\text{sat}}} \right). \tag{42.20}$$

42.2.4 Efficiency Evaluation

From the current versus voltage characteristics, the electro-optical efficiency of a solar cell can be obtained from

$$\eta_{\text{sc}} = \max \left(\frac{IV}{P_{\text{inc}}} \right), \tag{42.21}$$

where the denominator is the total, spectrally integrated incident solar power. Obtaining a high efficiency translates into maximizing the open-circuit voltage V_{oc} and the short-circuit current I_{sc} as defined in Section 42.2.3, which maximizes the product IV. A maximum short-circuit current is achieved by capturing the incident broadband solar radiation and absorbing the photons in the solar cell. According to Equation 42.17, this requires a large absorption coefficient α, which is a material property, and a good absorptivity a, which is the ability to capture light into the solar cell, an electromagnetic property. In bulk solar cells, the absorptivity a can be maximized by broadband antireflection coatings. If the solar cell thickness is in the order of a few optical wavelengths, waveguiding or even evanescent wave effects can help maximize a. It will be shown in Section 42.2.4.2 how this is applied to nanowire solar cells.

The evaluation of the open-circuit voltage involves a maximum optical generation current I_{gen} and a minimum reverse saturation current I_{sat}, which is a challenge to fulfill. As can be seen from Equation 42.16, the major ingredient to I_{sat} is the optical DOS of the solar cell.

42.2.4.1 Bulk Solar Cell

First, the efficiency evaluation is shown for a bulk solar cell. In principle, all current commercial technologies fall into this category, also the thin-film designs. For a homogeneous medium, the optical DOS is

$$D_{\text{opt}}^{\text{hom}} = \left(\frac{n_r}{c} \right)^3 \frac{1}{4\pi\hbar^3} (\hbar\omega)^2 \Omega, \tag{42.22}$$

where Ω is the solid angle of the radiation emission and absorption, respectively.

For the bulk solar cell, the solid angle of illumination is $\Omega = \pi$, and therefore

$$D_{\text{opt}}^{\text{vac}} = \left(\frac{1}{c}\right)^3 \frac{1}{4\hbar^3}(\hbar\omega)^2. \tag{42.23}$$

The solid angle of emission at any location inside the solar cell is $\Omega = 4\pi$, leading to

$$D_{\text{opt}}^{\text{sc}} = \left(\frac{n_r}{c}\right)^3 \frac{1}{\hbar^3}(\hbar\omega)^2. \tag{42.24}$$

For the absorptivity a, a simple ray optics model can be applied, which leads to

$$a = (1 - r)(1 - \exp(-\alpha d)) \approx 1, \tag{42.25}$$

with r being the reflectivity at the air-semiconductor interface, and d the cell thickness. By design, one can realize a good antireflective coating and sufficient cell thickness, leading to the relation $a \approx 1$. The short-circuit current can therefore be maximized by engineering the cell.

The open-circuit voltage contains the ratio of $\frac{I_{\text{gen}}}{I_{\text{sat}}}$, which is approximately proportional to the ratio $\frac{D_{\text{opt}}^{\text{vac}} n_r^{\text{sc}}}{D_{\text{opt}}^{\text{sc}} n_r^{\text{vac}}}$. This results in a factor $\frac{1}{4n_r^2}$ in the short-circuit current calculation. The physics behind this factor is the light intensity concentration. The factor of 4 in the denominator can be reduced by efficient light management, as it comes from the photon collection from the angle π but photon emission into an angle of 4π [10]. Introducing a dielectric environment that allows emission into an angle π only (by, e.g., photonic crystals or conventional mirrors) increases the open-circuit voltage. Therefore, light management is an important physical mechanism for advanced solar cell design [10,11].

42.2.4.2 Optical Microconcentrators

Waveguides, resonators, or optical antennas can influence the optical absorption behavior of a solar cell. As first case, an optical concentrator that acts as absorber is taken that focuses the incoming solar radiation to the absorber area A_1, whereas the illuminated (large) area is A_0 (for illustration, see Figure 42.1). This leads to an absorptivity $a \propto \frac{A_0}{A_1} = f$, as the absorber concentrates the energy flux density to the area A_1. As a consequence, the short-circuit current is obtained with a much smaller volume of absorption compared to a nonconcentrated case (see integral in Equation 42.18). At the same time, the reverse saturation current (Equation 42.16) is also integrated over a smaller volume compared to a bulk solar cell. The ratio of $\frac{I_{\text{gen}}}{I_{\text{sat}}}$, and therefore the open circuit voltage, increases. However, two assumptions need to be made: First, the reverse saturation current density needs to be independent of the local carrier density, which is, within the Shockley theory, a valid assumption. Second, the optical concentrator modifies the optical DOS of the absorber. This needs to be taken into account in the open-circuit voltage calculation. Therefore, the open-circuit voltage is only increased if the following relation holds:

$$\frac{D_{\text{opt}}^{\text{vac}}}{D_{\text{opt}}^{\text{mc}}} f > \frac{D_{\text{opt}}^{\text{vac}}}{D_{\text{opt}}^{\text{bulk}}}, \tag{42.26}$$

or

$$\frac{D_{\text{opt}}^{\text{bulk}}}{D_{\text{opt}}^{\text{mc}}} f > 1, \tag{42.27}$$

where $D_{\text{opt}}^{\text{bulk}}$ is the optical DOS for the bulk cell and $D_{\text{opt}}^{\text{mc}}$ is the optical DOS for the micro concentrated cell, respectively; their ratio can be interpreted as the Purcell factor. This result shows an interesting design opportunity for high-efficiency nanostructured solar cells: If optical elements are introduced that focus the

solar electromagnetic energy to a small volume, the open-circuit voltage, and therefore the efficiency, can be increased compared to a bulk or thin-film solar cell. However, the Purcell factor, or the ratio of optical DOS, should not be increased at the same rate, as it compensates the concentration effect.

42.2.4.3 Electronic Nanostructures

In the field of lasers and light-emitting diodes, electronic nanostructures such as quantum wells and quantum dots have increased the performance considerably [12]. In solar cells, the application of electronic nanostructures has not been used as a standard feature to improve the cell efficiency. In general, the introduction of nanostructures alters the electronic band structure. It leads to the following consequences:

- The effective bandgap, i.e., the absorption edge, is shifted to the lowest subband energy difference.
- The absorption coefficient increases, as the electronic DOS is higher at the band edge compared to a bulk material, and in addition, the overlap of the envelope wave functions enters the absorption coefficient.
- The absorption coefficient becomes anisotropic due to the anisotropy of the underlying carrier wave functions and DOS.

As a drawback, electronic nanostructures typically have dimensions around 10 nm, which is much smaller than the optical wavelength or a typical absorption path length. Therefore, in order to obtain efficient absorption, multiple nanometer-sized regions need to be stacked in the direction of the light path or efficient focusing of the light onto the absorbing nanostructure needs to be done. Fundamentally, the anisotropy of the absorption and emission can be used to restrict the absorptivity α to the solid angles of the solar radiation, and to reduce the reverse saturation current by a maximum theoretical factor of 4 (see explanation in Section 42.2.4.1) [13]. In addition, the increase of the absorption coefficient leads to smaller active regions, which increases the ratio of short-circuit current to reverse saturation current as outlined in Section 42.2.4.2.

Finally, some advanced concepts have been pursued with electronic nanostructures such as intermediate band solar cells [14] or hot-carrier contacts [15]. The basic idea is to introduce additional energy levels with nanostructures in order to overcome the single bandgap limitation of the semiconductor. Such structures require nonequilibrium carrier transport models and are not discussed here.

42.3 Computational Models

In the previous sections, some fundamental properties of nanostructured solar cells have been described within the semianalytical framework of the Shockley–Queisser model. In general, the incoming light generates electron–hole pairs by absorption in the semiconductor. In Section 42.2, it has been shown that the efficiency of a solar cell depends on the solar optical generation $G^{s}(\vec{x})$, the generation in equilibrium $G^{0}(\vec{x})$, and the recombination under illumination $R^{s}(\vec{x})$ (see Equation 42.14). In principle, those are all local quantities. In nanostructured solar cells, both light intensity and carrier density can be controlled on the scale of tens of nanometers by optical waveguides, plasmonic subwavelength structures, or quantum wells or dots. This introduces a strong local variation of the physical quantities contributing to the cell efficiency. Computational models can capture the local variations, and a correct determination of the efficiency is feasible.

42.3.1 Optical Models

The governing optical equations for solar cells with structures similar or smaller than the optical wavelength are Maxwell's equations. The solar illumination is modeled as a broadband current source, and the resulting electromagnetic field contributes to the local carrier generation rate. There is a multitude of models to solve Maxwell's equations either analytically, semianalytically, or numerically.

A popular analytical solution is the plane wave ansatz, which after some simplifications leads to the transfer matrix method (TMM). Using TMM, light propagation in multilayer structures can be evaluated or the even simpler Beer–Lambert law can be obtained:

$$I(x) = I_0 e^{-\alpha x}, \tag{42.28}$$

with the intensity I.

Most semianalytical methods expand the solution of Maxwell's equations in modes. For nanowire solar cells, which are essentially cylinders, the modes of a cylindrical optical fiber can be used, and the solution is then represented as superposition of the fiber modes [16].

Numerical methods essentially discretize the simulation domain and solve the respective equations only on parts of the domain, e.g., vertices or edges. The finite-difference time-domain (FDTD) method [17] is a popular method that discretizes space and time using explicit finite differences. With sources placed in the simulation domain, the propagation of electromagnetic waves can be computed iteratively in space and time. The FDTD method is straightforward to implement and can be adapted for parallel computing in a natural manner. The original FDTD kernels [17] require rectangular tensor product meshes, which can be restrictive for complex geometries. Moreover, the solutions are only second-order accurate in space and time, which requires many grid points per wavelength for accurate solutions, in particular for optical frequencies. A rather elegant approach to overcome these limitations is the discontinuous Galerkin time-domain method [18]. It combines explicit time stepping with a local finite-element space discretization, and therefore powerful high-order accuracy and flexible meshes can be used. The local elements are connected with constraints for the numerical flux in order to ensure consistency and convergence to the exact solution. For solar cell simulations, which use the broadband solar spectrum as source, the time-domain methods allow the use of spectral broadband sources, with only one simulation run for the entire spectrum. Dispersive materials need special treatment in the time domain, and therefore additional computational resources.

A second class of numerical methods solve Maxwell's equations in the frequency domain. Here, monochromatic harmonic waves replace the time derivatives with an imaginary frequency factor, and any multifrequency solution can be built as superposition in the linear case. Maxwell's equations can be reduced to the Helmholtz equation, which for the electric field $\vec{E}$ reads

$$\left(\Delta - \omega^2 \mu\epsilon\right) \vec{E} = -j\mu\vec{J}, \tag{42.29}$$

where μ and ϵ are the respective permeability and permittivity, ω is the frequency, and $\vec{J}$ is a current source. This equation is valid for homogeneous isotropic materials, and the permittivity is in general a dispersive complex quantity in optics. This equation can be solved by a finite-element method in three dimensions, with sophisticated meshes, and an approximation to the exact solution. In its homogeneous form (without current sources), it yields an eigenvalue problem that give the eigenfrequencies and eigenvectors of a given geometry. For complex permittivities, the eigenvalue problem becomes complex as well, and the imaginary part of the eigenfrequency gives the quality factor of the eigenmode [19,20].

In order to evaluate the efficiency of a solar cell beyond the analytical Shockley–Queisser limit, the framework presented in Section 42.2 as efficiency analysis for nanostructured solar cells requires evaluation of the classical electromagnetic field for the solar cell geometry in the presence of a complex refractive index. This enters both the absorptivity a (see Equation 42.12) and the optical DOS D_{opt} (Equation 42.7).

42.3.2 Electronic Models

Any electronic model describes the dynamics of the charge carriers inside the solar cell, i.e., the photogenerated carriers and the intrinsic carriers under applied bias. In a generalized description, solar cell operation means that the photogenerated carriers thermally relax to the band edges and are subsequently driven to

the contacts creating the photocurrent. In the radiative (i.e., Shockley–Queisser) limit, the carriers can recombine radiatively on their way to the contacts, and of course, the resulting photons can be reabsorbed. This gives an upper limit to the photocurrent. However, carriers can be lost by nonradiative recombination, which decreases the efficiency further. An excellent description of such an efficiency analysis can be found in [21]. In order to understand local features of the carrier transport, such as surface recombination effects, or defect recombination related to specific doping species or materials, a space-resolved carrier transport model is needed. In nanostructures, additional physical properties are potentially relevant, e.g., the thermal relaxation process is different from bulk processes, which can lead to ballistic or semiballistic carrier transport. If these heated carriers can be detected by a contact without thermal relaxation, a higher efficiency is achieved as the thermalization energy loss can be avoided [22].

Therefore, electronic properties of nanostructures require simulations by quantum transport models in order to include effects such as carrier tunneling, phase coherence (and carrier localization), and ballistic transport. These models have been successfully applied to nanowire devices by several groups [23] and established methods are the nonequilibrium Green's functions (NEGF) method or the Wigner function formalism [24]. The NEGF method has been applied to quantum solar cells [25,26]; however, these methods in general are computationally demanding, as they discretize both real and phase space. Therefore, simulation domains of only a few nanometers can be accounted for. In the future, the so called multiscale approaches might be feasible that combine computationally intense quantum transport methods with efficient semiclassical methods; several approaches have been made, with still many open questions [27].

In the following sections, the carrier transport in nanowire solar cells is analyzed. It will turn out that these solar cells have dimensions with diameters larger than 50 nm and lengths more than 1000 nm. In this regime, semiclassical models still have a decent validity, therefore the established and efficient drift-diffusion model will be used. It comprises the self-consistent solution of the current continuity equations for electrons and holes with the Poisson equation for the electrostatic potential.

42.4 Light Capture and Emission

Figure 42.2 shows the arrangement of the nanowire array for a solar cell. The wire diameter is d, the lattice constant is a, and the wire height is L. The first question is how such an array can incouple and absorb light. Solar radiation spans from the ultraviolet to the infrared wavelength range, and the solar cell therefore needs to be a broadband absorber. It has been reported in the literature that an array of nanowires can act as a good broadband absorber [16,28]. For highest efficiency with a single bandgap material, the classical Shockley–Queisser analysis predicts a band gap of the semiconductor material between $E_g \approx 1.3$ eV and 1.4 eV [11].This bandgap optimum can be explained by the balance of having minimum carrier thermalization losses from high-energy photons on the one hand and minimizing the losses due

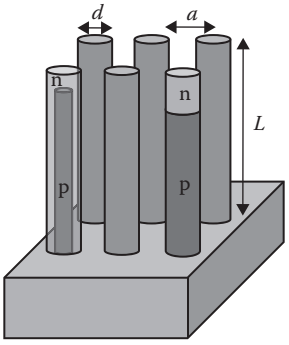

FIGURE 42.2 Illustration of an ordered nanowire array solar cell. The p-n junction can be arranged in a radial (left) or axial (right) manner. The periodicity can be hexagonal or cubic.

to the optical transparency of the semiconductor for low-energy photons. Indium phosphide (InP) with $E_g = 1.34eV$ matches this range very well; therefore, the following analysis will use this material. The findings are applicable to any other material as well.

42.4.1 Absorptivity

Using a three-dimensional finite-element solution of the complex vectorial Maxwell equations, the wire arrangement as shown in Figure 42.2 is illuminated from the top side with a linearly polarized sheet source. The source matches the power spectrum of AM1.5D solar power and is placed at a distance above the wires, which is beyond the evanescent mode decay length. The space between the wires is vacuum, and both the wire and the substrate material is InP, with the complex refractive indices following the dispersion relation given in [12]. The simulation is carried out for a quarter of the wire only, with periodic boundary conditions on the cut planes in order to represent an infinite periodicity of the array . Perfectly matched layer absorbing boundary conditions are placed at the top and the bottom of the simulation domain.

From the electromagnetic fields, the relative local absorptivity $a(\vec{x}, \lambda)$ is calculated according to Equation 42.12, which becomes, after integration of the wire volume, the spectral absorptivity of the entire wire $\tilde{a}(\lambda)$. The optical absorption in the substrate is not included in the integration and does not contribute to the efficiency of the cell. Figure 42.3 shows the spectral absorptivity for different wire periodicity and diameters; the wire height is constant at $L = 2$ μm. A geometric fill factor is defined that describes the volume of active material in a nanostructured cell relative to a bulk material volume. This fill factor can be expressed as

$$f_{nw} = \frac{\pi d^2}{4a^2}. \tag{42.30}$$

The highest fill factor can be achieved with $d = a$ and a fill factor of $f_{nw} \approx 0.79$, i.e., when the nanowires touch. For the simulations shown in Figure 42.3, combinations of periodicity and diameter have been

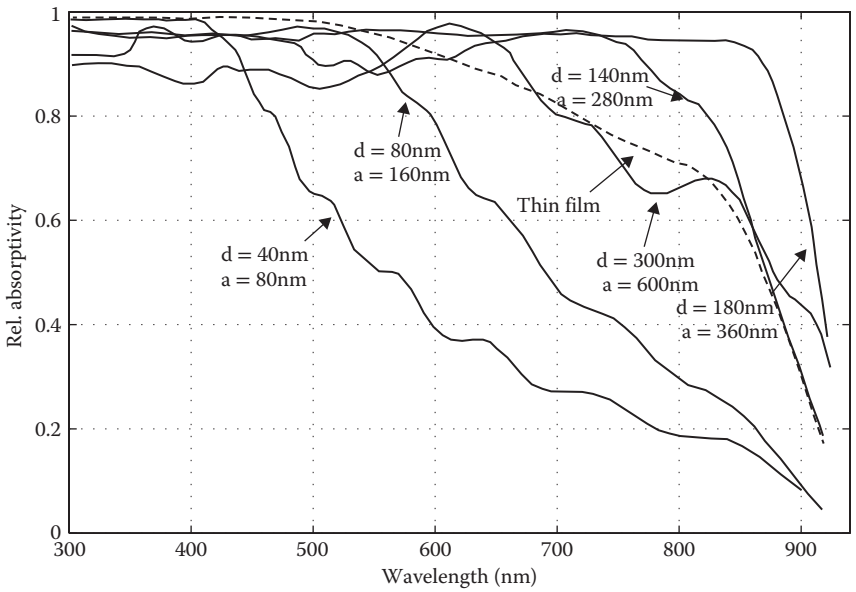

FIGURE 42.3 Spectral relative absorptivity of an InP nanowire array. The wire diameter is labeled d, and the wire distance is a.

chosen so that the material volume fill factor is identical at $f_{nw} = 0.196$. The arrangement with the smallest wire diameter of $d = 40$ nm and distance of $a = 80$ nm shows a substantial reduction of absorptivity above a wavelength of 500 nm. Nanowires with larger diameter and larger array pitch show a better long wavelength absorptivity, and the case of $d = 180$ nm and $a = 360$ nm has an absorptivity of more than 90% up to the bandgap wavelength of InP. The dashed curve shown in Figure 42.3 is the absorptivity of a homogeneous InP film with a thickness of $2\mu m \times 0.196 = 392$ nm, which is the same material volume as the nanowire array. This thin film, however, is covered by an ideal antireflective coating in order to eliminate reflections (which is not added to the wire array case). The decrease of absorptivity at wavelengths larger than 600 nm comes from the insufficient thickness of the InP layer.

42.4.2 Modal Analysis

In order to understand the findings of Section 42.4.1, the spectral behavior of the absorptivity of the nanowire array can be analyzed by a modal description of the electromagnetic field distribution. In [16], it has been shown that the absorptivity obtained from the three-dimensional calculation can be constructed from a separation of the power absorptivity parallel to the nanowire axis according to the Beer–Lambert law. The effective attenuation coefficient is obtained from the two-dimensional mode calculation of a cross-section of the wire array perpendicular to the wire axis. This assumes an infinite extension of the wires along their axes, and the calculation gives the eigenmodes with their complex eigenvalues $n_m^{\text{eff}} = n'_m + in''_m$, with m as the mode index. The real part of the eigenvalue describes the coupling efficiency of the electromagnetic incident TEM wave to the wire array and is inserted in the separation ansatz for the calculation of the reflected power. The imaginary part is the effective absorptivity that enters the Beer–Lambert expression. For an illustration, see Figure 42.4. The optical power at the interface between vacuum and the nanowire array top can be expressed as

$$P_m(z=0) = P_0 \frac{4n_0 n'_m}{(n_0 + n'_m)^2} \frac{\Gamma_m}{\sum_i \Gamma_i}, \tag{42.31}$$

with n_0 being the vacuum refractive index, P_0 the source power, and Γ_m is the overlap integral of the electric field of the incoming TEM mode with the mth mode of the wire array cross-section. The absorbed power in the nanowire material is then

$$P_{\text{abs}}(\lambda) = \sum_i P_i(z=0)\left(1 - \exp\left(\frac{4\pi}{\lambda} n''_i L\right)\right). \tag{42.32}$$

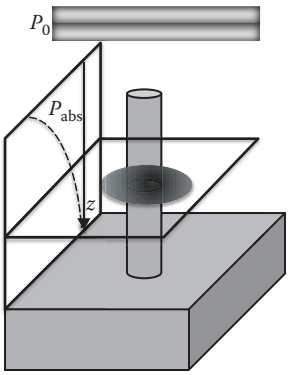

FIGURE 42.4 Illustration of the electromagnetic separation approach for a nanowire array. The incoming plane wave power P_0 is separated into two-dimensional cross-sectional modes and an exponential power decay in z-direction following a Beer–Lambert law.

Using this separation ansatz, the spectral absorptivity curves of the three-dimensional calculation can be reproduced with excellent accuracy, as shown in [28].

This procedure identifies the type of modes that couple the linearly polarized free-space electromagnetic field to the nanowire array. The unit cell of the array consists of the wire itself (which resembles a cylindrical fiber) and the periodic boundary to the neighboring wires that can be viewed as the Neumann boundary condition at the border of the unit cell.

The broadband absorptivity for the nanowire array can be described by only very few modes. For the case of $d = 180$ nm and $a = 360$ nm, only two modes are necessary to create the absorptivity characteristic as shown in Figure 42.3. These modes are plotted in Figure 42.5 with their electric field mode profile and their relative incoupling efficiency. The latter describes the ability of the wire array to act as an antireflective structure.The effect of relative incoupling in Figure 42.5 (bottom figure) is a combination of the confinement factor Γ and the real part of the effective index of the in-plane mode in the nanowire array n' (see Equation 42.31). For incoupling of 90%, the effective index therefore is close to 1 (i.e., the surrounding air), which for a fiber mode, is the regime of weak confinement. Still, most of the electromagnetic power couples into the wire array, and the field distribution as shown in in Figure 42.5 (e.g., mode A at 500 nm) shows a strong field between the wires, which is also absorbed according to the Beer–Lambert law. Between 300 nm and 700 nm, mode A couples close to 90% into the wire, and the electric field analysis shows that the electric field distribution resembles an fiber mode. Finally, the electric field distribution between the wires resembles a fundamental TEM mode, meaning parallel arrangement of the electric field vector. At 750 nm, mode A reaches its cutoff wavelength, and most of the light couples into mode B, which is of type HE_{11}. For further detail, the reader is referred to the optical fiber standard dispersion relations and eigenmodes, which are similar to the case of the sparse wire array.

It is surprising that only two modes are responsible for the broadband absorptivity of the electromagnetic power. This modal description helps understand the characteristics: If the wire spacing is increased from the ideal distance, the overlap integral of the wire modes decreases, and therefore the absorptivity decreases. If the wire spacing is decreased, the effective index of the wire array n' increases, and therefore the relative incoupling decreases. If, at the same distance, the wire diameter is modified, similar effects occur; an increase of the wire diameter moves the fiber mode dispersion relation to the multimode regime, where the effective indices are larger (a standard optical fiber dispersion relation illustrates this). A decrease of the wire diameter will decrease the mode number to a point where only the fundamental mode can be used.

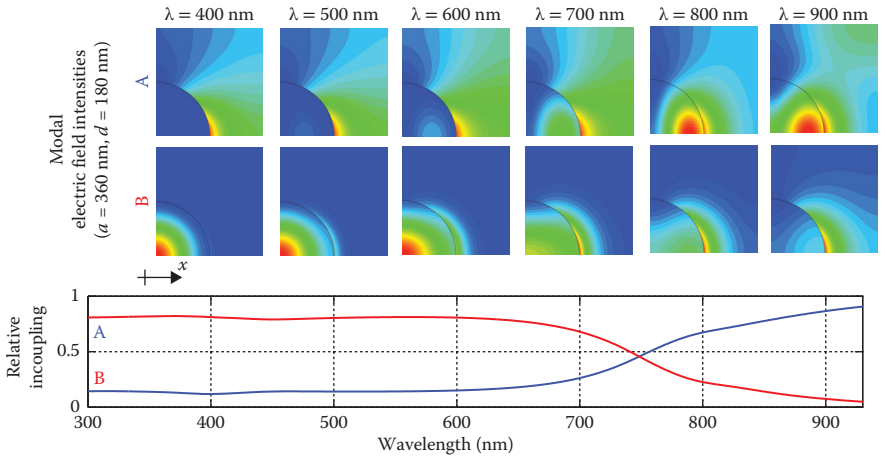

FIGURE 42.5 Top: Electric field distribution in the nanowire array for mode A and mode B, for different wavelengths. Bottom: Relative incoupling P_A/P_0 and P_B/P_0 (see Equation 42.31). (Reprinted from J. Kupec et al., *Optics Express*, 18, 27589–27605, 2010. With permission.)

As a rule of thumb, therefore, the wire distance and diameters should be chosen so that in the dispersion relation of the cross-sectional eigenvalue problem, the bandwidth of operation can span the fundamental and a few higher order modes. Moreover, the wire length should be chosen so that the imaginary part of the eigenvalue ensures full absorption of the optical power according to the Beer–Lambert law.

The modal analysis of the wire array can also be applied for solar radiation under an incident angle different from vertical illumination. For moderate incident angles (below 15°), the results do not change significantly.

42.4.3 Shockley–Queisser Efficiency of a Nanowire Array

Having calculated the relative absorptivity from electromagnetics, the Shockley–Queisser efficiency can subsequently be evaluated from Equation 42.21. Carrier transport effects are not included at this level of description and will be described later. Also, as approximation, the optical DOS in the semiconductor $D_{\text{opt}}^{\text{sc}}$ is taken to be a bulk DOS, for simplicity. This neglects the Purcell effect for the radiative emission, and therefore we get an upper limit for the solar cell efficiency. The impact of the modified optical DOS will be discussed in Section 42.4.4.

Figure 42.6 shows the calculated efficiency of the InP nanowire array with a square wire arrangement and a wire length of $L = 2\,\mu\text{m}$. The efficiency is plotted versus the volumetric fill factor. For small wire diameters of below 100 nm, the efficiency is highest for high fill factors and decreases for larger wire distances. Here, mostly the fundamental HE_{11} mode is responsible for the absorptivity; however, the overlap with the wire is too small for efficient absorption. For larger wire distance, the electromagnetic energy therefore is lost in the substrate. For thicker wires with diameters larger than 100 nm, the most dense array with highest fill factor shows the same efficiency as the thinner wires; however, with increasing wire distance, the efficiency increases up to values of 30% for the diameter of $d = 180$ nm. This increase can be explained by a reduced effective index of the wire array with lower fill factor, and therefore less back reflection from the vacuum–wire array interface. The electromagnetic energy is concentrated to the wires by the modal light capture, and the decrease of efficiency below fill factors of 10% is due to the decreasing coupling factor. As a result, a maximum efficiency of 30% can be reached with a fill factor of approximately 20%, a wire diameter of $d = 180$ nm, and a wire separation of $a = 360$ nm, without an additional anti-reflective coating. Hence, the wire array acts as a microconcentrator element as well as an active absorber at the same time. An InP bulk cell with ideal broadband antireflective coating gives a detailed balance efficiency of $\eta_{\text{ideal}} = 31\%$ for comparison.

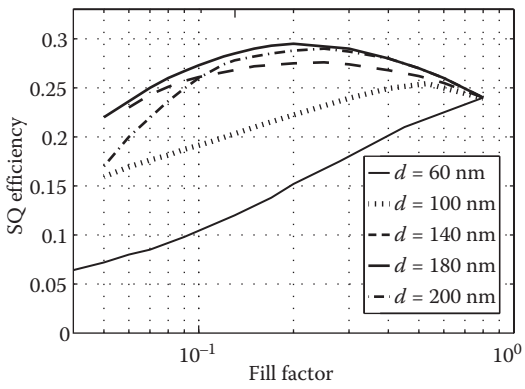

FIGURE 42.6 Efficiency of an InP nanowire array based on the Shockley–Queisser analysis in the radiative limit for different wire diameters d. An infinitely thick, perfect antireflective InP bulk cell has an efficiency of approximately 32%.

42.4.4 Purcell Effect

The nanowire array acts as a microconcentrator, and therefore, as discussed in Section 42.2.4.2, the open-circuit voltage can be increased compared to a bulk solar cell, if the concentration factor outweighs the increase in optical DOS. The reciprocity theorem states that any lens or concentrator not only increases the collection angle of the solar cell, but also the emission angle. In addition, the spontaneous emission rate (which in the bulk case is a material property) can be altered if the dielectric surrounding of the emission volume is modified, which changes the recombination rates. For the InP nanowire model array, the question is now how large the Purcell factor is and if it substantially alters the efficiency considerations described up to now.

In order to calculate the optical DOS, dipole sources were placed in the nanowire array, with vertical (z-) and horizontal (xy-) polarization. As position, the cross-section of the nanowire was sampled at a height of $z_0 = 1.9$ μm, where the absorptivity is large. A wavelength sample of $\lambda = 700, 800$, and 900 nm was used, which is close to the spontaneous emission maximum in InP. A description similar to Equation 42.7 has been used in order to calculate the local optical density of states (LDOS) (for a detailed procedure, please see [19]). In Figure 42.7, the LDOS is plotted, normalized to the free-spaced DOS. It basically describes the enhancement or suppression factor for the spontaneous emission rate; a number larger than 1 would mean that spontaneous emission is enhanced. From the calculation, the z-polarized normalized LDOS reaches values around 2–3, close to the wire surface, where the maximum absorption takes place (see Figure 42.5 for comparison). This is a photonic crystal effect, as the z-polarized dipole emits predominantly in the horizontal direction. The xy-LDOS is 0.5–0.8, close to unity. As a consequence, the averaged LDOS over all directions leads to a normalized LDOS very close to unity, which according to Equation 42.27 allows the full benefit of the optical microconcentration effect to enter the open-circuit voltage calculation.

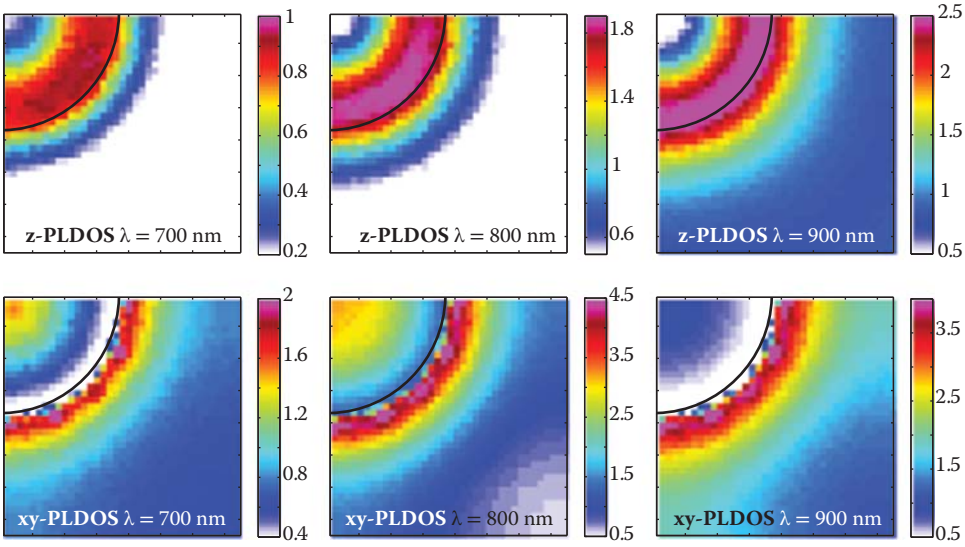

FIGURE 42.7 Local optical density of states (LDOS) relative to the bulk density of states in a nanowire array with diameter of $d = 180$ nm and pitch $a = 360$ nm. The top row is from an emitter polarized parallel to the wire axis, the bottom row for a dipole emitter with horizontal polarization. (Reprinted from J. Kupec et al., *Optics Express*, 18, 27589–27605, 2010. With permission)

42.5 Electronic Properties

42.5.1 Radial versus Axial Junction Arrangement

The nanowire geometry by nature has a high aspect ratio of length to diameter, and the p-n junction can be arranged in radial (i.e., vertical) or axial (i.e., horizontal) manner generating large or small quasi-neutral regions. Both implementations are shown in Figure 42.2 schematically and also have been demonstrated experimentally by epitaxial growth techniques. This p-n junction placement introduces another important aspect to nanostructured solar cell design. Assuming an abrupt junction, and the depletion approximation, the p-n junction can be separated into three distinct regions: the depleted region, and the quasi-neutral p- and n-regions. Photogenerated carriers in the depleted region are separated by the electric field and constitute a local photo current. In the quasi-neutral regions, the carriers diffuse, ideally to the contacts, if a local gradient is present. The cell current at the contacts is reduced if photogenerated carriers recombine on their way to the contact. This can be internal radiative or nonradiative recombination, consisting of spontaneous emission, Auger recombination, and Shockley–Read–Hall recombination. As a general guideline for minimizing the influence of nonradiative mechanisms, the recombination limited carrier lifetime needs to be much larger than the transit time of the carriers to the contacts. This principle is used in the thin-film solar cell technology.

In the analysis in Section 42.2.3, only the radiative recombination was included in the detailed balance limit. The analysis here is also done in the detailed balance limit as well, as it gives an upper boundary for the efficiency.

As example, the InP nanowire geometry is used, with a diameter of $d = 180$ nm and a wire height of 2 μm. The top n-region has a thickness of 100 nm, and the core diameter of the core-shell implementation is $d = 120$ nm. As p-doping, an active concentration of $p = 1{\times}10^{18} cm^{-3}$ and n-doping of $n = 5{\times}10^{18} cm^{-3}$ is used. The junction arrangement of the respective axial and core-shell structure is shown in Figure 42.2. The space and frequency resolved absorptivity profile has been calculated solving the vectorial Helmholtz equation under AM1.5D illumination; an illustration can be seen in Figure 42.8 after integration over frequency.

The wire array fill factor is 20%, which gives a maximum Shockley–Queisser efficiency. After spectral integration of the absorptivity, a local optical generation rate can be evaluated. The electronic simulation has been carried out only for a single wire by solving the continuity equations for holes and electrons coupled to the Poisson equation for the electrostatic potential. This means that any interaction between the

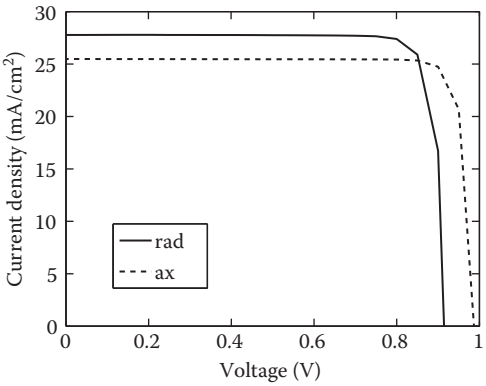

FIGURE 42.8 Current density versus voltage characteristics of axial and radial junction nanowire array solar cells made of InP, illuminated with the AM15.D solar spectrum. (ax, axial; rad, radial.)

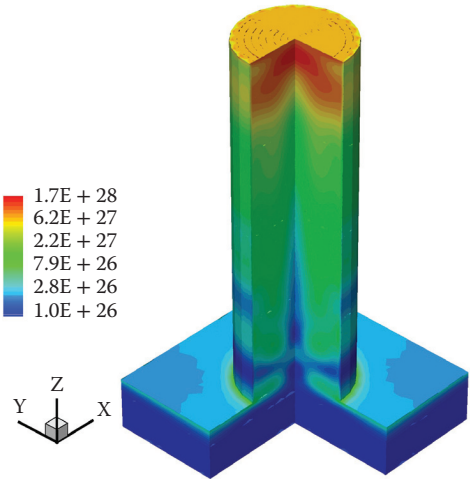

FIGURE 42.9 Optical generation rate profile in a nanowire array (only single nanowire shown) with a solar AM1.5D illumination spectrum (in units 1/s).

wires has been neglected for the carrier transport. Figure 42.9 shows the current versus voltage characteristics for the two designs. The axial p-n junction has a lower short-circuit current than the radial junction; however, it also shows a larger open-circuit voltage. The resulting maximum efficiencies are therefore close: 22% for the axial cell and 21% for the radial cell. In the axial wire, the photocurrent is generated from two effects. For one, the electric field in the depletion zone leads to drift current of the photogenerated carriers. This drift current removes carriers from the depletion zone, and in turn results in a diffusion current of the photogenerated carriers as excess minority carriers toward the junction. For the doping values in this wire, the depletion region is less than 15-nm thick at zero voltage, which would yield almost no photocurrent from direct drift. However, the simulations show that 600 nm of the quasi-neutral p-region contributes to the electron photocurrent by diffusion, and the entire quasi-neutral top n-region of 100 nm contributes to the hole photocurrent. The thickness of these diffusion regions depends critically on the minority carrier lifetime, and therefore impurity densities in the quasi-neutral regions need to be minimized. A more detailed analysis of this effect has been shown in [4]. In our simulation of the axial structure (with only radiative recombination present), a total of 700 nm of the wire height therefore can be viewed as the active region for photocurrent generation [29], which is 35% of the total length of the wire. In contrast, the radial junction contains a depletion region along the entire wire, and the charge separation due to drift and diffusion currents takes place in the entire wire volume. This leads to a higher photocurrent compared to the axial wire. The difference is not as large as the active volume ratio though, as the optical generation rate decreases exponentially from top to bottom, and the high generation top region is an active region for both structures. In [4], it has been found that the efficiency of a radial nanowire solar cell is much less prone to degradation if the minority carrier lifetime decreases, which can be directly explained by the increased depleted volume in the radial cell. It should be noted that the n-doped shell thickness of 30 nm for the radial wire has been chosen since it is larger than the depletion region so that a vertical photocurrent to the n-contact can be established.

The active region analysis explains the increased photocurrent for the radial structure, but it also can explain the reduction in open-circuit voltage compared to the axial device. The open-circuit voltage is given by Equation 42.20, with the fraction $\frac{I_{gen}}{I_{sat}}$ as contribution. Including the above described carrier transport effects, the volume integration for the calculation of the optical generation current I_{gen} and the reverse saturation current I_{sat} should only be done for the active volume. In the case of the axial junction, it is the depletion volume plus the diffusion zones in the quasi-neutral regions, which, in our specific case, account

to the top third of the wire. The radial junction wire collects photocarriers in the entire wire volume, and therefore the integration volume is the entire wire. As stated, this maximizes the optical generation current, and at the same time, also increases the reverse saturation current. Toward the bottom of the radial junction wire, the integration adds the same amount of reverse saturation current as in the upper region; however, the optical generation contribution decreases to a negligible amount. This is due to the exponential decay of the relative absorptivity a (see Equation 42.12). As result, the open-circuit voltage decreases if the wire is too long. This is plotted in Figure 42.10, where the wire array efficiency of a radial junction cell has been calculated for different wire lengths. For wire lengths below 2 μm, the solar spectrum is not fully absorbed within the wire (especially the long wavelengths), which limits the short-circuit current. For lengths above 2 μm, the reverse saturation current increases more strongly than the photocurrent, which decreases the open-circuit voltage. It should be noted that the values for the wire length and resulting efficiencies are specific to the material, dimensions, and doping concentrations that have been chosen for this example; however, the physical principles are valid for the general case of nanowire solar cells. For silicon nanowires, the optical absorption coefficient is smaller, and therefore the optimum wire length would be larger.

42.5.2 Surface Recombination

Up to now, all the calculations have been presented with radiative recombination only, excluding any nonradiative processes. In this section, the impact of surface recombination is analyzed. Nanostructures show a large surface to volume ratio, and any surface effect has a pronounced impact on the characteristic behavior of a device. For solar cells, surface recombination contributes to the nonradiative recombination mechanisms and decreases the cell efficiency, as discussed in the previous section. Here, the goal is to study the impact of surface recombination in the cell efficiency within the framework of the carrier continuity equation and the Poisson equation. The surface recombination rate is calculated via a Shockley–Read–Hall recombination with 2D trap densities [30]:

$$R_s = \frac{n_s p_s - n_i^2}{1/v_p(n_s + n_1) + 1/v_n(p_s + p_1)}, \tag{42.33}$$

where n_i is the intrinsic carrier concentration, v_n and v_p represent the surface recombination velocities, and n_1 and p_1 represent the surface trap densities. This model is used in the simulation procedure. The surface carrier densities are n_s and p_s, respectively.

The radial and axial structures have been simulated with constant surface recombination velocity of $v_s = 10^4$ cm/s, which is a realistic value for InP [31], and surface trap densities n_1 and p_1 set to a value for midgap states . Figure 42.11 shows the current–voltage characteristics with (dashed lines) and without surface recombination (solid lines). Turning on surface recombination, the maximum efficiency drops

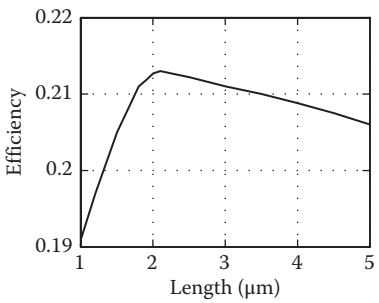

FIGURE 42.10 Cell efficiency of a radial junction nanowire array versus nanowire length.

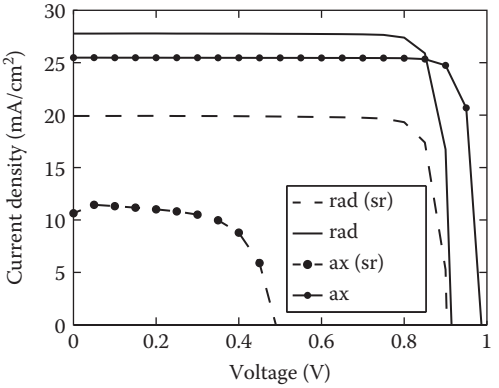

FIGURE 42.11 Current versus voltage relationship of a radial (rad) and axial (ax) junction nanowire array with and without surface recombination (sr).

from 21% to 15.5% for the radial structure and from 22% to 3.5% for the axial structure. It can be shown that the maximum surface recombination rate is achieved for the condition $v_n n_s = v_p p_s$ [32]. Assuming similar recombination velocities for electrons and holes for simplicity, the maximum surface recombination is hence located where electron and hole densities are both high at the same location. For a p-n junction, this occurs in the depletion region. It explains why the axial junction solar cell shows a substantial decrease in efficiency; in this geometry, the depletion region is exposed to the surface. This can be avoided by adding a passivation layer to the surface. For the radial junction, the depletion region thickness is smaller than the outer n-doped shell thickness, which makes the surface a quasi-neutral region. In that respect, the radial junction arrangement is vastly superior and much less prone to degradation from surface recombination effects.

42.6 Summary

Photovoltaic cells have become a fast growing technology for electrical power generation. Conversion efficiency and cost have been major drivers for the development. In that context, the use of nanostructures in solar cells has not lead to a major performance advantage up to now, and most implementations rely on classical p-n junction electronics and classical ray optics. In this chapter, the use of periodic nanowire arrays and their impact on solar cell efficiency has been analyzed. An efficiency analysis is presented that includes the nature of wave optics based on electromagnetic propagation and the optical DOS. The former leads to resonance and interference effects, which in turn causes localization of the electromagnetic energy and spectral dispersion. As main result, the efficiency of a bulk cell with ideal antireflective coating can be achieved with wires of only 20% of material without antireflective coating. In addition, efficient light management improves the solar cell efficiency via a modification of the optical DOS. This leads to a higher open-circuit voltage compared to a classical solar cell.

Acknowledgements

This work has been supported by the EU FP7 grant AMON-RA (grant agreement FP7-214814-1). The author acknowledges the contribution of J. Kupec, S. Yu on the numerical simulation results.

Appendix 42.A: Material Parameters

The InP material parameters used in the simulations are listed here. The refractive index dispersion is shown in Figure 42A.1.

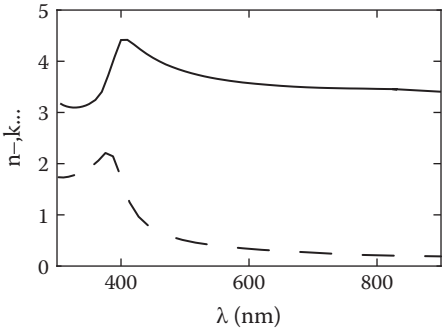

FIGURE 42A.1 InP refractive index n and extinction k versus wavelength (From S. L. Chuang, *Physics of Photonic Devices. Wiley Series in Pure and Applied Optics*, 2009. Copyright Wiley-VCH Verlag GmbH & Co. KGaA).

As mobility model for the electronic simulations, the Arora model has been used:

$$\mu = \mu_{\min} + \frac{\mu_D}{1 + (N_i/N_0)^A}, \tag{42A.1}$$

with N_i being the impurity density, and the remaining parameters have been chosen as [33].

	$\mu_{min}(m^2/V \cdot s)$	$\mu_D(m^2/V \cdot s)$	$N_0(m^{-3})$	A
el	0.040	0.520	3.00e23	0.47
hl	0.001	0.017	4.87e23	0.62

The intrinsic carrier density is $n_i = 1.2 \times 10\ m^{14}m^{-3}$, the bandgap energy is $E_G = 1.34$ eV, and the radiative recombination coefficient is $1.2 \times 10\ m^{-16}m^3/s$.

References

1. L. Tsakalakos, J. Balch, J. Fronheiser, M.-Y. Shih, S. F. Leboeuf, M. Pietrzykowski, P. J. Codella, B. A. Korevaar, O. Sulima, J. Rand, and A. Davuluru. Strong broadband optical absorption in silicon nanowire films. *Instrumentation*, 1:1–10, 2007.
2. E. Garnett and P. Yang. Light trapping in silicon nanowire solar cells. *Nano Letters*, 10:1082–1087, 2010.
3. J. Wallentin, N. Anttu, D. Asoli, M. Huffman, I. Berg, M. H. Magnusson, G. Siefer, P. Fuss-Kailuweit, F. Dimroth, B. Witzig-Mann, H. Q. Xu, L. Samuelson, K. Deppert, and M. T. Borgstrm. InP nanowire array solar cells achieving 13.8% exceeding the ray optics limit. *Science*, 339(6123):1057–1060, 2013.
4. B. Kayes, H. Atwater, and N. Lewis. Comparison of the device physics principles of planar and radial pn junction nanorod solar cells. *Journal of Applied Physics*, 97(11):114302, 2005.
5. C. Boecklin, R. G. Veprek, S. Steiger, and B. Witzigmann. Computational study of an InGaN/GaN nanocolumn light-emitting diode. *Physical Review B*, 81:155306, 2010.
6. G. Mariani, A. C. Scofield, C.-H. Hung, and D. L. Huffaker. GaAs nanopillar-array solar cells employing in situ surface passivation. *Nature Communications*, 4:1497, 2013.

7. W. Shockley and H. J. Queisser. Detailed balance limit of efficiency of pn junction solar cells. *Journal of Applied Physics*, 32(3):510–519, 1961.
8. E. M. Purcell, H. C Torrey, and R. V Pound. Resonance absorption by nuclear magnetic moments in a solid. *Physical Review*, 69(1–2):37–38, 1946.
9. Ivan S. Nikolaev, Willem L. Vos, and A. Femius Koenderink. Accurate calculation of the local density of optical states in inverse-opal photonic crystals. *Journal Optical Society of America B*, 26(5):987–996, 2009.
10. E. Yablonovitch. Statistical ray optics. *Journal of the Optical Society of America*, 72(7):899–907, 1982.
11. J. Nelson. *The Physics of Solar Cells*. London: Imperial College Press, 2003.
12. S. L. Chuang. *Physics of Photonic Devices. Wiley Series in Pure and Applied Optics*, 2nd edition. New York, NY: Wiley, 2009.
13. K. Barnham and G. Duggan. A new approach to high-efficiency multi-band-gap solar cells. *Journal of Applied Physics*, 67:3490–3493, 1990.
14. A. Luque and A. Marti. Increasing the efficiency of ideal solar cells by photon induced transitions at intermediate levels. *Physical Review Letters*, 78:5014–5017, 1997.
15. R. T. Ross and A. J. Nozik. Efficiency of hot-carrier solar energy converters. *Journal of Applied Physics*, 53(5):3813–3818, 1982.
16. J. Kupec and B. Witzigmann. Dispersion, wave propagation and efficiency analysis of nanowire solar cells. *Optics Express*, 17(12):10399–10410, 2009.
17. A. Taove and S. C. Hagness. *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd edition. Norwood, MA: Artech House, 2005.
18. B. Cockburn and C.-W. Shu. The Runge-Kutta discontinuous Galerkin method for conservation laws v: Multidimensional systems. *Journal of Computational Physics*, 141:199–224, 1998.
19. F. Roemer, B. Witzigmann, O. Chinellato, and P. Arbenz. Investigation of the Purcell effect in photonic crystal cavities with a 3D finite element nanowire solar cells: Electro-optical performance Maxwell solver. *Optical and Quantum Electronics*, 39(4–6):341–352, 2007.
20. F. Roemer and B. Witzigmann. Spectral and spatial properties of the spontaneous emission enhancement in photonic crystal cavities. *Journal Optical Society of America B*, 25(1):31–39, 2008.
21. O. D. Miller, E. Yablonovitch, and S. R. Kurtz. Strong internal and external luminescence as solar cells approach the Shockley–Queisser limit. *IEEE Journal of Photovoltaics*, 2(3):303–311, 2012.
22. D. Knig, K. Casalenuovo, Y. Takeda, G. Conibeer, J. F. Guillemoles, R. Patterson, L. M. Huang, and M.A. Green. Hot carrier solar cells: Principles, materials and design. *Physica E: Low-Dimensional Systems and Nanostructures*, 42(10):2862–2866, 2010.
23. M. Luisier and G. Klimeck. Atomistic full-band simulations of silicon nanowire transistors: Effects of electron-phonon scattering. *Physical Review B*, 80:155430, 2009.
24. W. Frensley. Wigner-function model of resonant-tunneling semiconductor device. *Physical Review B*, 36(3):1570–1580, 1987.
25. U. Aeberhard and R. Morf. Microscopic nonequilibrium theory of quantum well solar cells. *Physical Review B*, 77(12):125343, 2008.
26. U. Aeberhard. Simulation of nanostructure-based and ultra-thin film solar cell devices beyond the classical picture. *Journal of Photonics for Energy*, 4(1):042099, 2014.
27. M. Auf der Maur, G. Penazzi, G. Romano, F. Sacconi, A. Pecchia, and A. Di Carlo. The multiscale paradigm in electronic device simulation. *IEEE Transactions on Electron Devices*, 58(5):1425–1432, 2011.
28. J. Kupec, L. R. Stoop, and B. Witzigmann. Light absorption and emission in nanowire array solar cells. *Optics Express*, 18(26):27589–27605, 2010.
29. S. Yu, J. Kupec, and B. Witzigmann. Efficiency analysis of III-V axial and core-shell nanowire solar cells. *Journal of Computational and Theoretical Nanoscience*, 9(5):1–8, 2012.
30. F. Roemer and B. Witzigmann. Modelling surface effects in nanowire optoelectronics. *Journal of Computational Electronics*, 11(4):431–439, 2012.

31. S. Bothra, S. Tyaqi, S. K. Ghandhi, and J. M. Borrego. Surface recombination velocity and lifetime in InP. *Solid-State Electronics*, 34:47–50, 1991.
32. P. P. Altermatt, A. G. Aberle, J. Zhao, A. Wang, and G. Heiser. A numerical model of p-n junctions bordering on surfaces. *Solar Energy Materials and Solar Cells*, 74:165–174, 2002.
33. M. Sotoodeh, A. H. Khalid, and A. A. Rezazadeh. Empirical low-field mobility model for III–V compounds applicable in device simulation codes. *Journal of Applied Physics*, 87(6), 2000.